Semi-Blind Channel Estimation for LTE DownLink

Laureando:
Nicolò Michelusi

Facoltà di
Ingegneria delle Telecomunicazioni

Supervisori:
Tomaso Erseghe (DEI)
Lars Christensen (Nokia, Denmark)
Ole Winther (DTU, Denmark)

September 2009
In a MIMO system the number of channel parameters is much larger than in a typical SISO scenario, making the channel estimation task particularly critical. In fact, this increase in the number of channel parameters translates into a smaller estimation accuracy, which is counteracted by transmitting a longer pilot sequence. This in turn negatively impacts the bandwidth efficiency of the system, making pilot based approaches less attractive.

In this thesis we investigate the Semi-Blind approach to channel estimation in MIMO-OFDM systems, and in particular for LTE downlink. This technique, by exploiting the observations associated to the unknown symbols other than the pilot sequence to perform the channel estimate, potentially leads to an improvement in the estimation accuracy compared to the typical pilot based estimation approach, without requiring a long pilot sequence, despite the large number of parameters typical of a MIMO scenario.

Through simulations performed on the LTE system we show that the proposed Semi-Blind approaches lead to significant improvements in the estimation accuracy, both from an MSE and BER perspective, compared to the typical pilot based technique. However, exploiting the true discrete distribution of the unknown symbols is computationally demanding, therefore we propose the use of two approximations on the unknown symbols: the Gaussian and the Constant Modulus assumptions. These, though sub-optimal from a point of view of the estimation accuracy, still lead to significant improvements with respect to the pilot based approach, while reducing the computational overhead incurred when using true discrete distribution of the unknown symbols.
Contents

Abstract iii

List of Figures vii

1 Introduction 1
   1.1 Channel Estimation in MIMO systems 3
   1.2 MIMO-OFDM principles and system model 4
      1.2.1 MIMO model 4
      1.2.2 MIMO-OFDM model 5
      1.2.3 Model Assumptions 8
   1.3 Problem Formulation 9

2 Training sequence channel estimation of MIMO-OFDM FIR channels 11
   2.1 Maximum-Likelihood channel estimation of a MIMO-OFDM FIR channel 12
      2.1.1 Channel Identifiability Conditions 16
      2.1.2 Properties of ML channel estimator 18
         2.1.2.1 Bias of Maximum Likelihood channel estimator 19
         2.1.2.2 Variance of Maximum Likelihood channel estimator 20
      2.1.3 White Gaussian Noise at the receiver 22

3 Semi-Blind channel estimation 27
   3.1 General formulation of Semi-Blind ML estimation of MIMO-OFDM FIR channels 28
      3.1.1 Brief introduction to the EM-Algorithm 33
      3.1.2 ML solution through EM-algorithm 34
   3.2 Semi-Blind ML estimation: true discrete distribution of the unknown symbols 38
      3.2.1 ML solution through EM-algorithm 39
   3.3 Semi-Blind ML estimation: Gaussian approximation for the unknown symbols 40
      3.3.1 ML estimate through EM Algorithm 43
   3.4 Semi-Blind ML estimation: Constant Modulus approximation for the unknown symbols 44
      3.4.1 ML solution through EM-algorithm 45

4 Joint Semi-Blind Estimation of channel and noise covariance matrix 55
   4.1 Noise Model 56
## CONTENTS

1.2 Noise Covariance matrix Estimation ........................................... 61
1.3 Joint Semi-Blind Estimation of channel and noise covariance matrix ........................................... 66
   1.3.1 Pilot based approach ......................................................... 66
   1.3.2 Semi-Blind approach ......................................................... 68

5 Simulation Results and Discussion ................................................. 75
5.1 LTE frame structure ................................................................. 75
5.2 Simulation setup .................................................................. 77
5.3 Comparison of Semi-Blind and pilot based approaches for different antenna setups ........................................... 78
   5.3.1 $1T \times 1R$ MIMO ................................................................. 79
   5.3.2 $1T \times 2R$ MIMO ................................................................. 81
   5.3.3 $2T \times 1R$ MIMO ................................................................. 83
   5.3.4 $2T \times 2R$ MIMO, transmission rank $S = 1$ ......................................... 86
   5.3.5 $2T \times 2R$ MIMO, transmission rank $S = 2$ ......................................... 88
5.4 Estimation accuracy as a function of the sub-carriers ........................................... 90
5.5 Estimation accuracy as a function of the constellation order ........................................... 91
5.6 Convergence of the EM-Algorithm, Gaussian approximation ........................................... 93
5.7 Joint Estimation of Channel and noise covariance matrix ........................................... 95

6 Conclusion ........................................................................... 97

A Complex derivatives ................................................................ 101

B Computation of the posterior mean of constant modulus symbols ........................................... 103

C Cramér–Rao lower bound ......................................................... 107
   C.1 Unbiased Cramér–Rao lower bound for Complex parameters ........................................... 107
   C.2 Unbiased CRLB for pilot based estimator of MIMO-FIR channels ........................................... 110
      C.2.1 The Fisher Information Matrix for the estimation of $h$ ........................................... 111
   C.3 Unbiased CRLB for Semi-Blind estimation of MIMO-OFDM FIR Channels ........................................... 113

Bibliography ........................................................................ 119
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Illustration of a $2T \times 2R$ MIMO system, with channel estimator</td>
<td>2</td>
</tr>
<tr>
<td>3.1</td>
<td>$\tilde{g}_N(x)$ for different values of $N$</td>
<td>50</td>
</tr>
<tr>
<td>3.2</td>
<td>Plot of function $g(x)$ and its approximation $1 - e^{-1.067x}$</td>
<td>51</td>
</tr>
<tr>
<td>3.3</td>
<td>Gaussian approximation versus CM with uniform phase approximation, standard deviation on the posterior expectation: $N = L = 1, R = T = 1$</td>
<td>52</td>
</tr>
<tr>
<td>5.1</td>
<td>LTE frame structure</td>
<td>75</td>
</tr>
<tr>
<td>5.2</td>
<td>Pilot allocation on one resource block (12 sub-carriers times 7 OFDM symbols) for the cases 1, 2 and 4 transmitting antennas</td>
<td>76</td>
</tr>
<tr>
<td>5.3</td>
<td>Comparison of pilot based and Semi-Blind approaches (MSE), $1T \times 1R$</td>
<td>80</td>
</tr>
<tr>
<td>5.4</td>
<td>Comparison of pilot based and Semi-Blind approaches (BER), $1T \times 1R$</td>
<td>81</td>
</tr>
<tr>
<td>5.5</td>
<td>Comparison of pilot based and Semi-Blind approaches (MSE), $1T \times 2R$</td>
<td>82</td>
</tr>
<tr>
<td>5.6</td>
<td>Comparison of pilot based and Semi-Blind approaches (BER), $1T \times 2R$</td>
<td>83</td>
</tr>
<tr>
<td>5.7</td>
<td>Comparison of pilot based and Semi-Blind approaches (MSE), $2T \times 1R$</td>
<td>84</td>
</tr>
<tr>
<td>5.8</td>
<td>Comparison of pilot based and Semi-Blind approaches (MSE), equivalent channel, $2T \times 1R$ MIMO-OFDM, 4-QAM, 72 sub-carriers</td>
<td>84</td>
</tr>
<tr>
<td>5.9</td>
<td>Comparison of pilot based and Semi-Blind approaches (BER), $2T \times 1R$</td>
<td>85</td>
</tr>
<tr>
<td>5.10</td>
<td>Comparison of pilot based and Semi-Blind approaches (MSE), $2T \times 2R$</td>
<td>86</td>
</tr>
<tr>
<td>5.11</td>
<td>Comparison of pilot based and Semi-Blind approaches (BER), $2T \times 2R$</td>
<td>87</td>
</tr>
<tr>
<td>5.12</td>
<td>Comparison of pilot based and Semi-Blind approaches (MSE), transmission rank 1, 4-QAM, 72 sub-carriers</td>
<td>88</td>
</tr>
<tr>
<td>5.13</td>
<td>Comparison of pilot based and Semi-Blind approaches (BER), $2T \times 2R$</td>
<td>89</td>
</tr>
<tr>
<td>5.14</td>
<td>Comparison of pilot based and Semi-Blind approaches for different number of sub-carriers (MSE), $1T \times 2R$ MIMO-OFDM, 4-QAM</td>
<td>90</td>
</tr>
<tr>
<td>5.15</td>
<td>Comparison of pilot based and Semi-Blind approaches for different number of sub-carriers (BER), $1T \times 2R$ MIMO-OFDM, 4-QAM</td>
<td>91</td>
</tr>
<tr>
<td>5.16</td>
<td>Comparison of pilot based and Semi-Blind approaches for different constellation orders (MSE), $1T \times 2R$ MIMO-OFDM, 72 sub-carriers</td>
<td>92</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>5.17</td>
<td>Evolution of MSE and BER over the iterations of the EM-algorithm, $1T \times 1R$ MIMO-OFDM, 4-QAM, 72 sub-carriers</td>
<td>93</td>
</tr>
<tr>
<td>5.18</td>
<td>Evolution of MSE and BER over the iterations of the EM-algorithm, $2T \times 2R$ MIMO-OFDM, transmission rank $S = 2$, 4-QAM, 72 sub-carriers</td>
<td>94</td>
</tr>
<tr>
<td>5.19</td>
<td>Evolution of MSE and BER over the iterations of the EM-algorithm, $1T \times 2R$ MIMO-OFDM, 4-QAM, 72 sub-carriers</td>
<td>94</td>
</tr>
<tr>
<td>5.20</td>
<td>Joint Estimation of channel and noise covariance matrix, MSE of channel estimator, $2T \times 2R$ MIMO-OFDM, transmission rank 2, 4-QAM, 72 sub-carriers</td>
<td>95</td>
</tr>
<tr>
<td>5.21</td>
<td>Joint Estimation of channel and noise covariance matrix (BER), $2T \times 1R$ MIMO-OFDM, 4-QAM, 72 sub-carriers</td>
<td>96</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

During the last few decades we have experienced an extraordinary growth of wireless communications, which lead to the definition of new mobile communication standards, with the aim of providing broadband ubiquitous access to the Internet. In this context, LTE (Long Term Evolution) is a 3GPP project under standardization, promising downlink data rates of up to 300Mbps. This is accomplished by employing advanced technologies at the physical layer, such as Orthogonal Frequency Division Multiplexing (OFDM) and Multiple-Input Multiple-Output (MIMO) to increase the capacity of the wireless channel.

Typically, the bandwidth available to wireless communication systems is limited by a series of factors, the most important of which is the nature of the wireless channel. A defining characteristic of the wireless channel is multipath fading, which consists in the variation of the channel strength over time and frequency, due to constructive and destructive superposition of multiple paths traveling from the transmitter to the receiver through the wireless medium.

The frequency variation of the channel is due to the fact that the signal propagates through distinct paths to the receiver, thus arriving at distinct times, causing a spreading of the channel impulse response over time, which is equivalent to frequency selectivity in the frequency domain.

The time variation of the channel is due to the fact that distinct paths encounter moving obstacles while propagating through the wireless medium. Moreover, transmitter and receiver might be moving entities. These effects cause the channel impulse response to vary over time. The time window during which the channel is assumed to be time-invariant is called time coherence, and is approximately inversely proportional to the speed of the receiver, typically of the order of magnitude of a few ms.
Typically, wireless communication is established between one transmitting and one receiving antenna (SISO, Single-Input Single-Output systems). However, the capacity achievable in such systems is severely limited by fading, since the signal is severely attenuated when the channel is in a deep fade.

In recent years, MIMO (Multiple-Input Multiple-Output) has emerged as the new antenna technology. This has been proposed as a technique to increase the capacity and the reliability of wireless channels through the adoption of multiple antennas at the transmitter and receiver sides. A basic representation of such system is depicted in figure 1.1: a sequence of bits is encoded into the transmitting antennas by means of the encoding function \( C \), and transmitted through the wireless medium. At the receiver side, the observations collected on the antenna array \((y_k^{(0)} \text{ and } y_k^{(1)})\) are processed by the detector (function \( D(\hat{h}) \)), which is responsible for recovering the original bits.

![Figure 1.1: Illustration of a $2T \times 2R$ MIMO system, with channel estimator](image)

By adopting multiple antennas at the receiver, multiple copies of the same signal propagates through independent channels. Globally, the probability that all the channels are in a deep fade is reduced, thus improving channel reliability. This technique is called Receive Diversity. A similar effect is achieved by adopting multiple antennas at the transmitter, a technique called Transmit Diversity. By adopting multiple antennas at both the transmitter and the receiver sides, multiple information streams can be multiplexed through the transmitting antenna array, a technique called Spatial Multiplexing. Compared to a SISO system, this technique allows an increase in the capacity of the overall channel by a factor proportional to the minimum between the number of receiving and the number of transmitting antennas (actually to the channel rank). We suggest the interested reader to read [1] for a thorough treatment of MIMO systems and the derivation of this result.

Although MIMO represent a solution to increase the capacity and the reliability of wireless channels, it is particularly challenging from a channel estimation perspective. This is explained in the following section.
1.1 Channel Estimation in MIMO systems

Typically, channel estimation is performed by inserting a sequence of symbols known at the receiver (termed pilot symbols) in the transmitted frame. At the receiver side then, by observing the output in correspondence of the pilot symbols, it is possible to estimate the channel. This knowledge is then fed into the detection process, to allow optimal detection of the data, as depicted in figure [1.1]. This approach is the most commonly used in communication systems, for its low computational complexity and robustness. Its drawback consists in the fact that the pilot symbols don’t carry useful information, therefore they represent a bandwidth waste. Moreover, most of the observations (those related to the unknown symbols) are discarded in the estimation process, thus representing a missed opportunity to enhance the accuracy of the channel estimate.

In a MIMO system, channel estimation is even more critical than in a SISO system. In fact, a $T \times R$ MIMO system (where $T$ and $R$ represent the number of transmitting and receiving antennas respectively) can be represented as a set of $RT$ independent SISO channels, one between each transmitting-receiving antenna pair. It is clear that the number of channel parameters to estimate in a MIMO system increases with the product $RT$. Under this condition, the pilot based channel estimation approach has a severe limitation: as we will also demonstrate in the course of the thesis, a larger number of parameters require the transmission of a longer pilot sequence. However, the transmission of a longer pilot sequence is not desirable in a communication system, since they don’t carry useful information and represent a bandwidth waste.

In this context, it becomes important to develop a new estimation approach capable of improving the channel estimation accuracy without the need to transmit a longer pilot sequence. In this thesis the solution proposed is Semi-Blind channel estimation, which consists in exploiting also the unknown information other than the pilot sequence to estimate the channel. The potential advantage, compared to the pilot based approach, consists in the fact that all the information available at the receiver is exploited in the estimation process, therefore there is a potential improvement in the achievable estimation accuracy. However, this comes at the cost of an increased receiver complexity with respect to a pilot based approach, as we will demonstrate in the course of the thesis.

We start the treatment by modeling in section [1.2] the MIMO-OFDM system, and introducing the model assumptions used throughout the thesis. In section [1.3] we briefly formalize the channel estimation problem in MIMO-OFDM systems. Then, in chapter [2] we derive a Maximum-Likelihood estimator of MIMO-OFDM channels using the typical pilot based approach. We derive in particular a relation between the estimation accuracy
Chapter 1 Introduction

and the order of the MIMO-OFDM system (that is the number of transmitting-receiving antennas), highlighting the weaknesses of this approach for MIMO systems.

In chapter 3, we treat in detail Semi-Blind channel estimation of MIMO-OFDM channels, studying in particular three cases, depending on the assumptions used for the unknown symbols in the estimation process: in the first case we exploit the true discrete distribution, in the second we approximate the distribution of the unknown symbols with a circular Gaussian distribution, in the third case we assume the symbols have constant amplitude and phase uniformly distributed in \([0, 2\pi)\) (valid only for Constant Modulus constellations). As we will show in the simulation results in chapter 5, these three assumptions represent a trade-off between estimation accuracy and complexity: while using the true discrete distribution of the unknown symbols is optimal from the point of view of the estimation accuracy, from the perspective of the computational complexity it is far too demanding; therefore the use of approximations represent a solution to reduce the computational overhead, in spite of a reduced estimation accuracy. In general, since in the Semi-Blind approach the Maximum Likelihood solution cannot be determined in closed form, the use of iterative algorithms to converge to a local maximum of the likelihood function is required. We propose the Expectation-Maximization algorithm as a general framework to solve this maximization problem, where the unknown symbols are treated as hidden variables.

So far, we have assumed that the statistical properties of the noise are known at the receiver. However, this is not the case in a real communication system. Moreover, the wireless channel is a shared medium. Consequently, the receiver undergoes interference from other users, whose statistical properties have to be estimated at the receiver. In chapter 4, we derive an algorithm for the joint estimation of the noise covariance matrix and of the channel using the Semi-Blind approach.

Finally, in chapter 5, we present some simulation results performed on the LTE system, comparing the performance achievable with the Semi-Blind approaches and the pilot-based approach described in the thesis.

1.2 MIMO-OFDM principles and system model

1.2.1 MIMO model

MIMO (Multiple-Input Multiple-Output) is the use of multiple antennas at the transmitter and receiver sides, with the purpose of combating fading and increasing the capacity of wireless communication systems.
Let $T$ and $R$ be the number of transmitting and receiving antennas, respectively. This MIMO system is labeled as $T \times R$ MIMO, and can be represented as a set of $RT$ SISO channels, one between each transmitting-receiving antenna pair. Now, let’s consider the signal at the receiver. Using, now and in the rest of the thesis, the equivalent discrete baseband model, and assuming that, during the time span we observe the evolution of the model, the channel is time-invariant (block-fading channel), each SISO channel, modeled as a Finite Impulse Response (FIR) filter of length $L$, is described by means of $L$ complex taps. Therefore, the signal received at antenna $r$ is given by the superposition of the signals transmitted by each antenna $t = 0 \ldots T - 1$, filtered through the SISO channel between antenna pairs $(r, t)$, plus the noise. This can be written as

$$y_r(k) = \sum_{l=0}^{L-1} \sum_{r=0}^{T-1} h_{r,t}(l) x_t(k-l) + \tilde{\eta}_r(k) \quad (1.1)$$

where $y_r(k)$ is the signal received on antenna $r$ at time $k$, $h_{r,t}(l)$ is the $l$th tap of the FIR SISO channel between antenna pairs $(r, t)$, $x_t(k)$ is the signal transmitted through antenna $t$ at time $k$ and $\tilde{\eta}_r(k)$ is the noise on receiving antenna $r$ at time $k$.

Now, stacking the observations, the transmitted signal and the noise at time $k$ on the column vectors $y(k)$, $\tilde{\eta}(k)$ and $x(k)$ respectively, and letting $h_l$ be an $R \times T$ matrix with entries given by the $l$th tap between each antenna pairs $(r, t)$, we can rewrite (1.1) in matrix form as

$$y(k) = \sum_{l=0}^{L-1} h_l(k-l) + \eta(k) \quad (1.2)$$

which is the Input-Output relation of a MIMO system.

### 1.2.2 MIMO-OFDM model

Now, we go one step further, and we define the input-output relation of a MIMO-OFDM system.

Orthogonal Frequency Division Multiplexing (OFDM) is a modulation technique which consists in subdividing the available spectrum into multiple sub-carriers orthogonal to each other. Each sub-carrier is then independently modulated with a low-rate data stream, and transmitted through the channel. However, by combining in the time domain the streams associated to each sub-carrier, the overall data rate achieved is much higher than the data rates associated to the single streams. The advantage of this approach consists in the fact that the frequency-selective channel is transformed into a set.
of flat-fading channels. This is possible because the bandwidth occupied by each sub-carrier is much smaller than the overall bandwidth, therefore each sub-carrier undergoes approximately a flat-fading channel.

In this thesis, we treat the implementation of OFDM using the DFT (Discrete Fourier Transform) and Cyclic Prefix, which is the actual implementation of OFDM in LTE.

Let’s consider a MIMO-OFDM system with $N$ sub-carriers, $T$ transmitting and $R$ receiving antennas ($T \times R$ MIMO). Let $X_n(k)$ be the MIMO signal transmitted on sub-carrier $n$ at time $k$ (this is a $T \times 1$ vector). With OFDM, the time domain signal is obtained with the Inverse DFT transformation, through the relation

$$x^{(k)}(p) = \frac{1}{\sqrt{N}} \sum_n X_n(k) e^{i2\pi \frac{pn}{N}} \quad p = -CP \ldots N - 1$$

Here $x^{(k)}(p)$ is the $p$th sample of the $k$th MIMO-OFDM symbol, where this latter term refers to the ordered set of the symbols transmitted on all the sub-carriers, that is \{X_n(k), n = 0 \ldots N - 1\}. These samples are then transmitted in sequence through the channel across the antennas array.

Observe that the time-domain signal is composed of two parts: $x^{(k)}(p), p = 0 \ldots N - 1$ is a whole period of the Inverse DFT, whereas $x^{(k)}(p), p = -CP \ldots -1$ is the Cyclic Prefix of length $CP$, which is added at the beginning of the time-domain stream to make the channel appear cyclic, as we show now. Notice that, since the DFT is a periodic signal with period $N$, we have $x^{(k)}(p) = x^{(k)}(N + p), p = -CP \ldots -1$, therefore the insertion of the Cyclic Prefix corresponds to the insertion of the last samples of the Inverse DFT at the beginning of the stream.

Now, observe the Input-Output relation of a MIMO system given by \[1.2\]. Since the channel is FIR of length $L$, the output of the model at time $k$ depends only on the transmitted symbols at times $k - L + 1 \ldots k$. Therefore, assuming that the Cyclic Prefix satisfies the condition $CP \geq L - 1$, the output in correspondence of the $k$th OFDM symbol, considering only the output samples $p = 0 \ldots N - 1$, depends solely on the symbols transmitted on the $k$th symbol. In fact

$$y^{(k)}(p) = \sum_{l=0}^{L-1} h_l x^{(p)}(p - l) + \tilde{\eta}^{(k)}(p)$$

$$= \frac{1}{\sqrt{N}} \sum_n \sum_{l=0}^{L-1} h_l e^{i2\pi \frac{(p-l)n}{N}} X_n(k) + \tilde{\eta}^{(k)}(p) \quad p = 0 \ldots N - 1$$

It is thus clear that Inter-Symbol Interference from previous OFDM symbols is eliminated by setting $CP \geq L - 1$. 

Then, using this assumption on the length of the Cyclic Prefix, and letting \( H_n \) be the frequency domain channel, defined as \( \sqrt{N} \) times the DFT of the time domain channel \( h_l \), we obtain

\[
y^{(k)}(p) = \frac{1}{\sqrt{N}} \sum_n H_n X_n(k) e^{i2\pi \frac{p}{N}} + \tilde{\eta}^{(k)}(p) \quad p = 0 \ldots N - 1 \tag{1.5}
\]

At the receiver, the time-domain signal is processed using the \( N \)-points DFT. On sub-carrier \( m \) we have

\[
Y_m(k) = \frac{1}{\sqrt{N}} \sum_{p=0}^{N-1} y^{(k)}(p) e^{-i2\pi \frac{pm}{N}}
\]

\[
= \frac{1}{N} \sum_n H_n X_n(k) \sum_{p=0}^{N-1} e^{i2\pi \frac{(n-m)p}{N}} + \frac{1}{N} \sum_{p=0}^{N-1} \tilde{\eta}^{(k)}(p) e^{i2\pi \frac{pm}{N}}
\]

\[
= \sum_n H_n X_n(k) \delta_{nm} + \eta_m(k) = H_m X_m(k) + \eta_m(k) \tag{1.6}
\]

where \( \eta_m \) is the noise vector on sub-carrier \( m \) at time \( k \). From this relation we see that the insertion of the Cyclic Prefix of length \( CP \geq L - 1 \) has transformed a frequency selective channel into a set of \( N \) flat-fading channels.

Finally, assuming \( K \) OFDM symbols are transmitted, and collecting the received and the transmitted signals and the noise at time \( k \) on a matrix, we have the following Input-Output relation for a MIMO-OFDM system:

\[
Y_n = H_n X_n + \eta_n \quad \text{for} \ n = 0 \ldots N - 1 \tag{1.7}
\]

Here, the subscript \( n \) represents the sub-carrier index, \( Y_n \) is the \( R \times K \) observation matrix with entries \( Y_n(r,k) \in \mathbb{C} \) representing the signal received on sub-carrier \( n \) at time \( k \) on receiving antenna \( r \), \( X_n \) is the \( T \times K \) matrix of the transmitted symbols with elements \( X_n(t,k) \in \mathbb{C} \) representing the symbol transmitted on sub-carrier \( n \) at time \( k \) from transmitting antenna \( t \), \( H_n \) is the \( R \times T \) channel matrix with entries \( H_n(r,t) \in \mathbb{C} \) representing the channel coefficient between antenna pair \( (r,t) \), and \( \eta_n \) is the \( R \times K \) noise matrix.

Now, let’s assume that the matrix of the transmitted symbols \( X_n \) is a collection of both pilot symbols, used at the receiver for performing the channel estimate, and information symbols. We assume also that, in order to suppress multi-antenna interference during the estimation process, at a generic time \( k \) on sub-carrier \( n \), either all the antennas are transmitting pilots or none of them (in this case they are all transmitting information symbols). With these assumptions, we can split the matrix of the transmitted symbols into the sum of two matrices, the former one carrying the contribution from the pilot
Chapter 1 Introduction

symbols \((X^{(tr)})\), with null entries in correspondence of the unknown symbols, the latter carrying the contribution from the unknown symbols \((X^{(bl)})\), with null entries in correspondence of the pilot symbols. Similarly, we can split the observation and noise matrices into the observation and noise matrices associated to pilot symbols \((Y^{(tr)}\) and \(\eta^{(tr)}\)) and unknown symbols respectively \((Y^{(bl)}\) and \(\eta^{(bl)}\)). Therefore, on each sub-carrier \(n\) we have the following decomposition of the observation, symbol and noise matrices:

\[
\begin{align*}
X_n &= X^{(tr)}_n + X^{(bl)}_n \\
Y_n &= Y^{(tr)}_n + Y^{(bl)}_n \\
\eta_n &= \eta^{(tr)}_n + \eta^{(bl)}_n
\end{align*}
\] (1.8)

Using this notation, we can split the Input-Output relation [1.7] as

\[
\begin{align*}
Y^{(tr)}_n &= H_n X^{(tr)}_n + \eta^{(tr)}_n \quad \text{for } n = 0 \ldots N - 1 \\
Y^{(bl)}_n &= H_n X^{(bl)}_n + \eta^{(bl)}_n \quad \text{for } n = 0 \ldots N - 1
\end{align*}
\] (1.9)

The first relation describes the input-output model associated to the pilot symbols, the second instead describes the input-output model associated to the information symbols. Notice that, in the pilot based approach to channel estimation, only the first input-output relation is considered, since only the pilot observations are used for the estimate. Conversely, in the Semi-Blind approach all the information is considered at the receiver, both \(Y^{(tr)}\) and \(Y^{(bl)}\).

1.2.3 Model Assumptions

Based on the MIMO-OFDM model described in the previous section, we now define the general assumptions used throughout the thesis. In particular, we define the assumptions on the unknown symbols and on the noise at the receiver.

As regards the unknown symbols, we assume that they are obtained by encoding across the transmitting antenna array a set of \(S\) independent streams. The model used is the following:

\[
X^{(bl)}_n = CV^{(bl)}_n
\] (1.10)

where \(C\) is a \(T \times S\) precoding matrix, which encodes \(S\) independent streams of symbols into the \(T\) transmitting antennas array, and \(V^{(bl)}_n\) is the \(S \times K\) matrix of the information symbols. The entries of this matrix are assumed to be drawn uniformly from a discrete constellation \(\mathcal{C}\), independently and identically distributed, with zero mean and mean power \(\sigma_s^2\). Therefore we have \(E[V_n(k)V_n(k)^H] = \sigma_s^2 I_S\).
Notice that matrix $C$ encodes the symbols only across the transmitting antennas, not across time. Its columns represent a set of Hadamard vectors, with the property that $C^H C = I_S$, where $I_S$ is the $S \times S$ identity matrix. Therefore, also the transmitted symbols are independent across time and across sub-carriers, but they are not necessarily independent across the transmitting antennas.

In our treatment we assume $S \leq \min\{R, T\}$, since detector performance would be severely reduced in the case $S > \min\{R, T\}$ and ‘good’ approximate detector design is significantly harder for this case. This assumption is coherent with the fact that the capacity of a MIMO system linearly increases with the minimum between the number of receiving and the number of transmitting antennas, which corresponds to the rank of the channel matrix (assuming there is enough diversity in the wireless medium to make the channel matrix full-rank).

As regards the noise, we assume it is a zero mean multivariate Gaussian process, statistically independent across time and across sub-carriers, with covariance matrix on each sub-carrier $E[\eta_n(k)\eta_n(k)^H] = \text{Cov}(\eta_n)$ (or equivalently precision matrix $\mathbf{B}_{\eta_n} = \text{Cov}(\eta_n)^{-1}$).

Finally, observe that an OFDM system is designed to support a channel of length up to the length of the Cyclic Prefix, in order to suppress Inter-Symbol Interference at the receiver. Therefore, in the course of our treatment, we always assume that the condition $CP \geq L - 1$ is fulfilled. Moreover, in the study of the channel estimators carried on in the following chapters, we always assume that the channel length $L$ is known at the receiver.

### 1.3 Problem Formulation

Now that we have defined the system model and the assumptions used throughout the thesis, before proceeding with the treatment it is important to formulate the problem of channel estimation in MIMO-OFDM systems.

**Problem Statement 1.1** (Channel Estimation in MIMO-OFDM systems). Based on a set of observations corresponding to pilot symbols ($Y^{(tr)}$) and to the unknown symbols ($Y^{(bl)}$), and based on the sequence of transmitted pilots $X^{(tr)}$, the channel estimator attempts to approximate the unknown channel taps \{\(H_n, n = 0 \ldots N - 1\)\}.

In our case, the channel is assumed to be a FIR filter of length $L$, therefore there is a functional dependency of the channel taps in the frequency domain, given by the
Discrete Fourier Transform. We can in fact write
\[ H_n = \sum_{l=0}^{L-1} h_le^{-i2\pi \frac{ln}{N}} = f_n(h) \quad \forall \quad n = 0 \ldots N - 1 \quad \text{(1.11)} \]

where \( f_n(.) \) is a function expressing the dependency of \( H_n \) on the time domain channel \( h \).

This fact has to be taken into account in the estimation process, as we will do in the course of the thesis.

In order to read and understand the following chapters, the reader should be confident with the basics of estimation theory, in particular with Maximum-Likelihood estimation and its properties. The interested reader is suggested to read [2] or [3] for a general introduction to estimation theory.
Chapter 2

Training sequence channel estimation of MIMO-OFDM FIR channels

In this chapter we derive a Maximum Likelihood estimator of MIMO-OFDM FIR channels based solely on the transmission of a pilot sequence and on the observation of the corresponding output. We study the case of Gaussian noise at the receiver, independent across sub-carriers and across time, with covariance matrix $\text{Cov}(\eta_n)$ on each sub-carrier $n$. Then we apply the results to the simpler case of white Gaussian noise at the receiver, with variance $\sigma^2_w$ on each receiving antenna, in order to better understand the limits of applicability of the pilot based approach to MIMO systems.

In this chapter and in the following, where we treat the Semi-Blind approach, we assume to have perfect knowledge of the statistics of the noise at the receiver. However, observe that this assumption does not hold true in practice, therefore the noise covariance matrix needs to be estimated at the receiver as well. This issue is analyzed in chapter 4 in detail.

This chapter is organized as follows: based on the system model and on the assumptions described in the previous chapter in section 1.2 in section 2.1 we derive a pilot based Maximum Likelihood (ML) estimator of MIMO-OFDM FIR channels. We also analyze the properties of such estimator, in terms of its mean and its Mean Square Error and we compare it with the Cramér–Rao lower bound (which is derived in section C.2 of the Appendix). In particular, we analyze the case of white Gaussian noise at the receiver, since this gives a deeper insight on the limits of the pilot approach when applied to a MIMO system.
We also derive the necessary condition for the identifiability of the MIMO-OFDM FIR channel, and we determine the minimal pilot structure which satisfies these identifiability conditions.

### 2.1 Maximum-Likelihood channel estimation of a MIMO-OFDM FIR channel

In this section, we derive a ML estimator of a MIMO-OFDM FIR channel using the pilot based approach. As such, only the observations corresponding to pilot symbols are considered for the estimate \((Y^{(tr)})\), therefore the blind observations \(Y^{(bl)}\) are discarded in this chapter.

Since the channel is FIR of length \(L\), there is a functional dependency of the channel taps in the frequency domain, expressed through the DFT

\[
H_n = \sum_{l=0}^{L-1} h_l e^{-j2\pi ln/N} \quad n = 0 \ldots N - 1
\] (2.1)

Therefore the Maximum-Likelihood solution is determined with respect to the channel taps in the time-domain (collected on the parameter matrix \(h\)), since these represent an unconstrained set of parameters, from which the frequency domain channel is determined through the linear transformation given above.

Since the noise at the receiver is statistically independent across sub-carriers and across time, with covariance matrix \(\text{Cov}(\eta_n)\) (or equivalently precision matrix \(\mathcal{B}_{\eta_n} = \text{Cov}(\eta_n)^{-1}\)), the likelihood of the observations, conditioned on the transmitted pilots and on the time-domain channel \(h\), is given by

\[
p(Y^{(tr)}|h, X^{(tr)}) = \prod_{n=0}^{N-1} \left( \frac{1}{\pi R_{\eta_n}} |\mathcal{B}_{\eta_n}| \right)^{K_n^{(tr)}} \cdot 
\cdot \prod_{n=0}^{N-1} \exp \left\{ -\text{trace} \left[ \mathcal{B}_{\eta_n} (Y_n^{(tr)} - H_n X_n^{(tr)}) (Y_n^{(tr)} - H_n X_n^{(tr)})^H \right] \right\}
\] (2.2)

where \(K_n^{(tr)}\) is the number of pilot symbols transmitted on sub-carrier \(n\).
Chapter 2 Training sequence channel estimation of MIMO-OFDM FIR channels

The maximization of the likelihood function with respect to its arguments is equivalent to the minimization of the negative log-likelihood, given by

\[
-\ln p\left(Y^{(tr)} \middle| h, X^{(tr)}\right) = -\sum_{n=0}^{N-1} K_n^{(tr)} \ln \left(\frac{1}{\pi R} |\mathcal{B}_{\eta_n}| \right) + \\
+ \sum_{n=0}^{N-1} \text{trace} \left[ \mathcal{B}_{\eta_n} \left(Y_n^{(tr)} - H_n X_n^{(tr)}\right) \left(Y_n^{(tr)} - H_n X_n^{(tr)}\right)^H \right]
\] (2.3)

In order to enforce the channel length constraint, the minimization of 2.3 is performed with respect to the time-domain channel matrix \(h\). Keeping only the terms depending on \(h\), the ML estimate, \(\hat{h}\), is solution to the following minimization problem:

\[
\hat{h} = \min_h \left\{ \sum_{n=0}^{N-1} \text{trace} \left[ \mathcal{B}_{\eta_n} \left(Y_n^{(tr)} - H_n X_n^{(tr)}\right) \left(Y_n^{(tr)} - H_n X_n^{(tr)}\right)^H \right] \right\}
\] (2.4)

Since this problem will be encountered often in the course of this thesis, we express the above equation in a more general form, by defining the two matrices \(\Lambda_{xx}^{(n)}\) and \(\Lambda_{yx}^{(n)}\) as

\[
\begin{align*}
\Lambda_{xx}^{(n)} &= X_n^{(tr)} X_n^{(tr)H} \\
\Lambda_{yx}^{(n)} &= Y_n^{(tr)} X_n^{(tr)H}
\end{align*}
\] (2.5)

Then, we can rewrite 2.4 as

\[
\hat{h} = \min_h \left\{ \sum_{n=0}^{N-1} \text{trace} \left[ H_n^H \mathcal{B}_{\eta_n} H_n \Lambda_{xx}^{(n)} \right] - 2\text{real} \sum_{n=0}^{N-1} \text{trace} \left( H_n^H \mathcal{B}_{\eta_n} Y_n^{(tr)} X_n^{(tr)H} \right) \right\}
\] (2.6)

where we have defined the cost function

\[
f(h) = \sum_{n=0}^{N-1} \text{trace} \left( H_n^H \mathcal{B}_{\eta_n} H_n \Lambda_{xx}^{(n)} \right) - 2\text{real} \sum_{n=0}^{N-1} \text{trace} \left( H_n^H \mathcal{B}_{\eta_n} \Lambda_{yx}^{(n)H} \right)
\] (2.7)

The minimization is carried out by computing the derivative of 2.7 with respect to the channel entries \(\{h_l(r,t), l = 0 \ldots L - 1, r = 0 \ldots R - 1, t = 0 \ldots T - 1\}\), and equaling this derivative to zero. The complex derivative (defined in Appendix A) with respect to
entry $h_l(r, t)^*$ of the time-domain channel is given by

$$\frac{\partial f(h)}{\partial h_l(r, t)^*} = \sum_{n=0}^{N-1} \text{trace} \left[ \delta(t, r) B_{\eta_n} H_n \Lambda^{(n)}_{xx} - \delta(t, r) B_{\eta_n} \Lambda^{(n)}_{yx} \right] e^{i2\pi \frac{ln}{N}}$$

$$= \sum_{n=0}^{N-1} \left[ B_{\eta_n} \left( H_n \Lambda^{(n)}_{xx} - \Lambda^{(n)}_{yx} \right) \right]_{rt} e^{i2\pi \frac{ln}{N}} = 0$$ (2.8)

where $\delta(t, r)$ is a matrix whose entries are equal to zero except for the entry at position $(t, r)$ which is equal to 1.

Rewriting the above equation in matrix form we have

$$\sum_{n=0}^{N-1} B_{\eta_n} \Lambda^{(n)}_{yx} e^{i2\pi \frac{ln}{N}} = \sum_{n=0}^{N-1} B_{\eta_n} H_n \Lambda^{(n)}_{xx} e^{i2\pi \frac{ln}{N}}$$ (2.9)

Now, expressing $H_n$ as the Fourier transform of the time-domain channel, we obtain

$$\sum_{n=0}^{N-1} B_{\eta_n} \Lambda^{(n)}_{yx} e^{i2\pi \frac{ln}{N}} = \sum_{p=0}^{L-1} \sum_{n=0}^{N-1} B_{\eta_n} h_p \Lambda^{(n)}_{xx} e^{i2\pi \frac{(l-p)n}{N}}$$ (2.10)

In order to determine a solution to this problem, let’s consider the entry at position $(r, t)$ of the above system of equations, and let’s make explicit the matrix product operation in the following way

$$\left( \sum_{n=0}^{N-1} B_{\eta_n} \Lambda^{(n)}_{yx} e^{i2\pi \frac{ln}{N}} \right)_{rt} = \left( \sum_{p=0}^{L-1} \sum_{n=0}^{N-1} B_{\eta_n} h_p \Lambda^{(n)}_{xx} e^{i2\pi \frac{(l-p)n}{N}} \right)_{rt}$$

$$= \sum_{p=0}^{L-1} \sum_{n=0}^{N-1} \sum_{r_1 t_1} B_{\eta_n}(r, r_1) h_p(r_1, t_1) \Lambda^{(n)}_{xx}(t_1, t) e^{i2\pi \frac{(l-p)n}{N}}$$

$$= \sum_{p=0}^{L-1} \sum_{r_1 t_1} \left( \sum_{n} B_{\eta_n}(r, r_1) \Lambda^{(n)}_{xx}(t_1, t) e^{i2\pi \frac{(l-p)n}{N}} \right) h_p(r_1, t_1)$$ (2.11)

where for convenience we dropped the extrema of the sum over the sub-carrier number $n$.

Now, let $\Gamma^{(fr)}_{xx}$ be an $LRT \times LRT$ matrix with elements

$$\Gamma^{(fr)}_{xx}(RTl + Tr + t; RTp + Tr_1 + t_1) = \sum_{n} B_{\eta_n}(r, r_1) \Lambda^{(n)}_{xx}(t_1, t) e^{i2\pi \frac{(l-p)n}{N}}$$

$$= \left( \sum_{n} B_{\eta_n} \otimes \Lambda^{(n)}_{xx} e^{i2\pi \frac{(l-p)n}{N}} \right)_{Tr+t,Tr_1+t_1}$$ (2.12)
where the notation $A \otimes B$ represents the Kronecker product between $A$ and $B$, that is, assuming $A$ is an $M \times N$ matrix

$$A \otimes B = \begin{bmatrix}
A(0,0) \cdot B & A(0,1) \cdot B & \cdots & A(0,N-1) \cdot B \\
A(1,0) \cdot B & A(1,1) \cdot B & \ddots & \vdots \\
\vdots & \ddots & \ddots & \vdots \\
A(M-1,0) \cdot B & A(M-1,1) \cdot B & \cdots & A(M-1,N-1) \cdot B
\end{bmatrix}$$ (2.13)

Then, let’s redefine $h$ as an $LRT \times 1$ column vector with entries

$$h(RTl + Tr + t) = h_l(r,t)$$ (2.14)

and similarly let $\Gamma^{(tr)}_{yx}$ be a column vector with entries

$$\Gamma^{(tr)}_{yx}(RTl + Tr + t) = \left( \sum_{n=0}^{N-1} B_{yn} \Lambda^{(n)}_{yx} e^{i2\pi ln} \right)_{rt}$$ (2.15)

Then, we can rewrite 2.10 in matrix form as

$$\Gamma^{(tr)}_{yx} = \Gamma^{(tr)}_{xx} h$$ (2.16)

Finally, assuming that $\Gamma^{(tr)}_{xx}$ is full rank, and therefore invertible (we will discuss about the necessary conditions in section 2.1.1), the Maximum Likelihood estimate of the time-domain channel is given by

$$\hat{h}^{(tr)} = \Gamma^{(tr)-1}_{xx} \Gamma^{(tr)}_{yx}$$ (2.17)

Then, letting $\hat{H}^{(tr)}$ be an $NRT$ dimensional column vector with elements

$$\hat{H}^{(tr)}(RTn + Tr + t) = H_n(r,t)$$ (2.18)

and $\tilde{U}_N$ an $N \times L$ matrix obtained by taking the first $L$ columns of the $N \times N$ Fourier matrix $U_N$ with elements $U_N(n,l) = \frac{1}{\sqrt{N}} e^{-i2\pi ln}$, we can write the frequency domain channel estimate as

$$\hat{H}^{(tr)} = \sqrt{N} \left( \tilde{U}_N \otimes I_{RT} \right) \hat{h}^{(tr)}$$ (2.19)

where $I_K$ is the $K \times K$ identity matrix.

Since this estimator will be used often in the course of our treatment of Semi-Blind estimators, it is convenient to include all the operations involved in the estimation of the time-domain channel into a Black-Box, that is a function $\mathcal{H}$, taking as input the
symbol autocorrelation $\Lambda^{(n)}_{XX}$, the correlation between the observations and the transmitted symbols $\Lambda^{(n)}_{YX}$ and the noise precision matrix $B_{\eta n}$ on each sub-carrier (the channel length $L$, the number of receiving and transmitting antennas $R$ and $T$ are dropped for convenience), and returning the time-domain channel estimate. That is

$$\hat{h}^{(tr)} = \mathcal{H} \left( \Lambda^{(n)}_{XX}, \Lambda^{(n)}_{YX}, B_{\eta n}, n = 0 \ldots N - 1 \right)$$

(2.20)

With these inputs we can easily compute 2.12 and 2.15, which can then be used to estimate the channel using 2.17.

We now present a result on the necessary conditions for the identifiability of the channel (said in another way, the channel is identifiable if there are no ambiguities on the determination of the ML estimate).

### 2.1.1 Channel Identifiability Conditions

**Theorem 2.1** (Necessary identifiability condition of the channel for the pilot based approach through ML estimate). *The necessary (but not sufficient) condition for the identifiability of the channel is*

$$\sum_{n=0}^{N-1} \text{rank} \left( X_n^{(tr)} X_n^{(tr)H} \right) \geq LT$$

(2.21)

*Proof.* From equation 2.17 we see that the channel matrix $h$ is identifiable if and only if $\Gamma^{(tr)}_{XX}$ is invertible, or equivalently, if and only if it is full rank.

Observe that $\Gamma^{(tr)}_{XX}$ can be rewritten in the following form:

$$\Gamma^{(tr)}_{XX} = N \left( \hat{U}_N \otimes I_{RT} \right)^H \Lambda \left( \hat{U}_N \otimes I_{RT} \right)$$

(2.22)

where $\Lambda$ is a block diagonal $NRT \times NRT$ matrix obtained by stacking the matrices $B_{\eta n} \otimes \Lambda^{(n)}_{XX}$ along the diagonal.

Then for the rank of $\Gamma^{(tr)}_{XX}$, using the product rule we have

$$\text{rank} \left( \Gamma^{(tr)}_{XX} \right) \leq \min \left\{ \text{rank} \left( \hat{U}_N \otimes I_{RT} \right), \text{rank}(\Lambda) \right\}$$

(2.23)

Now, matrix $\hat{U}_N$ is full rank, since its columns belong to a set of orthonormal vectors (the columns of the Fourier matrix $U_N$ represent a set of orthonormal vectors), therefore we have:

$$\text{rank} \left( \hat{U}_N \otimes I_{RT} \right) = LRT$$

(2.24)
\( \Lambda \) is a block diagonal matrix, therefore its rank is equal to the sum of the ranks of its diagonal blocks:

\[
\text{rank} (\Lambda) = \sum_{n=0}^{N-1} \text{rank} \left\{ B_{\eta_n} \otimes \Lambda^{(n)}_{xx} \right\} \tag{2.25}
\]

and since \( B_{\eta_n} \) is a full-rank square matrix with rank \( R \) we can rewrite, using the fact that \( \text{rank} (A \otimes B) = \text{rank} (A) \text{rank} (B) \):

\[
\text{rank} (\Lambda) = R \sum_{n=0}^{N-1} \text{rank} \left( \Lambda^{(n)}_{xx} \right) \tag{2.26}
\]

Finally, using (2.23) we have:

\[
\text{rank} \left( \Gamma^{(tr)}_{xx} \right) \leq \min \left\{ LT, \sum_{n=0}^{N-1} \text{rank} \left( \Lambda^{(n)}_{xx} \right) \right\} \tag{2.27}
\]

Therefore, substituting \( \Lambda^{(n)}_{xx} = X_n^{(tr)} X_n^{(tr)H} \) the necessary condition for \( \Gamma^{(tr)}_{xx} \) to be full rank is:

\[
\sum_{n=0}^{N-1} \text{rank} \left( X_n^{(tr)} X_n^{(tr)H} \right) \geq LT \tag{2.28}
\]

which completes the proof. \( \square \)

We now present another broader necessary condition for the channel identifiability, determining the minimum number of pilot symbols necessary for the channel to be identifiable.

**Lemma 2.2.** The minimum number of pilots necessary for the channel to be identifiable is

\[
\sum_{n} \min \left\{ K_n^{(tr)}, T \right\} \geq LT \tag{2.29}
\]

**Proof.** In fact from channel identifiability condition (2.21) we have

\[
\sum_{n=0}^{N-1} \text{rank} \left( X_n^{(tr)} X_n^{(tr)H} \right) \geq LT \tag{2.30}
\]
Now, for the pilot correlation terms we have

\[ X_n^{(tr)}X_n^{(tr)H} = \sum_k X_n^{(tr)}(k)X_n^{(tr)H}(k) \]  \hspace{1cm} (2.31)

where \( X_n^{(tr)}(k) \) is the vector of the pilots transmitted at time \( k \), or zero if no pilots are transmitted at time \( k \). Observe that each matrix \( X_n^{(tr)}(k)X_n^{(tr)H}(k) \) has rank one if a pilot is transmitted at time \( k \), or rank zero otherwise, therefore, since on sub-carrier \( n \) \( K_n^{(tr)} \) pilots are transmitted, the rank of the correlation matrices is given by

\[ \text{rank} \left( X_n^{(tr)}X_n^{(tr)H} \right) = \text{rank} \left( \sum_k X_n^{(tr)}(k)X_n^{(tr)H}(k) \right) \leq \min \left\{ K_n^{(tr)}, T \right\} \]  \hspace{1cm} (2.32)

Finally we obtain

\[ \sum_{n=0}^{N-1} \text{rank} \left( X_n^{(tr)}X_n^{(tr)H} \right) \leq \sum_n \min \left\{ K_n^{(tr)}, T \right\} \]  \hspace{1cm} (2.33)

From the above inequality we see that, if \( \sum_n \min \left\{ K_n^{(tr)}, T \right\} \) < \( LT \), then necessarily condition 2.21 is not satisfied. Therefore a necessary condition on the number of pilots is

\[ \sum_n \min \left\{ K_n^{(tr)}, T \right\} \geq LT \]  \hspace{1cm} (2.34)

which proves the lemma.

However observe that, even if condition 2.29 of the lemma is satisfied, the necessary condition 2.21 may still not be satisfied. This is a consequence of the inequality used in 2.32.

Assuming \( K_n^{(tr)} \geq T \) on all the \( N_{tr} \geq N \) sub-carriers carrying pilots, the above lemma reduces to the condition \( N^{(tr)} \geq L \).

### 2.1.2 Properties of ML channel estimator

In this section we study the properties of the Maximum Likelihood channel estimator given by equation 2.19 in terms of its Bias and Variance. However, only for the calculation of the bias, we assume that the noise precision matrix \( B_{\eta_n} \) used for estimating the channel is not necessarily equal to the true noise precision matrix. This result will be used later during the thesis. Therefore, let \( B_{\eta_n} \) be the true noise precision matrix on sub-carrier \( n \), and \( \tilde{B}_{\eta_n} \) the one actually used to estimate the channel.
2.1.2.1 Bias of Maximum Likelihood channel estimator

Calculating the expectation of $2.19$ with respect to the observations we obtain

$$E \left[ \hat{H}^{(tr)} \right] = \sqrt{N} \left( \tilde{U}_N \otimes I_{RT} \right) \Gamma^{(tr)-1} \left[ \Gamma^{(tr)} \right]_{xx}$$  \hspace{1cm} (2.35)

Now, using $\tilde{B}_{\eta_n}$ instead of $B_{\eta_n}$ for the expression of $\Gamma^{(tr)}_{xx}$ and $\Gamma^{(tr)}_{yx}$, for the entries of $E \left[ \Gamma^{(tr)}_{yx} \right]$ from $2.15$ we have

$$E \left[ \Gamma^{(tr)}_{yx} (RTl + Tr + t) \right] = \left( \sum_{n=0}^{N-1} \tilde{B}_{\eta_n} H_n X_n^{(tr)} X_n^{(tr)^H} e^{2\pi i \frac{nl}{N}} \right)_{rt}$$

$$= \sum_{p=0}^{L-1} \sum_{n=0}^{N-1} \tilde{B}_{\eta_n} H_p X_n^{(tr)} X_n^{(tr)^H} e^{2\pi i \frac{(l-p)n}{N}}$$  \hspace{1cm} (2.36)

where in the last equality we expressed the frequency domain channel as the Fourier transform of the time domain channel. Then, making explicit the matrix products we obtain

$$E \left[ \Gamma^{(tr)}_{yx} (RTl + Tr + t) \right] = \sum_{p=0}^{L-1} \sum_{r=0}^{L-1} \sum_{t=0}^{N-1} \tilde{B}_{\eta_n} h_p X_n^{(tr)} X_n^{(tr)^H} e^{2\pi i \frac{(l-p)n}{N}} h_p(r_1,t_1)$$  \hspace{1cm} (2.37)

Recognizing and substituting in the above expression the entries of $\Gamma^{(tr)}_{xx}$ given by $2.12$ we can rewrite the above expression as

$$E \left[ \Gamma^{(tr)}_{yx} \right] = \Gamma^{(tr)}_{xx} h$$  \hspace{1cm} (2.38)

Finally, substituting into the bias of the estimator

$$E \left[ \hat{H}^{(tr)} \right] = \sqrt{N} \left( \tilde{U}_N \otimes I_{RT} \right) \Gamma^{(tr)-1} \Gamma^{(tr)}_{xx} h = \sqrt{N} \left( \tilde{U}_N \otimes I_{RT} \right) h = H$$  \hspace{1cm} (2.39)

which demonstrates that the ML estimator is unbiased.

This result shows also that the ML channel estimator is unbiased even if we don’t use the true noise covariance matrix for the estimate. This result will be used in chapter 4 for the joint estimation of the channel and of the noise covariance matrix on each sub-carrier.
2.1.2.2 Variance of Maximum Likelihood channel estimator

Now, we define the Mean Square Error of the estimator as the sum of the Mean Square Error for the estimation of each entry of the channel matrix, divided by the number of entries. This can be thought as the Mean Square Error for the estimation of the channel matrix entries, averaged over all the entries. For the case under consideration, the Mean Square Error corresponds to the variance of the estimator since the estimator is unbiased, as demonstrated in 2.1.2.1. Therefore for the variance we have

\[
\text{Var} \left( \hat{H}^{(tr)} \right) = \frac{1}{NRT} E \left\{ \text{trace} \left[ \left( \hat{H}^{(tr)} - H \right) \left( \hat{H}^{(tr)} - H \right)^H \right] \right\} 
\]

\[
= \frac{1}{RT} E \left\{ \text{trace} \left[ \left( \hat{H}_N \otimes I_{RT} \right) \left( \hat{H}^{(tr)} - h \right) \left( \hat{H}^{(tr)} - h \right)^H \left( \hat{H}_N \otimes I_{RT} \right)^H \right] \right\} 
\]

\[
= \frac{1}{RT} E \left\{ \text{trace} \left[ \left( \hat{h}^{(tr)} - h \right)^H \left( \hat{h}^{(tr)} - h \right) \right] \right\} 
\]

(2.40)

where in the last equality we used the fact that \(\text{trace}(AB) = \text{trace}(BA)\), and \(\hat{H}_N^H \hat{H}_N = I_L\).

Then, substituting into \(\hat{h}^{(tr)}\) the ML solution given by 2.17 we obtain

\[
\text{Var} \left( \hat{H}^{(tr)} \right) = \frac{1}{RT} \text{trace} \left\{ \Gamma_{xx}^{(tr)-1} E \left[ \left( \Gamma_{yz}^{(tr)} - E \left[ \Gamma_{yz}^{(tr)} \right] \right) \left( \Gamma_{yz}^{(tr)} - E \left[ \Gamma_{yz}^{(tr)} \right] \right)^H \right] \right\} 
\]

(2.41)

where we used the fact that \(h = \Gamma_{xx}^{(tr)-1} E[\Gamma_{yz}^{(tr)}] \) from 2.38 and \(\Gamma_{xx}^{(tr)}\) is Hermitian.

For the entries of the term \(E \left[ \left( \Gamma_{yz}^{(tr)} - E \left[ \Gamma_{yz}^{(tr)} \right] \right) \left( \Gamma_{yz}^{(tr)} - E \left[ \Gamma_{yz}^{(tr)} \right] \right)^H \right] \) we have

\[
E \left\{ \left[ \Gamma_{yz}^{(tr)} - E \left[ \Gamma_{yz}^{(tr)} \right] \right) \left( \Gamma_{yz}^{(tr)} - E \left[ \Gamma_{yz}^{(tr)} \right] \right)^H \right\}_{(l,r_1,t_1;p,r_2,t_2)} 
\]

\[
= \sum_{n=0}^{N-1} E \left\{ \mathcal{B}_{\eta_n} \left( Y_n^{(tr)} - H_n X_n^{(tr)} \right) X_n^{(tr)H} \right\}_{r_1 t_1} \cdot \left[ \mathcal{B}_{\eta_n} \left( Y_n^{(tr)} - H_n X_n^{(tr)} \right) X_n^{(tr)H} \right]^{*}_{r_2 t_2} e^{j2\pi \left( \frac{l-p}{N} \right)} 
\]

(2.42)
Now, making explicit the matrix products we obtain

\[ E \left\{ \left[ \left( \Gamma_{tr}^{(tr)} - E \left[ \Gamma_{tr}^{(tr)} \right] \right) \left( \Gamma_{tr}^{(tr)} - E \left[ \Gamma_{tr}^{(tr)} \right] \right)^H \right] \right\}_{(l,r_1,t_1;p,r_2,t_2)} = 2.43 \]

\[ = \sum_{n=0}^{N-1} \sum_{s_1,s_2} \sum_{k} \sum_{l} E \left[ \left( Y_n^{(tr)} - H_n X_n^{(tr)} \right) s_1 k \left( Y_n^{(tr)} - H_n X_n^{(tr)} \right)^*_s \right] . \]

\[ B_{\eta_n} (r_1, s_1) X_n^{(tr)*} (t_1, k) B_{\eta_n} (s_2, r_2) X_n^{(tr)} (t_2, k) e^{j2\pi \frac{(l-p) n}{N}} \]

\[ = \sum_{n=0}^{N-1} \sum_{s_1,s_2} B_{\eta_n} (r_1, s_1) \left( \text{Cov} (\eta_n) \right)_{s_1 s_2} B_{\eta_n} (s_2, r_2) \sum_{k} X_n^{(tr)*} (t_1, k) X_n^{(tr)} (t_2, k) e^{j2\pi \frac{(l-p) n}{N}} \]

where in the last equality we used the definition of noise covariance matrix \( \text{Cov} (\eta_n) \).

Finally, observing that

\[ \sum_{s_1 \neq s_2} B_{\eta_n} (r_1, s_1) \left( \text{Cov} (\eta_n) \right)_{s_1 s_2} B_{\eta_n} (s_2, r_2) = (B_{\eta_n} \text{Cov} (\eta_n) B_{\eta_n})_{r_1 r_2} = B_{\eta_n} (r_1, r_2) \]

we obtain

\[ E \left\{ \left[ \left( \Gamma_{tr}^{(tr)} - E \left[ \Gamma_{tr}^{(tr)} \right] \right) \left( \Gamma_{tr}^{(tr)} - E \left[ \Gamma_{tr}^{(tr)} \right] \right)^H \right] \right\}_{(l,r_1,t_1;p,r_2,t_2)} = 2.45 \]

\[ = \sum_{n=0}^{N-1} B_{\eta_n} (r_1, r_2) X_n^{(tr)*} X_n^{(tr)} e^{j2\pi \frac{(l-p) n}{N}} \]

and comparing the above expression with the entries of \( \Gamma_{tr}^{(tr)} \) in 2.12 we can rewrite

\[ E \left\{ \left[ \Gamma_{tr}^{(tr)} - E \left[ \Gamma_{tr}^{(tr)} \right] \right) \left( \Gamma_{tr}^{(tr)} - E \left[ \Gamma_{tr}^{(tr)} \right] \right)^H \right\} = \Gamma_{tr}^{(tr)} \]

Finally, substituting this expression into the expression for the variance of the estimator in 2.41 we obtain the following result:

\[ \text{Var} \left( \hat{H}^{(tr)} \right) = \frac{1}{RT} \text{trace} \left( \Gamma_{tr}^{(tr)} \right) \]

which represents the variance (MSE) of the Maximum Likelihood estimator.

In conclusion, the Maximum Likelihood estimator derived in the previous section is unbiased with variance \( \text{Var} \left( \hat{H}^{(tr)} \right) = \frac{1}{RT} \text{trace} \left( \Gamma_{tr}^{(tr)} \right) \).

In the Appendix, in section C.2 we derive the Cramér–Rao lower bound for the pilot based approach, showing that the ML estimator achieves the CRLB for any configuration of the pilot grid.
2.1.3 White Gaussian Noise at the receiver

In the previous section we derived the expression for the ML FIR channel estimator for Gaussian noise at the receiver with covariance matrix \( \text{Cov}(\eta_n) \) on each sub-carrier \( n \), and we derived its properties in terms of bias and variance.

In order to better understand the estimation accuracy achievable with the pilot based approach, it is interesting to study the particular case of white Gaussian noise at the receiver, with variance \( \sigma^2_w \) on all sub-carriers and on all receiving antennas.

In this case the covariance matrix is given by \( \text{Cov}(\eta_n) = \sigma^2_w I_R \) \( \forall n \), or equivalently the precision matrix is given by \( B_{\eta_n} = \frac{1}{\sigma^2_w} I_R \) \( \forall n \).

Moreover, we also assume a typical scenario where the pilots are allocated on sub-carrier \( n_0 < S_{tr} \) and then on the following sub-carriers spaced by \( S_{tr} \), where \( S_{tr} \) is the pilot sub-carriers spacing, a divisor of \( N \). We also assume that on all these sub-carriers and on all the transmitting antennas the total power \( \rho \) assigned to the pilots is the same, and that the pilot sequence is orthogonal across the transmitting antennas array. This can be mathematically written as

\[
\begin{align*}
X^{(tr)}_n X^{(tr)H}_n &= \rho I_T \quad n = n_0 + k S_{tr}, \forall k = 0 \ldots \frac{N}{S_{tr}} - 1 \\
X^{(tr)}_n X^{(tr)H}_n &= 0 \quad \text{otherwise}
\end{align*}
\]  

(2.48)

Since only \( \frac{N}{S_{tr}} \) sub-carriers over \( N \) are used for the allocation of pilots, and the rank of \( X^{(tr)}_n X^{(tr)H}_n \) is either \( 0 \) (no pilots allocated on sub-carrier \( n \)) or \( T \) (sub-carrier \( n \) is used for allocating pilots), the necessary identifiability condition becomes:

\[
\sum_{n=0}^{N-1} \text{rank} \left( X^{(tr)}_n X^{(tr)H}_n \right) = \frac{NT}{S_{tr}} \geq LT
\]

(2.49)

or equivalently \( \frac{N}{S_{tr}} \geq L \), which is the same result obtained in lemma 2.2 assuming that \( K^{(tr)}_n \geq T \) on the sub-carriers carrying pilots. In order to enforce the orthogonality of the pilot sequence across the transmitting antennas, one solution is to transmit a set of orthogonal vectors of symbol. For example, on the sub-carriers carrying pilots we may transmit \( T \) pilots in \( T \) distinct MIMO-OFDM symbols, where one only antenna transmits at a time with a power equal to \( \rho \), while the others are silent, and each antenna...
transmits a pilot on one of the $T$ time-slots. This can be written mathematically as

$$X^{(tr)}_n = \begin{pmatrix} \sqrt{\rho e^{j\theta_0}} & 0 & 0 & 0 \\ 0 & \sqrt{\rho e^{j\theta_1}} & 0 & 0 \\ 0 & 0 & \sqrt{\rho e^{j\theta_2}} & 0 \\ 0 & 0 & 0 & \sqrt{\rho e^{j\theta_3}} \end{pmatrix}$$ (2.50)

This is the solution used also to allocate the pilots in the LTE slots and in the course of our simulations.

Substituting (2.48) into the expression for $\Gamma^{(tr)}_{xx}$, we obtain

$$\Gamma^{(tr)}_{xx} = \frac{N\rho}{\sigma_{tr}\sigma_w} I_{LRT}$$ (2.51)

Notice that in this case the condition $\frac{N}{N_{tr}} \geq L$ represents not only a necessary but also a sufficient condition for the identifiability of the channel.

It can be shown that this pilot allocation method is optimal in the case of white Gaussian noise, since it minimizes the variance of the estimator. In fact, let’s assume we have a pilot power constraint, that is

$$\sum_n \text{trace} \left( X^{(tr)}_n X^{(tr)H}_n \right) = P$$ (2.52)

This translates into a constraint on the trace of matrix $\Gamma^{(tr)}_{xx}$, in fact

$$\text{trace} \left( \Gamma^{(tr)}_{xx} \right) = \frac{1}{\sigma_w^2} L R \sum_n \text{trace} \left( X^{(tr)}_n X^{(tr)H}_n \right) = \frac{1}{\sigma_w^2} L R P = \sum_{p=0}^{LRT-1} \lambda_p$$ (2.53)

where in the last equality we used the fact that the trace of $\Gamma_{xx}$ is equivalent to the sum of its eigenvalues $\{\lambda_p\}$.

The optimization of the pilot structure is performed by minimizing the variance of the estimator, given by (2.47) under the constraint (2.53) using the Lagrange multipliers in order to enforce the constraint we have the following cost function:

$$f = \frac{1}{RT} \text{trace} \left( \Gamma^{(tr) - 1}_{xx} \right) + \mu \left( \sum_{p=0}^{LRT-1} \lambda_p - \frac{1}{\sigma_w^2} L R P \right) = \frac{1}{RT} \sum_{p=0}^{LRT-1} \frac{1}{\lambda_p} + \mu \left( \sum_{p=0}^{LRT-1} \lambda_p - \frac{1}{\sigma_w^2} L R P \right) =$$ (2.54)
Then, calculating the derivative of the cost function with respect to the eigenvalue $\lambda_r$ and equaling it to zero we have

$$\lambda_r = \frac{1}{\sqrt{RT\mu}}$$  (2.55)

Finally, enforcing the power constraint we obtain

$$\lambda_r = \frac{\mathcal{P}}{\sigma_w^2 T} \quad \forall r$$  (2.56)

which demonstrates that the optimal $\Gamma_{xx}$ minimizing the variance of the estimator is given by

$$\Gamma^{(tr)}_{xx} = \frac{\mathcal{P}}{\sigma_w^2 T} I_{LRT}$$  (2.57)

This is achieved with the pilot allocation method 2.48 by setting $\mathcal{P} = \frac{N_T\rho}{\Sigma T_T}$.

For the variance of the estimator, substituting $\Gamma^{(tr)}_{xx}$ into 2.47 we have

$$\text{Var}\left(\hat{H}_{tr}\right) = \frac{LS_T\sigma^2_w}{N\rho}$$  (2.58)

Notice that in a common system the average transmission power per sub-carrier $\sigma^2_{Tx}$ is the same on all sub-carriers, and is equally distributed across the transmitting antennas. Then, using this assumption, since $\rho$ is the total power assigned to the pilots on each transmitting antenna and on each sub-carrier, $\rho\frac{N_T}{\Sigma T_T} = \sigma^2_{Tx} N_{TOT}$, where $N_{TOT}$ is the total number of pilot symbols allocated on the OFDM grid. Therefore we can rewrite the variance as

$$\text{Var}\left(\hat{H}_{tr}\right) = \frac{\sigma^2_w LT}{\sigma^2_{Tx} N_{TOT}} = \frac{LT}{\text{SNR} \cdot N_{TOT}}$$  (2.59)

where $\text{SNR} = \frac{\sigma^2_{Tx}}{\sigma^2_w}$ is the signal to noise ratio of the system.

The above expression highlights some important limitations of the pilot based approach. Observe that the variance of the estimator grows proportionally to the number of transmitting antennas, and inversely to the number of pilots $N_{TOT}$. This implies that a larger number of transmitting antennas has to be compensated by a longer pilot sequence, in order to achieve a given estimation accuracy while keeping fixed the other parameters. This behavior can be easily understood by inspecting the pilot allocation structure 2.50 which we showed to be optimal since it minimizes the variance of the estimator: only one antenna transmits at a time, since in this way the interference from the other antennas is suppressed, and each receiving antenna is able to effectively estimate the SISO channel between itself and the antenna transmitting the pilot symbol. Therefore $T$ pilot symbols...
are needed for all the antennas to transmit one pilot, in other words the number of pilot symbols necessary to achieve a given estimation accuracy is proportional to the number of transmitting antennas.

It is clear that, as the order of the MIMO system increases, while keeping fixed the other parameters, in order to achieve an acceptable estimation accuracy more pilots have to be collected at the receiver. This in turn is achieved either enlarging the observation time, or allocating more pilots on the OFDM grid. However, the first approach (larger observation time) compromises the ability of the receiver to track fast-varying channels, which is not acceptable in a Mobile Communication System. On the other hand, the second approach (more pilots on the OFDM grid) compromises the bandwidth efficiency of the system, since the pilots represent a waste of bandwidth. Therefore, it becomes important to exploit also other information at the receiver than relying solely on pilots. The approach studied in this thesis for improving the estimation accuracy consists in exploiting also the unknown symbols at the receiver (semi-blind channel estimation). In the next chapter we study different Semi-Blind approaches and algorithms, and we compare them with the pilot based approach studied in this chapter.
Chapter 3

Semi-Blind channel estimation

In chapter 2 we have derived a Maximum Likelihood estimator of a MIMO-OFDM FIR channel based exclusively on pilot symbols, assuming Gaussian noise at the receiver with covariance matrix $\text{Cov}({\eta}_n)$ on sub-carrier $n$. We have also shown that the estimation accuracy of such an estimator equals the corresponding Cramér–Rao lower bound and, in the case of white Gaussian noise at the receiver, and orthogonal pilots, equally spaced across the sub-carriers, the variance of the estimator is given by equation 2.59 and reported here

$$\text{Var}(\hat{H}_{tr}) = \frac{\sigma_w^2LT}{\sigma_T^2N_{TOT}} = \frac{LT}{\text{SNR} \cdot N_{TOT}} \quad (3.1)$$

where $N_{TOT}$ is the total number of pilots used for the estimate. From this result it is clear that, in order to improve the estimation accuracy, for a given signal to noise ratio and number of transmitting antennas, a larger number of pilot observations have to be collected at the receiver.

In a MIMO-OFDM system it is required to estimate a larger number of parameters with respect to a simple SISO system. This negatively impacts the accuracy of the estimator. In fact, observing the expression given above, we see that the variance for the estimation of the entries of the channel matrix increases linearly with the number of transmitting antennas. Moreover, notice that the one given above represents the average variance per entry of the channel matrix. If we sum the variance for the estimation of each channel entry, instead of averaging it over the number of entries, the overall variance is a quadratic function of the number of transmitting antennas and a linear function of the number of receiving antennas. This dependency on the dimension of the MIMO-OFDM system, in the case of estimation techniques based on pilots alone, translates into the need for a longer training sequence with respect to a simple SISO in order to achieve a given performance. This is achieved either by enlarging the observation time,
or by allocating more pilots on the OFDM grid. However, as explained in the previous chapter, a longer observation time is not desirable since it compromises the possibility to track fast-varying channels. On the other hand, allocating more pilots on the OFDM grid is not good since this comes to the disadvantage of the bandwidth efficiency of the system.

From these considerations, it is clear that it is important to develop a new class of estimators, which does not only exploit the known symbols for the estimate but also blind information in order to enhance the estimation accuracy, without the need for a longer observation time and with the minimum utilization of bandwidth for the allocation of pilots. This class of estimators is known as Semi-Blind estimators, and allow for the estimation of the channel parameters using all the available information at the receiver, with the potential for improving the estimation accuracy.

In this section we develop a Semi-Blind Maximum Likelihood estimator of a MIMO-OFDM FIR channel. The chapter is organized as follows: in section 3.1 we introduce ML semi-blind channel estimation of MIMO-OFDM FIR channels from a general perspective, that is we don’t make any prior assumption on the distribution of the transmitted signal, and we propose the Expectation-Maximization algorithm as a general framework to solve the maximization problem. Then in section 3.2 we apply the results derived in section 3.1 to the case when the true discrete distribution of the unknown symbols is exploited at the receiver for the estimate. However, this leads to an high computational overhead, which can be reduced with the use of approximations. Therefore, in section 3.3 we use the Gaussian assumption, that is we assume the unknown symbols are circular Gaussian distributed. Finally, in section 3.4 we use the Constant Modulus approximation for the unknown symbols, that is we assume they have constant amplitude and phase uniformly distributed in $[0, 2\pi)$. However, for this last case, we will show that its applicability is limited to Constant Modulus constellations (like 4-QAM or QPSK), and rank-one transmission.

### 3.1 General formulation of Semi-Blind ML estimation of MIMO-OFDM FIR channels

In this section we derive a general treatment of ML estimation of MIMO-OFDM FIR channels. This generality derives from the fact that we don’t make any prior assumption on the distribution of the transmitted symbols. Therefore the results we derive here can be applied to any particular case, either to training sequence, blind or semi-blind estimation techniques.
Let’s consider a $T \times R$ MIMO-OFDM system ($T$ and $R$ are the numbers of transmitting and receiving antennas, respectively), with $N$ sub-carriers. Let’s assume $K$ OFDM symbols are transmitted. The input-output relation of this system is given by

$$Y_n = H_n X_n + \eta_n \quad \forall \quad n = 0 \ldots N - 1$$

(3.2)

where $Y_n$ is the $R \times K$ observation matrix, $H_n$ is the channel matrix, $X_n$ is the $T \times K$ matrix of the transmitted symbols, and $\eta_n$ is the noise matrix at the receiver on sub-carrier $n$. We don’t make any assumption on the distribution of the transmitted symbols. Therefore $X_n$ may carry either pilots, unknown symbols, or both.

The Maximum Likelihood solution is obtained by maximizing the likelihood, or equivalently minimizing the negative log-likelihood of the observations with respect to the parameters of the model. As we did in the previous chapter for the pilot based approach, since the channel is FIR of length $L$, in order to enforce the functional constraint of the frequency domain channel taps, the ML solution is determined with respect to the channel coefficients in the time domain, that is $\{h_l(r,t), \forall \quad l,r,t\}$. Then ,stacking the time domain channel entries on the column vector $h$, with entries $h(RTl + Tr + t) = h_l(r,t)$, the likelihood of the observations conditioned on $h$ is given by:

$$p(Y|h) = E_X[p(Y|X,h)]$$

(3.3)

where the notation $E_\alpha[f(\alpha,\beta)]$, represents the expectation of $f(\alpha,\beta)$ with respect to the prior distribution of $\alpha$, whereas the notation $E_\alpha[f(\alpha,\beta)|\beta]$ represents the expectation of $f(\alpha,\beta)$ with respect to the distribution of $\alpha$ conditioned on $\beta$ (this expectation is a function of $\beta$).

Under regularity conditions (differentiability of the likelihood function with respect to its argument $h$), a necessary condition for the ML solution is that it is solution to the likelihood equation, which is obtained by calculating the gradient of $-\ln p(Y|h)$ with respect to the parameter vector $h$ (the gradient operator is indicated with the notation $\Delta_h$), and equaling it to zero. We obtain

$$-\Delta_h \ln p(Y|h) = -\frac{1}{p(Y|h)} \Delta_h p(Y|h) = -\frac{1}{p(Y|h)} \Delta_h E_X[p(Y|X,h)]$$

(3.4)

where we used the fact that $\frac{\partial \ln f(\alpha)}{\partial \alpha} = \frac{1}{f(\alpha)} \frac{\partial f(\alpha)}{\partial \alpha}$. 


Then, since the prior distribution of the transmitted symbols does not depend on the channel entries, we can move the gradient within the expectation term, obtaining

$$\Delta h \ln p(Y|h) = -\frac{1}{p(Y|h)}E_X [\Delta_h p(Y|X,h)] = \frac{1}{p(Y|h)}E_X [p(Y|X,h)\Delta_h \ln p(Y|X,h)]$$  \hspace{1cm} (3.5)

Finally, using the fact that $E_X \left[ \frac{p(Y|X,h)}{p(Y|h)} f(X) \right] = E_X [f(X)|Y,h]$, and equaling the gradient to zero, the likelihood equation can be written as

$$\Delta h \ln p(Y|h) = -E_X [\Delta_h \ln p(Y|X,h)|Y,h] = 0$$  \hspace{1cm} (3.6)

Now, let’s assume that the noise at the receiver is a zero mean Gaussian process, independent across the sub-carriers and across time, with covariance matrix $\text{Cov}(\eta_n)$ (or equivalently precision matrix $B_{\eta_n} = \text{Cov}(\eta_n)^{-1}$) on sub-carrier $n$. Under this assumption, when conditioned on the transmitted symbols and on the channel, the observations are independent across sub-carriers and across time, therefore we can split the probability density function (PDF) $p(Y|X,H)$ into the product of the PDFs of the observations on each sub-carrier, and equivalently we can express $\ln p(Y|X,H)$ as the sum of those densities. Then, making explicit the probability density function on each sub-carrier we obtain

$$-\ln p(Y|X,h) = -\sum_{n=0}^{N-1} K_n \ln \left( \frac{|B_{\eta_n}|}{\pi R} \right) + \sum_{n=0}^{N-1} \text{trace} \left[ B_{\eta_n} (Y_n - H_n X_n)(Y_n - H_n X_n)^H \right]$$  \hspace{1cm} (3.7)

where $K_n$ is the number of observations used for the estimate on sub-carrier $n$, and $H_n$ is the frequency domain channel tap on sub-carrier $n$, whose entries are linear functions of the parameter vector $h$ through the DFT transform.

The derivative of this term with respect to the time-domain channel matrix entry $h_l(r,t)$ is given by

$$-\frac{\partial \ln p(Y|X,h)}{\partial h_l(r,t)} = \sum_{n=0}^{N-1} \text{trace} \left[ B_{\eta_n} \delta(r,t) X_n (Y_n - H_n X_n)^H \right] e^{-i2\pi \frac{ln}{N}}$$

$$= \sum_{n=0}^{N-1} \left[ X_n (Y_n - H_n X_n)^H B_{\eta_n} \right]_{lr} e^{-i2\pi \frac{ln}{N}}$$  \hspace{1cm} (3.8)
Finally, equaling the derivative to zero, we obtain the entries of the likelihood equation 3.6

$$-\frac{\partial \ln p(Y|h)}{\partial h_l(r,t)} = - \sum_{n=0}^{N-1} E_{X_n} \left[ X_n (Y_n - H_n X_n)^H B_{\eta_n} \left| Y, h \right\rangle \right]_{tr} e^{-i2\pi \frac{\theta}{N}} = 0 \quad (3.9)$$

Since the above equation has to be satisfied for all the transmitting-receiving antennas pairs \((r,t)\) and for all channel taps \(l\), we can rewrite it in matrix form with respect to the indexes \(t\) and \(r\), obtaining the following set of equations

$$-\frac{\partial \ln p(Y|h)}{\partial h^*_l} = - \sum_{n=0}^{N-1} B_{\eta_n} E_{X_n} \left[ Y_n X_n^H - H_n X_n X_n^H \left| Y, h \right\rangle \right] e^{i2\pi \frac{\theta}{N}} = 0 \quad \forall \ l = 0 \ldots L - 1 \quad (3.10)$$

The ML estimate of the channel is solution to equation 3.10. However, observe that a solution to the above equation is not necessarily the ML solution. In fact, all the solutions to equation 3.10 are stationary points of the negative log-likelihood function, however they are not guaranteed to be absolute minima of the function. Furthermore, observe that the solution depends on the posterior expectation and the posterior correlation of the transmitted symbols after observing \(Y\). However, except for the case where the symbols are known at the receiver (pilot based estimation approach), these terms are a function of the channel, therefore in general there is no closed form solution to the above equation. Notice also that this equation is very general, since we didn’t use any assumption on the prior symbols, we have only used the fact the noise at the receiver is Gaussian, independent across sub-carriers and across time. Therefore any particular case can be deduced from it.

It is interesting to observe that, in the case of training sequence estimation, \(X_n\) is the matrix containing solely the pilot symbols, which is a deterministic quantity independent of the channel realization and of the observations, therefore for this case the above equation reduces to:

$$\sum_{n=0}^{N-1} B_{\eta_n} H_n X_n X_n^H e^{i2\pi \frac{\ln}{N}} = \sum_{n=0}^{N-1} B_{\eta_n} Y_n X_n^H e^{i2\pi \frac{\ln}{N}} \quad (3.11)$$

which is the same equation we obtained in chapter 2 equation 2.9 for which a closed form solution exists and is given by 2.17 for the time-domain matrix.

With reference to the system model and the set of assumptions described in section 1.2 when both pilot symbols and blind information are used for the estimation, we can split equation 3.10 into the sum of the contribution coming from the pilot symbols and the contribution from the blind information, that is, using the superscripts \((tr)\) and \((bl)\) to
Chapter 3 Semi-Blind channel estimation

distinguish pilot from blind observations, symbols and noise, we can rewrite equation 3.10 as:

\[-\frac{\partial \ln p(Y|h)}{\partial h_l^*} = - \sum_{n=0}^{N-1} B_{\eta_n} \left( Y_n^{(tr)} - H_n X_n^{(tr)} \right) X_n^{(tr)H} e^{i2\pi \frac{ln}{N}}
- \sum_{n=0}^{N-1} B_{\eta_n} E_{V_n^{(bl)}} \left[ \left( Y_n^{(bl)} - H_n CV_n^{(bl)} \right) V_n^{(bl)H} \right] Y_n^{(bl), h} C^H e^{i2\pi \frac{ln}{N}}
= 0 \ \forall l = 0 \ldots L - 1 \]  

(3.12)

where we have used the fact that, according to the set assumptions defined in 1.2.3, the unknown symbols and the noise are independent across the sub-carriers, so that the unknown symbols on sub-carrier \( n \) are independent from the observations on the other sub-carriers.

Since this equation involves the calculation of the posterior expectation of the transmitted symbols and their correlation conditioned on the observations \( Y \), the solution to this equation then depends on the assumptions we use on the prior distribution of the unknown symbols. From the point of view of the estimation accuracy, the optimal solution consists in using the true discrete distribution of the symbols. However this solution is computationally very demanding, since it requires the computation of the posterior probabilities for any possible combination of transmitted symbols. Moreover, it is not scalable to MIMO systems since the number of symbol combinations grows exponentially with the transmission rank. As an example, while in a SIMO system, using 16-QAM as modulation format, we have to calculate 16 posterior probabilities associated to each transmitted symbol, in a two transmitting antennas system with rank two transmission the number of combinations grows to \( 16^2 = 256 \). Therefore, in order to reduce the computational overhead, we need to relax the true discrete distribution of the unknown symbols, and use some approximations. In the course of this thesis, we will consider in particular the Gaussian approximation for the unknown symbols, and the Constant Modulus approximation, which are treated in section 3.3 and 3.4 respectively.

Although computationally complex, the true discrete distribution is considered in section 3.2. This case can be used as a lower bound on the performance of the Semi-Blind estimators studied in sections 3.3 and 3.4, and is useful to understand the performance loss incurred when using approximations on the distribution of the unknown data.

Before considering the true discrete distribution of the unknown symbols, we describe the EM-algorithm as a general framework which can be used to determine the ML solution. As we did so far, we keep a level of generality on the distribution of the unknown symbols, so that any particular case can be deduced in a unified way. For a
general treatment of the EM-algorithm, we refer the interested reader to [4], [5] and [6]. However, we briefly introduce it, explaining the steps involved in the algorithm, before proceeding with the treatment.

### 3.1.1 Brief introduction to the EM-Algorithm

Let $p(Y|\theta)$ be the likelihood of the observations, conditioned on the parameter vector $\theta$, and let $X$ be a set of hidden variables, with prior distribution $p(X|\theta)$. These variables, as the name suggests, are not directly observed, but their knowledge provide further information about the observations. Then, it is straightforward to demonstrate the following equality:

$$\ln p(Y|\theta) = E_X[q] \left[ \ln \left( \frac{p(Y,X|\theta)}{q(X)} \right) \right] + E_X[q] \left[ \ln \left( \frac{q(X)}{p(X|Y,\theta)} \right) \right]$$  \hspace{1cm} (3.13)

where $q(X)$ is any distribution on the hidden variables, and the notation $E_X[q]$ is used to specify that the expectation is calculated with respect to this distribution.

Then, recognizing in the above equation the expression for the Kullback–Leibler divergence between the distribution $q(X)$ and the posterior distribution on the hidden variables $p(X|Y,\theta)$, which is a non negative quantity, we have the following lower bound to the likelihood function:

$$\ln p(Y|\theta) = E_X[q] \left[ \ln \left( \frac{p(Y,X|\theta)}{q(X)} \right) \right] + KL(q \parallel p) \geq E_X[q] \left[ \ln \left( \frac{p(Y,X|\theta)}{q(X)} \right) \right] = \mathcal{F}(q, \theta)$$  \hspace{1cm} (3.14)

The EM algorithm, instead of directly maximizing the log-likelihood function $\ln p(Y|\theta)$, maximizes the lower bound to the likelihood function, $\mathcal{F}(q, \theta)$, with respect to its arguments, the distribution $q(X)$ and the parameters $\theta$. In particular, starting from an initial guess $\theta^{(0)}$, the algorithm proceeds by iterating two steps: an expectation step ($E$-step) during which the lower bound is maximized with respect to the distribution $q(X)$ on the latent variables, given the current estimate of the parameters, and a maximization step ($M$-step), during which the lower bound is maximized with respect to the parameter vector, providing a new estimate of $\theta$. As a consequence of the maximization at each step and through multiple iterations, the lower bound increases at each step of the algorithm, converging to a local maximum of the likelihood function.

During the E-step, the distribution maximizing the lower bound is the posterior distribution of the hidden variables given the current estimate of the parameter vector $\theta^{(j)}$, and
that is \( q(X) = p(X|Y, \theta^{(j)}) \), since this solution is such that the Kullback–Leibler divergence term is equal to zero, therefore the lower bound equals the log-likelihood function. During the M-step, since the term \( q(X) \) is independent from \( \theta \), the new estimate of the parameter vector, \( \theta^{(j+1)} \) is given by

\[
\theta^{(j+1)} = \max_{\theta} \left\{ E_{X} \left[ \ln p(Y, X|\theta) \right] \right\}
\]

and substituting the expression for the current update of the distribution \( q(X) = p(X|Y, \theta^{(j)}) \) we obtain

\[
\theta^{(j+1)} = \max_{\theta} \left\{ E_{X} \left[ \ln p(Y, X|\theta, \theta^{(j)}) \right] \right\}
\]

Therefore, the M-step consists in the maximization of the expectation of the likelihood of the complete data (observations and hidden variables) in the log-domain. In many problems, this can be performed much more easily than the direct maximization of the likelihood function.

### 3.1.2 ML solution through EM-algorithm

This algorithm can be used in the Semi-Blind estimation approach. In fact, the unknown symbols can be considered are latent variables, since they are unobserved and their knowledge affects the distribution of the observations. Moreover, the log-likelihood of the observations, conditioned on the transmitted symbols, is a quadratic function of the channel parameters, therefore the update of the channel estimate during the M-step can be performed in closed form.

In order to keep the highest level of generality, let’s assume that the unknown symbols \( V^{(bl)} \) are mapped by means of an encoding function \( G \) into the transmitted symbols \( X \), that is

\[
X = G \left( V^{(bl)} \right)
\]

Therefore, considering the unknown symbols \( V^{(bl)} \) as hidden variables, we have the following lower bound to the log-likelihood of the observations conditioned on the channel:

\[
\ln p(Y|h) \geq E_{V^{(bl)}}^{(q)} \left[ \ln \left( \frac{p(Y, V^{(bl)}|h)}{q(V^{(bl)})} \right) \right] = F \left( q(V^{(bl)}), h \right)
\]

for any distribution on the hidden variables \( q(V^{(bl)}) \).
Now, using (3.16) the \((j+1)\)th update of the channel estimate during the \(M\)-step, given the current estimate \(h^{(j)}\) at the \(j\)th iteration of the EM-algorithm, is given by

\[
h^{(j+1)} = \max_h \left\{ E_{V^{(bl)}} \left[ \ln p \left( Y, V^{(bl)} \right| h \right] \right\}
\]

(3.19)

and using the fact that \(p \left( Y, V^{(bl)} \right| h \right) = p \left( Y | X = G \left( V^{(bl)} \right), h \right) p \left( V^{(bl)} \right)\), and that the prior distribution of the unknown symbols is independent from the channel entries, we can rewrite

\[
h^{(j+1)} = \max_h \left\{ E_{V^{(bl)}} \left[ \ln p \left( Y | X, h \right) \right| Y, h^{(j)} \right\}
\]

(3.20)

Since the noise is independent across the sub-carriers, the log-likelihood of the observations conditioned on the transmitted symbols and on the channel can be split into the sum of the log-likelihood terms on each sub-carrier. Therefore, making explicit the log-likelihood terms and discarding those independent of the channel entries we can write

\[
h^{(j+1)} = \min_h \left\{ \sum_n \text{trace} \left( H_n^H B_{\eta_n} H_n E_{V^{(bl)}} \left[ X_n X_n^H \right| Y, h^{(j)} \right] \right\} - 2 \text{real} \left\{ \sum_n \text{trace} \left( H_n^H B_{\eta_n} Y_n E_{V^{(bl)}} \left[ X_n^H \right| Y, h^{(j)} \right] \right\}
\]

(3.21)

By rewriting \(A_{xx}^{(n,j)} = E_{V^{(bl)}} \left[ X_n X_n^H \right| Y, h^{(j)} \right]\) and \(A_{yx}^{(n,j)} = Y_n E_{V^{(bl)}} \left[ X_n^H \right| Y, h^{(j)} \right]\) we have

\[
h^{(j+1)} = \min_h \left\{ \sum_n \text{trace} \left( H_n^H B_{\eta_n} H_n A_{xx}^{(n,j)} \right) - 2 \text{real} \sum_n \text{trace} \left( H_n^H B_{\eta_n} A_{yx}^{(n,j)} \right) \right\}
\]

(3.22)

This minimization problem was studied in chapter 2, and is equivalent to equation 2.6. Its solution is given by equation 2.20. Then we can write

\[
h^{(j+1)} = \mathcal{H} \left( A_{xx}^{(n,j)}, A_{yx}^{(n,j)}, B_{\eta_n}, n = 0 \ldots N - 1 \right)
\]

(3.23)

From the above equation, it is clear that only the terms \(A_{xx}^{(n,j)}\) and \(A_{yx}^{(n,j)}\) have to be calculated during the E-step. After that, the M-step is identical, independently from the distribution of the unknown symbols \(V^{(bl)}\).

Now, considering the system model and the set of assumptions described in 1.2, the encoding function is simply given by \(X_n^{(bl)} = CV_n^{(bl)}\). Moreover, since the noise and the unknown symbols are independent across sub-carriers and across time, the symbols on sub-carrier \(n\) depend solely on the blind observations on the same sub-carrier, therefore
the terms $\Lambda^{(n,j)}_{xx}$ and $\Lambda^{(n,j)}_{yx}$ can be rewritten as

$$
\Lambda^{(n,j)}_{xx} = E_{V^{(bl)}_n} \left[ X_n X_n^H | Y^{(bl)}_n, h^{(j)} \right] = X^{(tr)}_n X^{(tr)H}_n + C \Lambda^{(n,j)}_{vv} C^H \tag{3.24}
$$

$$
\Lambda^{(n,j)}_{yx} = E_{V^{(bl)}_n} \left[ Y_n X_n^H | Y^{(bl)}_n, h^{(j)} \right] = Y^{(tr)}_n X^{(tr)H}_n + Y^{(bl)}_n \tilde{V}^{(j)}_n C^H \tag{3.25}
$$

where we have defined

$$
\begin{align*}
\Lambda^{(n,j)}_{vv} &= E_{V^{(bl)}_n} \left[ V^{(bl)}_n V^{(bl)H}_n | Y^{(bl)}_n, h^{(j)} \right] \\
\tilde{V}^{(j)}_n &= E_{V^{(bl)}_n} \left[ V^{(bl)}_n | Y^{(bl)}_n, h^{(j)} \right] \quad \forall \ n = 0 \ldots N - 1
\end{align*}
$$

From this expression, it is clear that $\Lambda^{(n,j)}_{vv}$ is calculated using only the posterior second order moments of $V^{(bl)}_n$, whereas $\tilde{V}^{(j)}_n$ is the conditional mean (first order moment).

Therefore, each iteration of the EM-algorithm consists in calculating the first and second order statistics of the unknown symbols conditioned on the observations and on the current estimate of the channel (during the $E$-step), and in updating the estimate of the channel accordingly during the $M$-step.

In order to initialize the algorithm, we need a first estimate of the channel. One possible choice consists in using the training sequence estimate studied in the previous chapter. This doesn’t depend on the distribution of the unknown symbols, since it is determined using only the pilot observations.

Furthermore, we need also to define the termination conditions of the algorithm. Observing that the lower bound to the likelihood function $F(h, q)$ is ever increasing at each step and at each iteration of the EM-algorithm, and approaches a local maximum to the log-likelihood function, one possible approach for determining the convergence of the algorithm consists in calculating after each iteration (either after the M-step or after the E-step) the cost function $F(h, q)$ and comparing this value with the one obtained at the end of the previous iteration. If the new value differs by less than a certain threshold with respect to the value of the cost function in the previous iteration, the algorithm is assumed to have converged to a stationary point, and the algorithm is exited, otherwise another iteration is repeated, with the current channel estimate and posterior distribution of the unknown symbols as input. This is the approach used in the simulation results, when it was possible to compute the cost function.

Another approach consists in performing a fixed number of iterations. The advantage consists in the fact that there is no need to calculate the cost function, which represents a computational overhead. The disadvantage is that there is no possibility to control the closeness of the channel estimate to a local maximum of the likelihood function.

To sum up, the EM algorithm works as follows:
1. Initialize the channel to the training sequence estimator: $h^{(0)} = h^{(tr)}$. Set $j = -1$. Set the threshold for determining the convergence of the algorithm $\lambda$

2. Set $j := j + 1$

3. • **E-step**: compute the posterior mean and second order moment of the unknown symbols, using the current estimate of the channel, $h^{(j)}$

   \[
   \begin{align*}
   \Lambda^{(n,j)}_{xx} &= \mathbb{E}_{V_n^{(bl)}} \left[ V_n^{(bl)} V_n^{(bl)H} \middle| Y_n^{(bl)}, h^{(j)} \right] \\
   \tilde{V}^{(j)} &= \mathbb{E}_{V_n^{(bl)}} \left[ V_n^{(bl)} Y_n^{(bl)} H, h^{(j)} \right]
   \end{align*}
   \] (3.26)

   • **M-step**: update the channel matrix

   \[
   h^{(j+1)} = \mathcal{H} \left( \Lambda^{(n,j)}_{xx}, \Lambda^{(n,j)}_{yx}, B_{\eta_n}, n = 0 \ldots N - 1 \right)
   \] (3.27)

   with $\Lambda^{(n,j)}_{xx}$ and $\Lambda^{(n,j)}_{yx}$ given by 3.24

4. Calculate the new cost function $\mathcal{F}(h^{(j+1)}, q^{(j)})$ and the difference with respect to the one calculated in the previous iteration, that is

   \[
   \Delta^{(j)} = \mathcal{F}(h^{(j+1)}, q^{(j)}) - \mathcal{F}(h^{(j)}, q^{(j-1)})
   \] (3.28)

5. If $\Delta^{(j)} < \lambda$ the algorithm is assumed to have converged and is exited, otherwise another iteration is repeated (from step 2)

Once exited, the algorithm returns the current channel estimates, but also the posterior distribution of the unknown symbols, which can be used in the detection process.

The value assigned to the constant $\lambda$ determines the closeness of the channel estimate to a stationary point (local minimum) of the log-likelihood function. In fact, if the current channel estimate is relatively far away from a local maximum of the log-likelihood function, the difference $\Delta^{(j)}$ between the cost function evaluated at the current iteration and the cost function evaluated at the previous one is relatively large. Conversely, if the current channel estimate is relatively close to a local minimum of the log-likelihood function, $\Delta^{(j)}$ is relatively small. Therefore, the smaller is the value chosen for $\lambda$, the closer the channel estimate obtained with the EM-algorithm is to a local minimum of the log-likelihood function and the more accurate it is. However, the more closely we want to approach a local minimum to the log-likelihood function and the more iterations are needed to converge. Therefore the value for $\lambda$ is set by trading off estimation accuracy with convergence speed of the algorithm.

Now that we have derived a general treatment of the EM-algorithm applied to the Semi-Blind channel estimation problem, we investigate more in detail three cases: in the first
one the true discrete distribution is exploited for the estimate, in the second we use the Gaussian assumption for the unknown symbols, in the third we make the Constant Modulus assumption.

For all these three approaches, we use the EM-algorithm for the determination of a local maximum of the log-likelihood function. Observe that, once calculated the posterior first and second order moments of the unknown symbols, the M-step is identical for all the approaches. The difference resides in the E-step, since the calculation of the first and second order moments of the unknown symbols depends on their prior distribution and on the assumptions used. For this reason, in the next sections, when dealing with the EM-algorithm, we will develop in detail only the E-step, but we will not consider the M-step instead, since this is common to all cases.

3.2 Semi-Blind ML estimation: true discrete distribution of the unknown symbols

We start the study of the Semi-Blind channel estimation approach by considering the true discrete distribution of the unknown symbols.

With reference to the system model and the set of assumptions defined in 1.2, the unknown symbols are drawn uniformly from a finite discrete constellation $C^{S \times 1}$, where $S$ is the transmission rank, independently across the sub-carriers and across time.

Therefore, for all the unknown symbols $V_n^{(bl)}(k)$ we have

$$p \left( V_n^{(bl)}(k) \right) = \frac{1}{|C|^S} \quad \forall \quad V_n^{(bl)}(k) \in C^{S \times 1}$$  \hspace{1cm} (3.29)

Now, based on a set of observations, collected on the observation matrix $Y$, and a set of pilot symbols $X^{(tr)}$, the goal is to determine the ML estimate of the channel, which is solution to the likelihood equation, given by

$$\Delta_h \ln p \left( Y | X^{(tr)}, h \right) = 0$$  \hspace{1cm} (3.30)

where the gradient is calculated with respect to the time-domain channel matrix $h$, in order to enforce the channel length constraint.

As we saw in the general treatment 3.1 there is no closed form solution to this problem for the general Semi-Blind approach, therefore we seek for a local maximum to the log-likelihood function. We use the EM-algorithm to solve this maximization problem, which is treated in the following section.
3.2.1 ML solution through EM-algorithm

In section 3.1.2, we described the EM-algorithm for the determination of the ML solution, showing that the update of the channel estimate during the M-step depends only on the posterior first and second order statistics of the unknown symbols.

In fact, during the E-step we have to compute the posterior first and second order statistics

\[
\begin{align*}
\Lambda^{(n,j)}_{\nu} &= E_{V_{\nu}^{(M)}} \left[ V_{n}^{(bl)}(k) V_{n}^{(bl)H} | Y_{n}^{(bl)}(k), h^{(j)} \right] \quad \forall \ n = 0 \ldots N - 1 \\
\tilde{V}_{n}^{(j)} &= E_{V_{\nu}^{(M)}} \left[ V_{n}^{(bl)} | Y_{n}^{(bl)}(k), h^{(j)} \right] 
\end{align*}
\]  

(3.31)

In order to do it, we need the posterior distribution of the unknown symbols, given the current channel estimate \(h^{(j)}\). Therefore, for the unknown symbol on sub-carrier \(n\) at time \(k\), using Bayes’ rule we have

\[
p \left( V_{n}^{(bl)}(k) | Y_{n}^{(bl)}(k), h^{(j)} \right) = \rho p \left( Y_{n}^{(bl)}(k) | V_{n}^{(bl)}(k), h^{(j)} \right) p \left( V_{n}^{(bl)}(k) \right) 
\]  

(3.32)

where \(\rho\) is the normalization factor, independent of the value of the unknown symbol.

Now, \(V_{n}^{(bl)}(k)\) takes value from the discrete finite alphabet \(C^{S \times 1}\) with uniform distribution. Therefore \(p \left( V_{n}^{(bl)}(k) \right)\) is a constant independent of the value assumed by \(V_{n}^{(bl)}(k)\), which can therefore be included into the normalization factor. Finally, making explicit the probability density function \(p \left( Y_{n}^{(bl)}(k) | V_{n}^{(bl)}(k), h \right)\), and writing only the terms depending on \(V_{n}^{(bl)}(k)\), the posterior distribution of the unknown symbol on sub-carrier \(n\) at time \(k\), labeled with the notation \(q_{nk}^{(j)}\), can be written as

\[
q_{nk}^{(j)}(\beta) = \frac{\exp \left\{ \text{trace} \left[ B_{\nu} \left( Y_{n}^{(bl)}(k) - H_{n}^{(j)} C \beta \right) \left( Y_{n}^{(bl)}(k) - H_{n}^{(j)} C \beta \right)^{H} \right] \right\}}{\sum_{\alpha \in C^{S \times 1}} \exp \left\{ \text{trace} \left[ B_{\nu} \left( Y_{n}^{(bl)}(k) - H_{n}^{(j)} C \beta \right) \left( Y_{n}^{(bl)}(k) - H_{n}^{(j)} C \alpha \right)^{H} \right] \right\}} \\
\forall \beta \in C^{S \times 1}
\]  

(3.33)

From the posterior distribution on the unknown symbols \(q_{nk}^{(j)}(\beta)\), we can calculate the two matrices \(\tilde{V}_{n}^{(j)}\) and \(\Lambda^{(n,j)}_{\nu}\) defined in (3.31) as

\[
\begin{align*}
\tilde{V}_{n}^{(j)}(k) &= \sum_{\beta \in C^{S \times 1}} \beta \cdot q_{nk}^{(j)}(\beta) \quad \forall \ k \\
\Lambda^{(n,j)}_{\nu} &= \sum_{\beta \in C^{S \times 1}} \beta \beta^{H} \cdot \sum_{k} q_{nk}^{(j)}(\beta)
\end{align*}
\]  

(3.34)

As regards the complexity of this algorithm, observe that the computation of the terms above requires the calculation of the posterior distribution for each point of the constellation \(C^{S \times 1}\). Letting \(M\) be the constellation order, \(M^{S}\) posterior probabilities have to
be calculated for each unknown symbols. With 16-QAM and transmission rank \( S = 2 \), this corresponds to 256 posterior probabilities to be calculated. If we also add the fact that, in order to converge to an optimal solution, we need to perform multiple iterations of the \( E \) and \( M\)-steps, it is clear that the computational overhead of this algorithm is very high. Moreover, this solution is not scalable to higher order MIMO systems, since the number of posterior probabilities which need to be computed grows exponentially with the transmission rank.

In order to limit the number of iterations of the algorithm, instead of using the convergence criterion defined in the general description of the algorithm in section \( 3.1.2 \) we use a fixed number of iterations in the simulations. 5-6 iterations are sufficient to achieve a good convergence of the algorithm.

In the following sections \( 3.3 \) and \( 3.4 \) we study two approximations on the distribution of the unknown symbols, which can potentially reduce the computational overhead. However, as we will see in the simulation results, this reduction in complexity goes to the disadvantage of the estimation accuracy.

### 3.3 Semi-Blind ML estimation: Gaussian approximation for the unknown symbols

In the previous section, we studied the case where the true discrete distribution of the unknown symbols is taken into account, demonstrating the high computational overhead incurred with such an approach.

In this section, we relax the discreteness of the unknown symbols by assuming that they are circular Gaussian distributed.

Observe that, assuming that the distribution of the unknown symbols is circular Gaussian, implies that the distribution of the observations conditioned on the channel matrix is a multivariate Gaussian. Actually, the observations are distributed as a mixture of multivariate Gaussians. In fact, the distribution of the observations conditioned on the transmitted symbols is a multivariate Gaussian, therefore the marginalization over the discrete distribution of the unknown symbols leads to a mixture of Gaussians. However, we can approximate this distribution with a single multivariate Gaussian.

It is interesting to derive what is the best multivariate Gaussian \( q(X) \) which can be used as an approximation of the true distribution \( p(X) \). A widely used measure of closeness of a distribution to another is the Kullback–Leibler divergence, which for continuous
Chapter 3 Semi-Blind channel estimation

distributions is defined as

\[ KL(p||q) = \int_D p(X) \ln \left( \frac{p(X)}{q(X)} \right) dX \] (3.35)

where \( p(X) \) is the true PDF and \( q(X) \) is the PDF we want to use to approximate \( p(X) \).

Let’s assume we want to approximate \( p(X) \) with a multivariate Gaussian \( q(X) \) with mean \( m \) and covariance matrix \( \Sigma \). Then, the best \( m \) and \( \Sigma \) are obtained by minimizing the Kullback–Leibler divergence with respect to \( m \) and \( \Sigma \). It can be easily shown, by calculating the derivative and equaling it to zero, that the solution is given by

\[
\begin{align*}
  m &= E[X] \\
  \Sigma &= E[(X - m)(X - m)^H]
\end{align*}
\] (3.36)

where the expectation is taken with respect to the true distribution \( p(X) \).

Translating this example to our estimation problem, we want to approximate the distribution of the observations corresponding to the unknown symbols with a multivariate Gaussian \( q(Y) \) with mean \( m_Y \) and covariance matrix \( \Sigma_Y \). Using the set of assumptions defined in 1.2.3, the noise, the unknown symbols and consequently the observations are statistically independent across sub-carriers and across time. Then, using (3.36) on sub-carrier \( n \) at time \( k \) the mean value of the observations is given by

\[ m_{Y_n} = E[Y_n(k)] = E[H_nX_n(k) + W_n(k)] = 0 \] (3.37)

where we used the fact that the noise and the unknown symbols are zero mean.

Similarly, for the covariance matrix we obtain

\[ \Sigma_{Y_n} = E[Y_n(k)Y_n(k)^H] = H_nE[X_n(k)X_n(k)^H]H_n^H + \text{Cov} (\eta_n) \] (3.38)

Therefore, it is clear that approximating the distribution of the blind observations with a Gaussian distribution with zero mean and covariance matrix given by (3.38) is equivalent to approximating the distribution of the unknown symbols with a Gaussian distribution with zero mean and covariance matrix \( E[X_n(k)X_n(k)^H] \). Moreover, this is the best Gaussian approximation of the distribution of the blind observations.

It is interesting to understand how well the Gaussian assumption approximates the true distribution of the observations: the higher is the noise variance at the receiver with respect to the power of the symbols, the larger is the lobe of each multivariate Gaussian, the more overlap there is between pairs of multivariate Gaussians, and the better the true mixture of Gaussians is approximated with one multivariate Gaussian. Therefore, we expect this approximation to perform well especially in the low-SNR regime. We also
expect this approximation to perform the better the higher is the constellation order. In fact, for a given transmission power, the bigger is the constellation order \( M \), the closer are the projections of the transmitted symbols on the observation space (that is the points \( \{ HCV \in \mathbb{C}^{R \times 1} \; \forall \; V \in \mathbb{C}^{S \times 1} \} \)), and the more overlap there is between pairs of multivariate Gaussians belonging to the mixture. This holds true also for the transmission rank, as long as the dimension of the observation space, corresponding to the number of antennas \( R \), is kept fixed. In fact, the higher is \( S \), the more multivariate Gaussians, the closer gets their centers, and the more they overlap.

Now, with reference to the system model described in \[1.2.3\], we have for the unknown symbols \( E[X_n(k)X_n(k)^H] = \sigma^2_sCC^H \), therefore the distribution of the observations corresponding to blind information is a multivariate Gaussian, with zero mean and covariance and precision matrices given by

\[
\begin{align*}
\Sigma_{Y_n} &= \sigma^2_sH_nCC^H_n + \text{Cov}(\eta_n) \\
B_{Y_n} &= \Sigma_{Y_n}^{-1}
\end{align*}
\] (3.39)

In order to better understand the potential benefit achievable with this Semi-Blind approach, let’s consider the simple case of one sub-carrier and channel length \( L = 1 \). Moreover, let’s assume \( C = 1 \), which corresponds to no encoding across antennas, and \( R \geq T \). Then, the distribution of the blind observations is a multivariate Gaussian with zero mean and covariance matrix \( \text{Cov}(Y(k)) = \sigma^2_sHH^H + \text{Cov}(\eta) \). Observe that, letting \( H = USV^H \) be the singular value decomposition of the channel matrix, and substituting it into the expression for the covariance matrix, we obtain:

\[
\text{Cov}(Y(k)) = \sigma^2_sUSS^TU^H + \text{Cov}(\eta)
\] (3.40)

Observe that the distribution of the observations does not depend on the right unitary matrix \( V \), which means that the channel matrix is identifiable up to a rotation factor if we base the estimation only on the blind observations. Assuming that we are provided with a long enough sequence of blind observations, we can accurately estimate the \textit{whitening matrix} \( W = US \). The right unitary matrix \( V \), can then be estimated using only pilot symbols. Observe that \( V \) is a \( T \times T \) matrix, and given its unitary constraints, it is parameterized by \( T^2 \) real parameters. Therefore the pilots are used to estimate only \( T^2 \) real parameters instead of the usual \( 2RT \) required to estimate the whole channel matrix \( H \), which represents a factor \( \frac{2R}{T^2} \) improvement (\[7\], \[8\]). Even in the case \( R = T \) this corresponds to 3dB improvement in the mean square error of the channel estimator. This decomposition of the channel matrix is used in the papers \[7\] and \[8\], where the authors propose an algorithm for the estimation of the right unitary matrix \( V \) based only on the pilot sequence, assuming perfect knowledge of the whitening matrix \( W \).
In the more general case with more than one sub-carrier, with a channel length constraint \( L \) to enforce, and non perfect knowledge of the whitening matrix, we still improve the estimation accuracy using this semi-blind approach, since the blind observations provide information for estimating part of the channel, up to some uncertainties, which can be resolved using the pilot observations.

Now, let’s consider the likelihood equation 3.12 since the posterior expectation of the unknown symbols is a function of the channel matrix, there is no closed form solution to this equation. Therefore we can only determine a local maximum to the likelihood function. We propose the Expectation-Maximization algorithm, which is discussed in the next section.

### 3.3.1 ML estimate through EM Algorithm

As we showed in the general treatment in section 3.1.2, the \( E \)-step consists in calculating the posterior distribution and second order statistics of the unknown symbols, given the current channel estimate \( h^{(j)} \).

Using Bayes’ rule, the posterior distribution of the unknown symbol on sub-carrier \( n \) at time \( k \) is given by

\[
p \left( V_n^{(bl)}(k) | Y_n^{(bl)}(k), h \right) = \mu p \left( Y_n^{(bl)}(k) | V_n^{(bl)}(k), h \right) p \left( V_n^{(bl)}(k) \right)
\]

where \( \mu \) is the normalization factor, which does not depend on the unknown symbols.

Now, \( p(Y_n^{(bl)}(k) | V_n^{(bl)}(k), h) \) is a Gaussian PDF with mean \( H_n C V_n^{(bl)}(k) \) and covariance \( \text{Cov}(\eta_n) \) (precision \( B_{\eta_n} \)), and the unknown symbols \( V_n^{(bl)}(k) \) are Gaussian distributed with zero mean and covariance \( \sigma_s^2 I_S \). Therefore, keeping only the terms depending on the symbol \( V_n^{(bl)}(k) \) and including the others in the normalization factor \( \mu \) we have

\[
p(V_n^{(bl)}(k) | Y_n^{(bl)}(k), h) = \mu \exp \left\{ -V_n^{(bl)}(k)^H \left( C^H H_n^H B_{\eta_n} H_n C + \frac{1}{\sigma_s^2} I_S \right) V_n^{(bl)}(k) \right\} \cdot \exp \left\{ 2 \text{real} \left( V_n^{(bl)}(k)^H C^H H_n^H B_{\eta_n} V_n^{(bl)}(k) \right) \right\}
\]

(3.42)

However, when conditioned on \( Y_n^{(bl)}(k) \) and \( h \), \( V_n^{(bl)}(k) \) is Gaussian distributed with mean \( m_{V_n}(k) \) and covariance matrix \( \Sigma_{V_n}(k) \). Therefore we have also:

\[
p \left( V_n^{(bl)}(k) | Y_n^{(bl)}(k), h \right) = \lambda \exp \left\{ -V_n^{(bl)}(k)^H \Sigma_{V_n}(k)^{-1} V_n^{(bl)}(k) \right\} \cdot \exp \left\{ 2 \text{real} \left( V_n^{(bl)}(k)^H \Sigma_{V_n}(k)^{-1} m_{V_n}(k) \right) \right\}
\]

(3.43)

where \( \lambda \) is the normalization factor.
Comparing the above expression with equation 3.42 we have the following two equalities for the posterior covariance matrix $\Sigma_{V_n}(k)$ and for the posterior mean $m_{V_n}(k)$ of the unknown symbols at time $k$ on sub-carrier $n$, given the current update of the channel matrix $h^{(j)}$:

$$\begin{cases}
\Sigma_{V_n}^{(j)} = \left( C^H H_n^{(j)H} B_{\eta_n} H_n^{(j)} C + \frac{1}{\sigma^2} I_S \right)^{-1} \\
 m_{V_n}(k)^{(j)} = \Sigma_{V_n} C^H H_n^{(j)H} B_{\eta_n} Y_n^{(bl)}(k)
\end{cases} \quad (3.44)$$

where for the covariance term we dropped the time index $k$ since it is independent from it.

Then, stacking the posterior mean of the unknown symbols on a matrix using the time $k$ as column index, and we have

$$m_{V_n}^{(j)} = \Sigma_{V_n}^{(j)} C^H H_n^{(j)H} B_{\eta_n} Y_n^{(bl)} \quad (3.45)$$

From the posterior mean and covariance we can calculate the posterior first and second order moments of the unknown symbols as

$$\begin{cases}
\tilde{V}_n^{(j)} = m_{V_n}^{(j)} \\
K_n^{(bl)}(n,j) = m_{V_n}^{(j)}m_{V_n}^{(j)H} + K_n^{(bl)} \Sigma_{V_n}^{(j)}
\end{cases} \quad (3.46)$$

These matrices are then used during the $M$-step to update the channel matrix, as described in the general treatment in section 3.1.2.

### 3.4 Semi-Blind ML estimation: Constant Modulus approximation for the unknown symbols

In this section we propose a Semi-Blind MIMO-OFDM FIR channel estimation technique based on the assumption that the unknown symbols are drawn from a constant modulus alphabet. By constant modulus, it is meant a modulation technique with the property that all the points in the constellation have the same amplitude. By constant modulus, it is meant a modulation technique with the property that all the points in the constellation have the same amplitude. In section 3.3 we studied a semi-blind channel estimator relying on the Gaussian approximation on the distribution of the unknown symbols. In fact, the Gaussian assumption means that we have two degrees of uncertainty on the transmitted symbols: amplitude and phase. Conversely, the points in a constant modulus constellation have only one degree of freedom, the phase, since the amplitude is fixed. While in the Gaussian assumption the phase of the symbols is uniformly distributed in the range $[0, 2\pi)$ and the amplitude is Rayleigh distributed, in the Constant Modulus assumption used throughout this section, the amplitude is fixed and known, while the phase of the symbols is assumed to be...
uniformly distributed in the range \([0, 2\pi]\). Therefore, given the less degree of freedom on the unknown symbols, we expect to achieve a more accurate estimate than the Gaussian assumption. This will be demonstrated in the Simulation Results, presented in chapter 5.

The challenge with the Constant Modulus assumption is that it is difficult to effectively exploit this property. Many Semi-Blind estimation approaches have been proposed relying on this assumption (see for example 9, 10 and 11). In particular, in 9 a Constant Modulus algorithm relying on higher order statistics of the observations has been proposed. However, this algorithm suffers from noise amplification, therefore it relies on averaging over long observation sequences; moreover its applicability is limited to SISO systems. In this thesis we propose an alternative algorithm, based on a Taylor series expansion of the posterior probabilities of the unknown symbols, for the limit case of the constellation order \(M\) going to infinity. This algorithm performs well also with a short sequence of blind observations, as we will show in the simulation results. However, its applicability is limited to MIMO-OFDM systems with transmission rank one (\(S = 1\)).

In section 3.1 we saw that the Maximum Likelihood estimate is solution to the following equation:

\[
- \frac{\partial \ln p(Y|H)}{\partial h_l} = - \sum_{n=0}^{N-1} B_{\eta_n} \left( Y_n^{(tr)} - H_n X_n^{(tr)} \right) X_n^{(tr)H} e^{i2\pi \frac{ln}{N}} + \\
- \sum_{n=0}^{N-1} B_{\eta_n} E_{V_n^{(b)}|Y_n^{(b)},h} \left[ \left( Y_n^{(b)} - H_n CV_n^{(b)} \right) V_n^{(b)H} C^H \right] e^{i2\pi \frac{ln}{N}} = 0
\]

\[\forall \ l = 0 \ldots L - 1 \quad (3.47)\]

Also with the assumption of Constant Modulus alphabet, as with the Gaussian and the discrete assumptions for the unknown symbols, the ML solution cannot be determined in closed form from the above likelihood equation, since the posterior distribution of the unknown symbols is a function of the channel. Therefore, again, we use the EM-algorithm to determine a local maximum to the log-likelihood function.

### 3.4.1 ML solution through EM-algorithm

From the general treatment provided in 3.1.2 we see that the calculations involved in the \textit{M-step} require only the first and second order moments of the unknown symbols, which are calculated during the \textit{E-step}.

Observe that, assuming rank-one transmission (\(S = 1\)), and assuming that the unknown symbols \(V_n(k)\) are drawn from a constant modulus alphabet, the term \(V_n(k)V_n(k)^H\) is
deterministically equal to the symbol power $\sigma_s^2$, independently of the observations and of the channel realization. Therefore:

$$
E_{V_n^{(bl)}} \left[ V_n^{(bl)} V_n^{(bl)H} \right| Y_n^{(bl)}, h] = K_n^{(bl)} \sigma_s^2
$$

(3.48)

For the other expectation term $E_{V_n^{(bl)}} \left[ V_n^{(bl)H} \right| Y_n^{(bl)}, h]$ there is not such simple property.

There are two possible approaches to calculate the posterior mean of the unknown symbols: the first one consists in calculating the posterior expectation based on the true discrete distribution of the input symbols. This case was considered in section 3.2, where we showed that, although optimal from the point of view of the estimation accuracy, since it takes into account the true distribution of the unknown symbols, it is a computationally demanding algorithm, since it requires the computation of $p(\alpha|Y, H)$ for any point $\alpha \in \mathcal{C}$. The second approach consists in relaxing the assumption of discreteness of the input symbols, and approximating the posterior mean by considering the limit case of the constellation order $M$ going to infinity, which is equivalent to assuming the symbols have constant amplitude and phase uniformly distributed in $[0, 2\pi)$. The latter is the approach used here.

Observe that, assuming for now $S \geq 1$, and letting $V_{nk} \in \mathcal{C}^{S \times 1}$ be the unknown symbol transmitted on sub-carrier $n$ at time $k$, and $Y_{nk}$ the corresponding observation, the posterior mean of the unknown symbol is given by

$$
E_{V_{nk}} [V_{nk}| Y_{nk}, h] = \sum_{\alpha \in \mathcal{C}^{S \times 1}} \alpha p(\alpha|Y_{nk}, h)
$$

(3.49)

Now, using Bayes’ rule, we can write the posterior distribution as

$$
p(\alpha|Y_{nk}, h) = \mu p(Y_{nk}|\alpha, h) p(\alpha)
$$

(3.50)

where $\mu$ is the normalization factor, independent from $\mu$, and the prior distribution $p(\alpha)$ is a constant with respect to $\alpha$, since the symbols are drawn uniformly from the alphabet, therefore $p(\alpha) = \frac{1}{|\mathcal{C}|} \pi$.

Then, under the assumption that the noise is Gaussian with zero mean and precision matrix $\mathcal{B}_{\eta_n} = \text{Cov}(\eta_n)^{-1}$, we have

$$
E_{V_{nk}} [V_{nk}| Y_{nk}, h] = \frac{\sum_{\alpha \in \mathcal{C}^{S \times 1}} \alpha \exp \left\{ -\text{trace} \left[ \mathcal{B}_{\eta_n} (Y_{nk} - H_n C\alpha) (Y_{nk} - H_n C\alpha)^H \right] \right\}}{\sum_{\alpha \in \mathcal{C}^{S \times 1}} \exp \left\{ -\text{trace} \left[ \mathcal{B}_{\eta_n} (Y_{nk} - H_n C\alpha) (Y_{nk} - H_n C\alpha)^H \right] \right\}}
$$

(3.51)
For the exponential term in the above expression we have

\[ \exp \left\{ -\text{trace} \left[ B_{\eta_n} (Y_{nk} - H_n C \alpha) (Y_{nk} - H_n C \alpha)^H \right] \right\} \]

\[ = \mu \exp \left\{ -\alpha^H C^H H_n^H B_{\eta_n} H_n C \alpha \right\} \exp \{ 2\text{real} (Y_{nk}^H B_{\eta_n} H_n C \alpha) \} \quad (3.52) \]

where \( \mu \) is a constant which does not depend on \( \alpha \).

Then, letting \( \gamma_{s1s2} = (C^H H_n^H B_{\eta_n} H_n C)_{s1s2} \) and \( \xi_s = (Y_{nk}^H B_{\eta_n} H_n C)_s \) we can rewrite the above exponential term as

\[ \exp \left\{ -\text{trace} \left[ B_{\eta_n} (Y_{nk} - H_n C \alpha) (Y_{nk} - H_n C \alpha)^H \right] \right\} \]

\[ = \mu \exp \left\{ -\sum_{s1} \gamma_{s1s1} |\alpha_{s1}|^2 \right\} \exp \left\{ -\sum_{s1,s2 \neq s1} (\gamma_{s1s2} \alpha_{s2} \alpha_{s1}^*) \right\} \exp \{ 2\text{real} \sum_s \alpha_s \xi_s \} \quad (3.53) \]

and using the constant modulus assumption we have \(|\alpha_{s1}|^2 = \sigma_s^2\), therefore, including the terms independent of \( \alpha \) in the factor \( \mu \) we obtain

\[ \exp \left\{ -\text{trace} \left[ B_{\eta_n} (Y_{nk} - H_n C \alpha) (Y_{nk} - H_n C \alpha)^H \right] \right\} = \]

\[ = \mu \exp \left\{ -\sum_{s1,s2 \neq s1} (\gamma_{s1s2} \alpha_{s2} \alpha_{s1}^*) \right\} \exp \{ 2\text{real} \sum_s \alpha_s \xi_s \} \quad (3.54) \]

Finally, we can rewrite (3.49) as:

\[ E_{V_{nk}} [V_{nk} | Y_{nk}, h] = \frac{\sum_{\alpha \in \mathbb{C}^{S \times 1}} \alpha \exp \left\{ -\sum_{s1,s2 \neq s1} (\gamma_{s1s2} \alpha_{s2} \alpha_{s1}^*) \right\} \exp \{ 2\text{real} \sum_s \alpha_s \xi_s \}} {\sum_{\alpha \in \mathbb{C}^{S \times 1}} \exp \left\{ -\sum_{s1,s2 \neq s1} (\gamma_{s1s2} \alpha_{s2} \alpha_{s1}^*) \right\} \exp \{ 2\text{real} \sum_s \alpha_s \xi_s \}} \quad (3.55) \]

In the case of transmission rank \( S = 1 \), the above expectation simplifies to

\[ E_{V_{nk}} [V_{nk} | Y_{nk}, h] = \frac{\sum_{\alpha \in \mathbb{C}} \alpha \exp \left\{ 2\text{real} (\alpha \xi) \right\}} {\sum_{\alpha \in \mathbb{C}} \exp \left\{ 2\text{real} (\alpha \xi) \right\}} \]

\[ = \frac{\sum_{\alpha \in \mathbb{C}} \alpha \exp \left\{ 2\text{real} (\alpha Y_{nk}^H B_{\eta_n} H_n C) \right\}} {\sum_{\alpha \in \mathbb{C}} \exp \left\{ 2\text{real} (\alpha Y_{nk}^H B_{\eta_n} H_n C) \right\}} \quad (3.56) \]

We see that in the case \( S > 1 \) there is one more term in the expression for the posterior expectation, given by \( \exp \left\{ -\sum_{s1,s2 \neq s1} (\gamma_{s1s2} \alpha_{s2} \alpha_{s1}^*) \right\} \), which keeps into account the correlation between the symbols across the transmission streams. Because of this term, it was not possible to derive a simple expression for the limit case of the constellation order \( M \) going to infinity for the general case \( S \geq 1 \), but only for the case \( S = 1 \), for which we see that this term fades away (equation 3.56). Moreover, for \( S > 1 \) property
3.48 doesn’t hold anymore, which is a further argument for considering only the case $S = 1$ in the rest of the treatment.

Assuming as justified above $S = 1$, and assuming the unknown symbols are drawn from a 4-QAM or $M$-PSK constellation of any order $M$, the idea is to perform a Taylor series expansion of the exponential term $\exp \{ 2 \text{real} (\alpha Y_n^H B_n H_n C) \}$ in 3.56 and then calculating the limit case of the constellation order going to infinity.

The computations involved are quite cumbersome, therefore we defer the interested reader to Appendix B for the derivations. Using this approach, in the appendix we show that the posterior expectation of the unknown symbols can be approximated with the following expression

$$E_{V_n^{(bl)}(k)} \left[ V_n^{(bl)}(k) \right | Y_n^{(bl)}(k), h] = \sigma_s e^{i\theta_n} \frac{\sum_{n=0}^{+\infty} \frac{1}{n!(n+1)!} (|\rho_{nk}|\sigma_s)^{2n+1}}{\sum_{n=0}^{+\infty} \frac{1}{(n)^2} (|\rho_{nk}|\sigma_s)^{2n}} = \sigma_s e^{i\theta_n} g(|\rho_{nk}|\sigma_s) \quad (3.57)$$

where we defined the complex term $\rho_{nk} = C^H H_n^H B_n Y_n^{(bl)}(k)$, and $\theta_{nk}$ is the phase of $\rho_{nk}$.

We have also defined the scalar function:

$$g(x) = \frac{\sum_{n=0}^{+\infty} \frac{1}{n!(n+1)!} x^{2n+1}}{\sum_{n=0}^{+\infty} \frac{1}{(n)^2} x^{2n}} \quad \forall \quad x \geq 0 \quad (3.58)$$

Notice that the approximation to the posterior expectation 3.57 has amplitude $\sigma_s g(|\rho_{nk}|\sigma_s)$ solely depending on the factor $|\rho_{nk}|\sigma_s$, and phase $\theta_{nk} = \text{phase} (\rho_{nk})$. The term $\sigma_s e^{i\theta_{nk}}$ has a clear significance: it is the Maximum Likelihood estimate of the symbol $V_n^{(bl)}(k)$, assumed to have constant amplitude $\sigma_s$ and uniform phase between 0 and $2\pi$.

In fact, writing the likelihood of the observation $Y_n^{(bl)}(k)$ conditioned on the channel and on the phase $\theta_{nk}$ of the transmitted symbol $V_n^{(bl)}(k) = \sigma_s e^{i\theta_{nk}}$, we have:

$$- \ln p \left( Y_n^{(bl)}(k) | \theta_{nk}, h \right) = - \ln \left( \frac{|B_m|}{\pi R} \right) +$$

$$+ \text{trace} \left[ B_{\eta_n} \left( Y_n^{(bl)}(k) - H_n C \sigma_s e^{i\theta_{nk}} \right) \left( Y_n^{(bl)}(k) - H_n C \sigma_s e^{i\theta_{nk}} \right)^H \right]$$

$$= \mu - \left( C^H H_n^H B_{\eta_n} Y_n^{(bl)}(k) \right) \sigma_s e^{-i\theta_{nk}} - \left( C^H H_n^H B_{\eta_n} Y_n^{(bl)}(k) \right)^* \sigma_s e^{i\theta_{nk}}$$

$$= \mu - \sigma_s \rho_{nk} e^{-i\theta_{nk}} - \sigma_s \rho_{nk}^* e^{i\theta_{nk}} \quad (3.59)$$
where $\mu$ is a constant term independent of $\theta_{nk}$ and in the last equality we used the definition of $\rho_{nk}$ given above.

Then, calculating the derivative with respect to $\theta_{nk}$ and equaling it to zero we obtain:

$$
- \frac{\partial \ln p \left( Y_n^{(bl)}(k) | \theta_{nk}, h \right)}{\partial \theta_{nk}} = i \sigma_s \rho_{nk} e^{-i \theta_{nk}} - i \sigma_s \rho_{nk}^* e^{i \theta_{nk}} \\
= -2 \sigma_s \text{imag} \left( \rho_{nk} e^{-i \theta_{nk}} \right) = 0 \quad (3.60)
$$

There are two solutions solution to the above equation:

$$
\begin{align*}
\theta_{nk}^{(0)} &= \text{phase} \left( \rho_{nk} \right) \\
\theta_{nk}^{(1)} &= \text{phase} \left( \rho_{nk} \right) + \pi 
\end{align*} \quad (3.61)
$$

However, other than solution to the likelihood equation, another necessary condition for the ML solution is that the second derivative of the negative log-likelihood function evaluated at the ML solution is greater than zero, since this condition forces the ML solution to be a minimum, not a maximum of the negative log-likelihood function. Therefore, calculating the derivative of $3.60$ again with respect to $\theta_{nk}$ we obtain:

$$
- \frac{\partial^2 \ln p \left( Y_n^{(bl)}(k) | \theta_{nk}, h \right)}{\partial \theta_{nk}^2} = \sigma_s \text{real} \left( \rho_{nk} e^{-i \theta_{nk}} \right) \quad (3.62)
$$

and calculating this function in correspondence of $\theta_{nk}^{(0)}$ and $\theta_{nk}^{(1)}$ we see that the ML solution is $\theta_{nk} = \text{phase} \left( \rho_{nk} \right)$.

Now, let’s consider the amplitude normalized to $\sigma_s$ of the posterior expectation, given by the function $g(|\rho|\sigma_s)$ in $3.57$:

$$
g(|\rho|\sigma_s) = \frac{\sum_{n=0}^{\pm \infty} \frac{1}{n!(n+1)!} (|\rho|\sigma_s)^{2n+1}}{\sum_{n=0}^{\pm \infty} \frac{1}{(n)!^2} (|\rho|\sigma_s)^{2n}} \quad (3.63)
$$

Since it is not possible to solve analytically the above sum, we seek for an approximation. Let $\tilde{g}_N(x)$ be the function obtained by taking the first $N$ terms of the numerator and denominator in $3.63$ that is:

$$
\tilde{g}_N(x) = \frac{\sum_{n=0}^{N} \frac{1}{n!(n+1)!} (x)^{2n+1}}{\sum_{n=0}^{N} \frac{1}{(n)!^2} (x)^{2n}} \quad (3.64)
$$

This function is plotted in figure $3.1$ for different values of $N$. 

Then we have also:

$$\lim_{N \to +\infty} \tilde{g}_N(x) = g(x) \quad (3.65)$$

We observe that the series of functions \( \{\tilde{g}_N\} \) approaches the black curve \( \tilde{g}_{20}(x) \) for growing values of \( N \), which is equal to zero for \( x = 0 \) and converges to one for growing values of \( x \). Therefore we expect \( \tilde{g}_{20}(x) \) to be a close approximation of \( g(x) \).

The behavior of this function can be intuitively understood by considering the statistical properties of the term \( \sigma_s \rho_{nk} \), in the low and high-SNR ranges. In fact, assuming for simplicity white Gaussian noise at the receiver with variance \( \sigma^2_w \) and considering the term \( \sigma_s \rho_{nk} \) as a random variable, its mean and variance are given by:

$$
\begin{align*}
E\left[\sigma_s \rho_{nk}\right] &= 0 \\
E\left[\sigma_s^2 \left|\rho_{nk}\right|^2\right] &= \sigma_s^2 C^H H_n^H B_{\eta_n} E\left[Y_n^{(b)}(k)Y_n^{(a)}(k)^H\right] B_{\eta_n} H_n C \\
&= \frac{\sigma_s^2}{\sigma_w^2} \left(C^H H_n^H H_n C\right) \left[\frac{\sigma_s^2}{\sigma_w^2} \left(C^H H_n^H H_n C\right) + 1\right]
\end{align*}
$$

(3.66)

where we used the Constant Modulus property and the assumption of independence of the transmitted symbols from the noise.

In the low-SNR regime we have \( \frac{\sigma_s^2}{\sigma_w^2} \ll 1 \), therefore for the variance of \( \sigma_s \rho_{nk} \) we have:

$$
E\left[\sigma_s^2 \left|\rho_{nk}\right|^2\right] \approx \frac{\sigma_s^2}{\sigma_w^2} C^H H_n^H H_n C \ll 1
$$

(3.67)

which means that \( \sigma_s \rho_{nk} \) is statistically small, and accordingly \( g(\sigma_s \rho_{nk}) \), that is the amplitude of the posterior expectation, is small (see figure 3.1 curve \( \tilde{g}_{20}(x) \) for small values
of $x$). This behavior is the one expected, since in the low-SNR regime the observations carry mostly noise, and very few information about the transmitted symbols, therefore the posterior mean is close to the prior mean, which is zero.

Conversely, in the high-SNR regime we have $\frac{\sigma^2_s}{\sigma^2_w} \gg 1$, therefore for the variance of $\sigma_s \rho_{nk}$ we have:

$$E \left[ \sigma^2_s | \rho_{nk} \right] = \frac{\sigma^4_s}{\sigma^4_w} (C^H H_n^H H_n C)^2 \gg 1 \quad (3.68)$$

which means that $\sigma_s \rho_{nk}$ is statistically large, and accordingly $g(\sigma_s | \rho_{nk})$ is close to 1. Similarly, this high-SNR regime behavior is the one expected, since the observations carry mostly information about the transmitted symbols, therefore the posterior mean is close to the true transmitted symbol, or equivalently it is close to the circle of amplitude $\sigma_s$.

Therefore, we can statistically associate large values of $\sigma_s | \rho_{nk}$ to the high-SNR regime, and small values to the low-SNR regime.

Since it is not practical to use the truncated series expansion, we want to approximate the curve $g(x)$ (or equivalently its truncated version $\tilde{g}_{20}(x)$) with another simpler function. We verified that one close approximation is of the form $\hat{g}(x, \alpha) = 1 - e^{-\alpha x}$, for some positive real $\alpha$. In fact this function is also equal to zero for $x = 0$, is strictly lower than one for $x > 0$ and converges to 1 for $x \to +\infty$. The coefficient $\alpha$ was determined by minimizing the Mean Square Error between the approximation and $\tilde{g}_{20}(x) \simeq g(x)$. Using this approach, we determined the optimum coefficient to be $\alpha = 1.0639$. Therefore, the
approximation to the posterior expectation of the unknown symbols can be written as

$$E_{V_n^{(bl)}(k)} \begin{bmatrix} V_n^{(bl)}(k) | Y_n^{(bl)}(k), h \end{bmatrix} \simeq \sigma_s e^{i\theta_{nk}} \left( 1 - e^{-1.0639 \cdot \sigma_s |\rho_{nk}|} \right)$$

(3.69)

In figure 3.2 we show curve $g_{20}(x)$ and the approximation $\hat{g}(x, 1.0639)$, as well as the error on the amplitude.

It is interesting to compare the closeness of the posterior expectation using the Gaussian approximation (MMSE detector) and using the Constant Modulus approximation for the transmitted symbols to the true posterior expectation calculated averaging over the true discrete distribution of the symbols. Figure 3.3 shows the standard deviation of the error between the true posterior expectation and the approximated posterior expectation for different SNR and different number of bits per symbol, for the two cases where the symbols are assumed to be Gaussian distributed (the approximation used in section 3.3) and where they are assumed to be Constant Modulus with phase uniformly distributed in $[0, 2\pi)$. In this latter case the posterior expectation is calculated using the approximation to the posterior mean given by (3.69). It is worth noticing that the Constant Modulus approximation proposed leads to a significant improvement compared to the Gaussian assumptions, even for a small number of bits (the 2 bits case is particularly interesting, since this corresponds to the 4-QAM constellation used in the LTE system). Moreover,
the standard deviation decreases over the number of bits, since the more bits there are, the more evenly the symbols are distributed on the unit circle, and the better their phase can be approximated as being uniform in $[0, 2\pi)$.

To sum up, during the $E$-step the posterior mean of the unknown symbols is calculated using the current estimate of the channel $h^{(j)}$ as:

$$
\begin{align*}
\rho_{nk}^{(j)} = & C H_n^{(j)H} B_n Y_n^{(bl)}(k) \\
\theta_{nk}^{(j)} = & \text{phase} \left( \rho_{nk}^{(j)} \right) \\
E_{V_n^{(bl)}(k)} \left[ V_n^{(bl)}(k) \bigg| Y_n^{(bl)}(k), h^{(j)} \right] & \simeq \sigma_s e^{i\theta_{nk}^{(j)}} \left( 1 - e^{-1.0639\sigma_s |\rho_{nk}^{(j)}|} \right) = \tilde{V}_n^{(j)}(k)
\end{align*}
$$

(3.70)

Similarly, from (3.48) we have

$$
\Lambda_{vv}^{(n,j)} = K_n^{(bl)} \sigma_s^2
$$

(3.71)

These terms are then fed into the $M$-step to produce a new estimate of the channel, as explained in the general treatment 3.1.2.
Chapter 4

Joint Semi-Blind Estimation of channel and noise covariance matrix

In the previous chapters we assumed that the statistical properties of the noise (the noise covariance matrices \( \{\text{Cov}(\eta_n), \forall \ n\} \)) were known at the receiver. This knowledge allows for a more accurate estimation of the channel, since there is less uncertainty on the parameters modeling the system, but is unrealistic, since the statistical properties of the noise need to be estimated at the receiver, and this must be performed jointly with the channel.

Observe that the channel estimators studied in chapters 2 and 3 take as an input the noise covariance matrix. Therefore, we expect that the non-perfect knowledge of the noise covariance matrix at the receiver negatively impacts the channel estimate. Moreover, as we will see in the course of this chapter, also the noise covariance estimator takes as input the current channel estimate, therefore there is an interdependency between the channel and the noise covariance estimators. This issue is resolved by performing a joint estimate of the channel and of the covariance matrices on each sub-carrier. This is the topic of the chapter.

This chapter is organized as follows. In the first section (section 4.1) we statistically model the noise at the receiver, in order to identify an unconstrained set of parameters modeling the noise covariance matrix on each sub-carrier, under the assumption that the noise at the receiver is given by two contributions: a white Gaussian process and multi-user interference. Then in section 4.2 we derive an algorithm for the estimation of the noise covariance matrix on each sub-carrier, assuming perfect knowledge of the channel.
Finally, in section 4.3 we derive an algorithm for the joint estimation of the channel and of the noise covariance matrix. In particular, the main focus is on Semi-Blind estimation, that is the parameters governing the system (channel and noise covariance matrix) are jointly estimated using all the information available at the receiver.

### 4.1 Noise Model

In this section we derive a model of the noise at the receiver. The importance of such parameterization, as we demonstrate, derives from the fact that there is a functional dependence of the covariance matrices across the sub-carriers, which can be exploited to enhance the estimation accuracy with respect to the case where the covariance matrices are estimated independently on each sub-carrier. Basically, with such parameterization, the covariance matrices are identified by a smaller number of parameters with respect to the case where they are assumed to be functionally independent across the sub-carriers.

With reference to the system model defined in 1.2.2 on each sub-carrier \( n \) we have the following input output relation:

\[
Y_n = H_n X_n + \eta_n
\]  

(4.1)

So far, we have assumed that the noise \( \eta_n \) is a zero mean Gaussian process, independent across sub-carriers and across time, with covariance \( \text{Cov}(\eta_n) \) which is perfectly known at the receiver. Now, we go one step further, and we try to model appropriately the noise covariance matrix, identifying the minimum set of parameters describing the statistics of the noise at the receiver.

In particular, we assume that \( \eta_n \) is given by two contributions: the first is a purely circular white Gaussian process, with variance \( \sigma_w^2 \) on all sub-carriers and on all receiving antennas, represented by matrix \( W_n \); the other is multi user interference.

For the second contribution, the multi-user interference, we assume that there are \( U \) interferers using a MIMO-OFDM system, and that the channel between each interferer and the receiver is a MIMO-FIR channel of length \( L \), with \( T_u \) transmitting and \( R \) receiving antennas. Furthermore, we assume that the interferers are synchronized with the receiver, in such a way that the transformation between the interfering transmitters and the receiver is still circular.
Chapter 4 Joint Semi-Blind Estimation of channel and noise covariance matrix

Under these assumptions, the interference received on sub-carrier \( n \) at time \( k \) from user \( u \) is given by:

\[
\gamma_{n}^{(u)}(k) = H_{n}^{(u)} X_{n}^{(u)}(k)
\]  

(4.2)

where \( X_{n}^{(u)}(k) \in \mathbb{C}^{T_{u} \times 1} \) is the symbol vector transmitted by interferer \( u \) at sub-carrier \( n \) at time \( k \), and \( H_{n}^{(u)} \in \mathbb{C}^{R \times T_{u}} \) represents the channel matrix between interferer \( u \) and the receiver on sub-carrier \( n \). \( X_{n}^{(u)}(k) \) is assumed to be a circular white Gaussian vector, independent across sub-carriers, across time, from the other interferers and from the white Gaussian process, with covariance matrix \( E[X_{n}^{(u)}(k)X_{n}^{(u)}(k)^{H}] = \sigma_{s}^{(u)2} I_{T_{u}} \).

Then, summing together the contribution of the white Gaussian noise and of the interferers we have the following expression for the noise at the receiver:

\[
\eta_{n}(k) = \sum_{u=0}^{U-1} \gamma_{n}^{(u)}(k) + W_{n}(k) = \sum_{u=0}^{U-1} H_{n}^{(u)} X_{n}^{(u)}(k) + W_{n}(k)
\]  

(4.3)

Since the symbols transmitted by the interferers \( X_{n}^{(u)}(k) \) and the noise \( W_{n}(k) \) are Gaussian distributed and independent random variables, independent across the sub-carriers and across time, then also the distribution of \( \eta_{n}(k) \) conditioned on the channels between the interferers and the receiver is a zero mean Gaussian vector, independent across the sub-carriers and across time, with covariance matrix:

\[
\text{Cov} (\eta_{n}(k)) = E \left[ \left( \sum_{u=0}^{U-1} H_{n}^{(u)} X_{n}^{(u)} + W_{n} \right) \left( \sum_{u=1}^{U} H_{n}^{(u)} X_{n}^{(u)} + W_{n} \right)^{H} \right] = \\
= \sum_{u=0}^{U-1} \sigma_{s}^{(u)2} H_{n}^{(u)} H_{n}^{(u)H} + \sigma_{w}^{2} I_{R}
\]  

(4.4)

Since the interferers’ channels are FIR of length \( L \), for each interferer on each sub-carrier we can write the channel matrix as:

\[
H_{n}^{(u)} = \sqrt{N} \left( I_{R} \otimes \tilde{U}_{N}^{(n)} \right) h^{(u)}
\]  

(4.5)

where \( \tilde{U}_{N}^{(n)} \) represents the \( n \)th row of matrix \( \tilde{U}_{N} \), which is obtained by taking the first \( L \) columns of the Fourier matrix \( U_{N} \) with entries \( U_{N}(n,m) = \frac{1}{\sqrt{N}} e^{-i2\pi \frac{nm}{N}} \). \( h^{(u)} \) is the time-domain channel matrix, obtained by stacking the \( L \) channel taps on a column.

Then, substituting the above expression for \( H_{n}^{(u)} \) into (4.4) we obtain:

\[
\text{Cov} (\eta_{n}(k)) = N \left( I_{R} \otimes \tilde{U}_{N}^{(n)} \right) \left( \sum_{u=0}^{U-1} \sigma_{s}^{(u)2} h^{(u)} h^{(u)H} \right) \left( I_{R} \otimes \tilde{U}_{N}^{(n)} \right)^{H} + \sigma_{w}^{2} I_{R}
\]  

(4.6)
Finally, using the fact that $\frac{U^{(n)}_N U^{(n)H}_N}{N} = \frac{L}{N}$, we can rewrite:

$$\text{Cov}(\eta_n(k)) = N \left( I_R \otimes \tilde{U}^{(n)}_N \right) \left( \sum_{u=0}^{U-1} \sigma^2_s(u) h(u) h^H(u) + \frac{\sigma_w^2}{L} I_{RL} \right) \left( I_R \otimes \tilde{U}^{(n)}_N \right)^H =$$

$$= N \left( I_R \otimes \tilde{U}^{(n)}_N \right) \Sigma \left( I_R \otimes \tilde{U}^{(n)}_N \right)^H \quad (4.7)$$

where we defined the $LR \times LR$ matrix $\Sigma$ as:

$$\Sigma = \sum_{u=0}^{U-1} \sigma^2_s(u) h(u) h^H(u) + \frac{\sigma_w^2}{L} I_{RL} \quad (4.8)$$

$\Sigma$ is an Hermitian positive definite matrix. In fact, from the definition of positive definite matrix, for any non null $x \in \mathbb{C}^{RL \times 1}$ we have, assuming $\sigma_w^2 > 0$

$$x^H \Sigma x = \sum_{u=0}^{U-1} \sigma^2_s(u) x^H h(u) h^H(u) x + \frac{\sigma_w^2}{L} x^H x \geq \frac{\sigma_w^2}{L} x^H x > 0 \quad (4.9)$$

Similarly, $\sum_{u=0}^{U-1} \sigma^2_s(u) h(u) h^H(u)$ is semi-definite positive, and letting $QDQ^H$ be its eigenvalue decomposition, with $Q$ unitary matrix and $D$ diagonal matrix with non-negative diagonal entries, we have:

$$\Sigma = QDQ^H + \frac{\sigma_w^2}{L} I_{RL} = Q \left( D + \frac{\sigma_w^2}{L} I_{RL} \right) Q^H \quad (4.10)$$

We observe that, if the number of interferers is $U = 0$, then $\Sigma = \frac{\sigma_w^2}{L} I_{RL}$ is parameterized by only one parameter, $\sigma_w^2$. Conversely, if the diagonal elements of $D$ are strictly positive, we allow full degree of freedom on the eigenvalues of $D$, and consequently on the eigenvalues of $\Sigma$, which means that $\Sigma$ can be any positive-definite matrix, therefore it needs the full parameterization of a positive definite matrix. In this case, since $\Sigma$ is positive definite, hence Hermitian matrix, it is parameterized by $(LR)^2$ real parameters: $LR$ real positive elements on the main diagonal, and $(LR)^2 - LR$ on the upper-right triangle (both real and imaginary part); the lower-left triangle is determined by the upper-right triangle from the Hermitian nature of $\Sigma$.

This full degree of freedom is achieved when the number of SIMO channels between the interferers and the receiver is greater than $LR$, that is:

$$\sum_{u=0}^{U-1} T_u \geq LR \quad (4.11)$$
In fact, letting $h^{(u,t)}$ be the SIMO channel between transmitting antenna $t$ of interferer $u$, we can write:

$$
\sum_{u=0}^{U-1} \sigma_s^{(u)} h^{(u)} h^{(u)H} = \sum_{u=0}^{U-1} \sum_{t=0}^{T_u-1} \sigma_s^{(u,t)} h^{(u,t)} h^{(u,t)H}
$$

(4.12)

whose rank is less or equal to $\sum_{u=0}^{U-1} T_u$.

Since we don’t know a priori how many users interfere with the communication, we always assume that there are enough users to give full-degree of freedom on $\Sigma$.

Notice that a sufficient condition for the covariance matrix on each sub-carrier to be positive definite is that $\Sigma$ is positive-definite. Therefore, any positive-definite $\Sigma$ satisfies the positive definite constraint on $\text{Cov}(\eta_n)$. In fact, for any non null $x \in \mathbb{C}^{R \times 1}$, using the definition of positive definite matrix we have

$$
x^H \text{Cov}(\eta_n(k)) x = N x^H \left( I_R \otimes \hat{U}^{(n)} \right) \Sigma \left( I_R \otimes \hat{U}^{(n)} \right)^H x = y^H \Sigma y > 0
$$

(4.13)

where we defined the non-null vector $y = \left( I_R \otimes \hat{U}^{(n)} \right)^H x$.

However, observe that $\Sigma$ doesn’t represent the minimal set of parameters from which the covariance matrix on each sub-carrier functionally depend. To show that, let’s rewrite explicitly the product in equation 4.7 with respect to the block matrices composing $\Sigma$

$$
\text{Cov}(\eta_n) = \sum_{l=0}^{L-1} \sum_{p=0}^{L-1-l} e^{i2\pi \frac{lp}{N}} \Sigma_{lp}
$$

(4.14)

where $\Sigma_{lp}$ is an $R \times R$ matrix with entries $\Sigma_{lp}(r_1, r_2) = \Sigma(Rl + r_1, Rp + r_2)$.

Then substituting $p - l$ with $k$ we have:

$$
\text{Cov}(\eta_n) = \sum_{l=0}^{L-1} \sum_{k=-l}^{L-1-l} e^{i2\pi \frac{lk}{N}} \Sigma_{l,k+l} = \sum_{l=0}^{L-1} \sum_{k=-L-1}^{L-1} e^{i2\pi \frac{lk}{N}} \Sigma_{l,k+l} \chi(-l \leq k \leq L - 1 - l)
$$

(4.15)

where $\chi(prop)$ is the $\chi$ function, equal to one if the proposition $prop$ is true, equal to zero otherwise.
Chapter 4 Joint Semi-Blind Estimation of channel and noise covariance matrix

Now, since the second sum does not depend on \( l \) anymore, we can swap the two sums and, after reordering the terms we obtain:

\[
\text{Cov}(\eta_n) = \sum_{k=-L+1}^{L-1} e^{2\pi \frac{kn}{N}} \sum_{l=\max\{-k,0\}}^{L-1} \Sigma_{l,k+l}
\]

\[
= \sum_{l=0}^{L-1} \Sigma_{ll} + \sum_{k=1}^{L-1} \left( e^{2\pi \frac{kn}{N}} \sum_{l=0}^{L-1-k} \Sigma_{l,k+l} + e^{-i2\pi \frac{kn}{N}} \sum_{l=0}^{L-1-k} \Sigma_{l,k+l}^H \right)
\]

\[
= \Gamma_0 + \sum_{k=1}^{L-1} \left( e^{2\pi \frac{kn}{N}} \Gamma_k + e^{-i2\pi \frac{kn}{N}} \Gamma_k^H \right) \tag{4.16}
\]

where in the last equality we used the fact that \( \Sigma_{k+l,l} = \Sigma_{l,k+l}^H \) and we defined \( \Gamma_k = \sum_{l=0}^{L-1-k} \Sigma_{l,k+l} \), which correspond to the sum of the block matrices on the \( k \)th block-line parallel to the main block-diagonal of \( \Sigma \).

From the above parameterization of the covariance matrices, we see that they depend solely on the \( R \times R \) matrices \( \Gamma_k \quad \forall \quad k = 0 \ldots L - 1 \). In order to determine the total number of parameters describing the noise statistics, observe that \( \Gamma_0 \) is Hermitian, therefore it is parameterized by \( R^2 \) real elements, whereas \( \Gamma_k, k \neq 0 \) doesn’t have this property, therefore they are parameterized by \( 2R^2 \) real parameters (both imaginary and real part). In total there are \( (2L-1)R^2 \) real parameters.

It is now clear the reason why we modeled the noise at the receiver. Let’s assume that, instead of using such parameterization, the covariance on each sub-carrier is functionally independent across the sub-carriers. Then, being each covariance matrix on each sub-carrier parameterized by \( R^2 \) parameters, there are a total of \( NR^2 \) real elements parameterizing the covariance matrices on all sub-carriers. Therefore, since \( N > 2L - 1 \), and practically \( N \gg L \), a smaller number of parameters need to be estimated with the parameterization given above, which represents a potential for improving the estimation accuracy.

Observe however that this parameterization doesn’t necessarily fulfill the positive definite nature of \( \text{Cov}(\eta_n) \). In fact, for any \( x \in \mathbb{C}^{R \times 1} \) we have

\[
x^H \text{Cov}(\eta_n(k)) x = x^H \Gamma_0 x + \sum_{k=1}^{L-1} \left( e^{2\pi \frac{kn}{N}} x^H \Gamma_k x + e^{-i2\pi \frac{kn}{N}} x^H \Gamma_k^H x \right)
\]

\[
= \gamma_0(x) + 2 \sum_{k=1}^{L-1} \text{real} \left( e^{i2\pi \frac{kn}{N}} \gamma_k(x) \right) \tag{4.17}
\]

where we defined \( \gamma_k(x) = x^H \Gamma_k x, \quad k = 0 \ldots L - 1 \). We observe that just imposing that \( \Gamma_0 \) is positive definite, while letting full degree of freedom on \( \Gamma_k, k \neq 0 \), doesn’t assure that \( \text{Cov}(\eta_n(k)) \) is positive definite. Therefore, while equation (4.16) represents a
minimal parameterization of the covariance matrix on each sub-carrier, it doesn’t give control on the fact that Cov(η_n) is positive-definite.

Conversely, this is possible through the parameterization given by equation [4.7] since, as we have shown, the positive-definite constraint is assured for any positive-definite Σ. For this reason, in the next section, where we propose an algorithm for the ML estimation of the noise covariance matrices on each sub-carrier, we use this parameterization of the noise statistics.

4.2 Noise Covariance matrix Estimation

In this section we deal with the estimation of the noise covariance matrix Cov(η_n), under the parameterization given in section 4.1. The algorithm discussed here represents an extension to [12], where the author presents an algorithm for the estimation of Band-Toeplitz covariance matrices. In fact, in our estimation problem, we have a Band-Circular constraint, which becomes clear when considering the lag τ correlation of the noise samples in the time-domain, which is equal to zero for |τ| ≥ L, due to the channel length L:

\[
E [\tilde{\eta}_p \bar{\eta}_{p-\tau}] = \sum_u \sum_{l=0}^{L-1} \sigma_s^{(u)2} h_l^{(u)} h_{l-\tau} + \delta_{\tau0} \sigma_w^2 I_R \tag{4.18}
\]

The circularity of the covariance matrix structure derives from the fact that, due to the insertion of the Cyclic Prefix at the transmitters, a full period of the noise process is available at the receiver.

The extensions to the paper derive from the fact that the each correlation term is a matrix, rather than a scalar. Moreover, we present an alternative parameterization of the covariance matrices, which enforces the positive definite constraint proper of covariance matrices.

We saw that the covariance matrix on sub-carrier n can be expressed as a function of an LR × LR Hermitian positive-definite matrix Σ, through the relation (see equation [4.7]):

\[
\text{Cov} (\eta_n) = N \left( I_R \otimes \tilde{U}_N^{(n)} \right) \Sigma \left( I_R \otimes \tilde{U}_N^{(n)} \right)^H \tag{4.19}
\]

Let’s assume that we want to perform a Maximum Likelihood estimate of the covariance matrices, under the functional constraint defined by equation [4.19]. Since the covariance matrix on each sub-carrier is a function of Σ, the constrained Maximum Likelihood solution is obtained by maximizing the likelihood of the observations with respect to Σ.
(under the constraint that it is positive-semidefinite), from which the ML estimate of the covariance matrices is obtained through relation $4.19$

The ML solution of $\Sigma$ is necessarily solution to the likelihood equation, which is obtained by calculating the gradient of the negative log-likelihood function with respect to the unconstrained elements parameterizing $\Sigma$ (the real diagonal elements, the real and imaginary part of the upper-right triangle), and equaling this derivative to zero. Unfortunately, there is no closed form solution to this maximization problem. However, the gradient can be used in a Gradient Descent Algorithm to converge to a local minimum of the negative log-likelihood function. The problem with this approach consists in the fact that the further positive-definite constraint on $\Sigma$ is difficult to enforce.

In fact, let’s consider the update of matrix $\Sigma$ during the Gradient Descent iterations. We have

$$
\Sigma^{(k)} = \Sigma^{(k-1)} - \mu_k \Delta_k
$$

(4.20)

where $\Sigma^{(k)}$ is the estimate of matrix $\Sigma$ at the $k$th iteration of the gradient descent algorithm, $\mu_k > 0$ is the step-size, and $\Delta_k$ is the gradient of the cost function calculated in correspondence of $\Sigma^{(k-1)}$. Notice that, from the properties of positive-definite matrices, if $\Sigma^{(0)} > 0$ and $\Delta_k \leq 0 \ \forall k$, than $\Sigma^{(k)} \geq \Sigma^{(k-1)} \geq \cdots \geq \Sigma^{(0)} > 0$, which implies that $\Sigma^{(k)} > 0 \ \forall k$. However, this is an absurd since in this case the eigenvalues of $\Sigma^{(k)}$ would diverge to infinity for growing values of $k$. Therefore, the gradient $\Delta_k$ is not necessarily semidefinite negative, which demonstrates that, even if we start from an initial positive definite estimate of $\Sigma$, at the $k$th iteration of the EM-algorithm we might not have a positive definite solution.

The solution proposed here consists in parameterizing matrix $\Sigma$ in such a way that the positive definite constraint is always enforced. Observe that any Hermitian $N \times N$ matrix $P$ is positive semi-definite if and only if it can be decomposed into the product $AA^H$ for some $N \times N$ matrix $A$, and it is strictly positive definite if and only if $A$ is full-rank. In fact, letting $P = QDQ^H$ be the eigenvalue decomposition of $P$, with $Q$ unitary matrix and $D$ diagonal matrix, if $P \geq 0$, then the diagonal entries of $D$ are non negative, and we can write $P = Q\sqrt{D}Q^H = AA^H$ for $A = Q\sqrt{D}$. If $P$ is strictly positive definite, then necessarily $A$ is full rank. Similarly, for any $N \times N$ matrix $A$, given the non null vector $x \in \mathbb{C}^{N \times 1}$ we have $x^HA^Hx = y^Hy \geq 0$. Therefore $P \geq 0$. Moreover, if $A$ is full rank, then necessarily we have $P = AA^H > 0$. 
Therefore we have

\[
\begin{align*}
    P \geq 0 & \iff P = AA^H \quad \text{for some square matrix } A \\
    P > 0 & \iff P = AA^H \quad \text{for some square full-rank matrix } A
\end{align*}
\] (4.21)

Therefore, since \( \Sigma \) is a positive definite matrix of dimension \( LR \times LR \), it can equivalently be decomposed into \( \Sigma = RR^H \) for some full-rank \( LR \times LR \) matrix \( R \).

This suggests that, instead of minimizing the negative log-likelihood function with respect to the positive-definite matrix \( \Sigma \), it is possible to perform the minimization with respect to \( R \). The difference consists in the fact that, while the minimization of the negative log-likelihood function with respect to \( \Sigma \) is constrained on the fact that \( \Sigma \) is positive-definite, the minimization with respect to \( R \) is unconstrained, since for any \( R \), \( \Sigma \) is positive-(semi)definite. Therefore, using such parameterization of \( \Sigma \), we transform the constrained minimization problem into an unconstrained one.

Assuming this decomposition of \( \Sigma \), and considering only the pilot observations for now, the minimization of the negative log-likelihood function with respect to \( R \) leads to

\[
\hat{R} = \min_R \left\{ -\ln p \left( Y^{(tr)} | X^{(tr)}, h, R \right) \right\}
= \min_R \left\{ -\sum_n K_n^{(tr)} \ln \left( \frac{|B_{\eta_n}|}{\pi^{LR}} \right) + \sum_n \text{trace} \left( B_{\eta_n} S_n^{(tr)} \right) \right\}
\] (4.22)

There is no closed form solution to this minimization problem, however the gradient of the above cost function with respect to matrix \( R \) can be used in a Gradient Descent algorithm to determine a local minimum.

The derivative of the above cost-function with respect to the entries of matrix \( R^* \) is given by

\[
- \frac{\partial \ln p(Y^{(tr)} | X^{(tr)}, h, R)}{\partial R(z, t)^*} = \sum_n \text{trace} \left[ B_{\eta_n} \frac{\partial \text{Cov}(\eta_n)}{\partial R(z, t)^*} \left( K_n^{(tr)} I_R - B_{\eta_n} S_n^{(tr)} \right) \right]
\] (4.23)

Now, using 4.19 we have

\[
\frac{\partial \text{Cov}(\eta_n)}{\partial R(z, t)^*} = N \left( I_R \otimes \tilde{U}_N^{(n)} \right) R \delta(t, z) \left( I_R \otimes \tilde{U}_N^{(n)} \right)^H
\] (4.24)
and substituting this into (4.23) we obtain

\[
- \frac{\partial \ln p(Y^{(tr)}|X^{(tr)}, h, R)}{\partial R(z, t)^*} = N \sum_n \text{trace} \left[ B_{\eta_n} \left( I_R \otimes \tilde{U}_N^{(n)} \right) R \delta(t, z) \left( I_R \otimes \tilde{U}_N^{(n)} \right)^H \left( K_n^{(tr)} I_R - B_{\eta_n} S_n^{(tr)} \right) \right]
\]

\[
= N \sum_n \left[ \left( I_R \otimes \tilde{U}_N^{(n)} \right)^H \left( K_n^{(tr)} I_R - B_{\eta_n} S_n^{(tr)} \right) B_{\eta_n} \left( I_R \otimes \tilde{U}_N^{(n)} \right) \right] \]_{zt} \tag{4.25}

Reordering the elements on the gradient matrix \( \Delta_R \) we obtain:

\[
\Delta_R = N \sum_n \left( I_R \otimes \tilde{U}_N^{(n)} \right)^H \left( K_n^{(tr)} I_R - B_{\eta_n} S_n^{(tr)} \right) B_{\eta_n} \left( I_R \otimes \tilde{U}_N^{(n)} \right) \] \tag{4.26}

Finally, let:

\[
\mathcal{P} = N \sum_n \left( I_R \otimes \tilde{U}_N^{(n)} \right)^H \left( K_n^{(tr)} I_R - B_{\eta_n} S_n^{(tr)} \right) B_{\eta_n} \left( I_R \otimes \tilde{U}_N^{(n)} \right) \] \tag{4.27}

where we highlight the dependence of \( \mathcal{P} \) on \( \Sigma \) (since the covariance matrix on each sub-carrier, hence the precision matrix \( B_{\eta_n} \), are functions of \( \Sigma \)).

Then, we can rewrite the gradient \( \Delta_R \) as:

\[
\Delta_R = \mathcal{P} \Sigma \] \tag{4.28}

Now, using the Gradient Descent algorithm for determining a local minimum of the negative log-likelihood function, we have the following update at the \( k \)th iteration:

\[
R^{(k)} = R^{(k-1)} - \mu_k \Delta_R \] \tag{4.29}

where \( R^{(k)} \) is the estimate of matrix \( R \) at the \( k \)th iteration of the gradient descent algorithm, \( \mu_k > 0 \) is the step-size, and \( \Sigma^{(k-1)} = R^{(k-1)} R^{(k-1)H} \) is the estimate of \( \Sigma \) in the previous iteration.

This translates into the following update of matrix \( \Sigma \):

\[
\Sigma^{(k)} = R^{(k)} R^{(k)H} = \left[ I_{LR} - \mu_k \mathcal{P} \left( \Sigma^{(k-1)} \right) \right] \Sigma^{(k-1)} \left[ I_{LR} - \mu_k \mathcal{P} \left( \Sigma^{(k-1)} \right) \right] \] \tag{4.30}

where we used the fact that \( \mathcal{P} \) is an Hermitian matrix.

Finally, using (4.19) the update to the covariance matrix on each sub-carrier is given by:

\[
\text{Cov} \left( \eta_n \right)^{(k)} = N \left( I_R \otimes \tilde{U}_N^{(n)} \right) \Sigma^{(k)} \left( I_R \otimes \tilde{U}_N^{(n)} \right)^H \] \tag{4.31}
It is clear that, even if we are minimizing the negative log-likelihood function with respect to $\mathcal{R}$, it is not needed to explicitly calculate matrix $\mathcal{R}$, since the update of $\Sigma$ does not explicitly depend on the previous estimate of $\mathcal{R}$, but only on the previous estimate of $\Sigma$. This is important, since it is not required to calculate the decomposition of $\Sigma$, and we can directly update $\Sigma$ using (4.30) instead.

Observe that the update (4.30) is such that $\Sigma^{(k)}$ is always positive definite, as long as the initialization of the Gradient Descent Algorithm is a positive-definite matrix.

In fact, for any non-null vector $x \in \mathbb{C}^{LR}$ we have

$$
x^H \Sigma^{(k)} x = x^H \left[ I_{LR} - \mu_k \mathcal{P} \left( \Sigma^{(k-1)} \right) \right] \Sigma^{(k-1)} \left[ I_{LR} - \mu_k \mathcal{P} \left( \Sigma^{(k-1)} \right) \right] x
$$

where we defined $y = \left[ I_{LR} - \mu_k \mathcal{P} \left( \Sigma^{(k-1)} \right) \right] x$. Therefore, if the previous estimate $\Sigma^{(k-1)}$ is positive-definite, also the new estimate $\Sigma^{(k)}$ is positive definite (as long as $\left[ I_{LR} - \mu_k \mathcal{P} \left( \Sigma^{(k-1)} \right) \right]$ is full-rank, which is a plausible assumption; otherwise it is semidefinite-positive, but never negative-definite). By induction, if $\Sigma^{(0)} > 0$, also $\Sigma^{(k)} \forall k$ is positive definite.

Therefore, we need an initial positive-definite estimate of matrix $\Sigma$. This is easily accomplished by assuming that the noise-covariance matrix is the same on all sub-carriers. Then we have $\text{Cov}(\eta_n) = \text{Cov}(\eta) \forall n$.

Under this assumption, the ML estimate can be determined in closed form, and corresponds to the sample covariance matrix, averaged over the sub-carriers, that is:

$$
\text{Cov}(\eta) = \frac{1}{N_{tr}} \sum_n S_n^{(tr)}
$$

where $N_{tr} = \sum_n K_n^{(tr)}$ is the total number of pilots.

This corresponds to an initialization of $\Sigma$ given by:

$$
\Sigma^{(0)} = I_L \otimes \left( \frac{1}{L N_{tr}} \sum_n S_n^{(0)} \right)
$$

Observe that $\text{Cov}(\eta)$, as defined in (4.33), is a positive definite matrix, since it is a sum of positive-definite matrices ($S_n^{(tr)}$). For the same reason, also $\Sigma^{(0)}$ is positive-definite, therefore it represents a valid initialization of the Gradient Descent algorithm.

As we did for the training sequence channel estimator, it is convenient to include all the operations involved in the estimation of the positive-definite matrix $\Sigma$ through the
Gradient Descent algorithm into a Black-Box, that is a function $\mathcal{G}$, taking as input the terms $S_{n}^{(tr)}$, the number of symbols used for the estimate on each sub-carrier $K_{n}^{(tr)}$, and the initialization of the Gradient Descent Algorithm $\Sigma^{(0)}$, and returning the ML estimate of matrix $\Sigma$. Therefore we define

$$\hat{\Sigma} = \mathcal{G} \left( \{ (S_{n}^{(tr)}, K_{n}^{(tr)}) , n = 0 \ldots N - 1 \} , \Sigma^{(0)} \right)$$

(4.35)

Based on the GD algorithm described in this section, in the next section we derive an algorithm for the joint estimation of channel and noise covariance matrix.

### 4.3 Joint Semi-Blind Estimation of channel and noise covariance matrix

So far, we have discussed the estimation of the noise covariance matrix on each sub-carrier, assuming the channel is known at the receiver, under the functional constraint given by (4.7). We showed that there is no closed form solution to this problem, therefore we suggested to use the Gradient Descent Algorithm for the determination of a local minimum of the negative log-likelihood function.

Now, we discuss about the joint estimation of channel and noise covariance matrix. We start first of all by discussing the pilot based approach, since the Semi-Blind approach, discussed in section 4.3.2 represents a natural extension, as we will show.

#### 4.3.1 Pilot based approach

The likelihood of the pilot observations, conditioned on the channel $h$ and on $\Sigma$ is given by

$$-\ln p \left( Y^{(tr)} | X^{(tr)}, h, \Sigma \right) = \sum_{n} K_{n}^{(tr)} \ln \left( \pi^{R} | \text{Cov}(\eta_{n}) \right) + \sum_{n} \text{trace} \left( B_{\eta_{n}} S_{n}^{(tr)} \right)$$

(4.36)

We know from chapter 2 that the ML estimate of the channel, based solely on pilot observations, and conditioned on the noise covariance matrix on each sub-carrier is given by (2.20), which is the unique solution to the likelihood equation. In the previous section we studied a Gradient Descent algorithm for the estimation of the noise covariance matrix, assuming the channel matrix $h$ is known. When neither the covariance matrices nor the channel matrix are known at the receiver, a joint ML solution is obtained by minimizing jointly the negative log-likelihood function (4.36) with respect to $h$ and $\Sigma$. 
This can be performed either by minimizing iteratively with respect to one unknown while keeping fixed the other till convergence, or by jointly minimizing with respect to $h$ and $\Sigma$ together. To understand the difference between the two approaches, let’s imagine for simplicity a function defined on a two dimensional space, $f(x, y)$ with $(x, y) \in \mathbb{R}^2$. With the first approach the minimization is performed, starting from the point $(x_0, y_0)$, firstly with respect to $x$ while keeping fixed $y = y_0$, then with respect to $y$ while keeping fixed $x = x_1$, and so on, iterating between these two steps until convergence; with the second approach the minimization is performed directly on $\mathbb{R}^2$, moving along the direction in $\mathbb{R}^2$ of fastest decrease of the function. The second approach seems to be optimal from a convergence point of view, since the Gradient Descent algorithm moves on the highest dimensional space identified by all the parameters governing the system, whereas with the first approach the Gradient Descent algorithm moves along the sub-space identified by keeping fixed some of the parameters while moving the others. However, the advantage of the first approach resides on the fact that the minimization with respect to the channel matrix while keeping fixed $\Sigma$ can be computed in closed form, therefore there is no need to use the Gradient Descent algorithm when minimizing with respect to $h$. For this reason, we choose the second approach for determining the joint ML solution.

Therefore, starting from an initial channel estimate $h^{(0)}$ and an initial estimate $\Sigma^{(0)}$, the algorithm proceeds by reestimating the channel keeping fixed the current estimate of $\Sigma$, then reestimating $\Sigma$ while keeping fixed the current channel estimate, and so on until convergence.

We see that for the initialization of the algorithm we need $h^{(0)}$ and $\Sigma^{(0)}$. The problem consists in the fact that the channel estimate and the noise covariance estimate depends on each other. However, observe that the channel estimator studied in chapter 2 has a nice property: even if the channel is estimated using a value for the noise covariance matrix which is different from the true noise covariance matrix, it is an unbiased estimator (see section 2.1.2.1 for the derivation of this result).

This means that we can perform an initial channel estimate assuming white Gaussian noise at the receiver with a given variance, for example $\sigma_w^2 = 1$, using 2.20. This estimate, even if it suffers from an higher variance with respect to the case where the channel is estimated using the true noise covariance matrix, is still unbiased.

With this initial channel estimate $h^{(0)}$, it is then possible to produce an initial estimate of the noise covariance matrix, using the GD algorithm described in the previous section and summarized in function 4.35.
Finally, the minimization with respect to the channel and with respect to $\Sigma$ are repeated until convergence. Convergence of the algorithm is determined by evaluating after each iteration the cost function in correspondence of the current estimates of the channel and of the noise covariance matrix, and comparing it with the cost function calculated at the end of the previous iteration: if the new cost function differs from the previous one by less than a certain threshold, the algorithm is exited, otherwise another iteration is repeated, using the current channel estimate and noise covariance estimate as inputs.

### 4.3.2 Semi-Blind approach

In the previous section, we showed how to jointly estimate the channel and the noise covariance matrix on each sub-carrier, using only the pilot observations. Now, we want to improve the estimation accuracy by including also the blind observations into the estimate.

Similarly to the procedure used in the previous chapter when dealing with the Semi-Blind channel estimators, we use the EM-algorithm, since we can model the unknown data as hidden variables.

For now, we don’t make any prior assumption on the distribution of the unknown symbol, since we want to treat EM in its general form, as we did in section 3.1.2 in the case of Semi-Blind channel estimators, so that we can then apply this algorithm to the particular cases, such as the Gaussian assumption, the Constant Modulus assumption, or the true Discrete assumption for the unknown symbols. As we will see, the update of the channel matrix and of the noise covariance matrices during the $M$-step depend only on the first and second order moments of the unknown symbols, similarly to the results obtained in section 3.1.2.

To start with, let’s consider the log-likelihood of the observations (pilot plus blind) conditioned on the transmitted pilots, on the channel realization $h$, and on matrix $\Sigma$ which parameterizes the noise covariance matrix on each sub-carrier.

From the general introduction to the EM-algorithm in section 3.1.1, we have the following lower bound to the log-likelihood function:

$$
\ln p \left( Y | X^{(tr)}, h, \Sigma \right) \geq E_{q(V^{(bl)})} \left[ \ln \left( \frac{p(Y, V^{(bl)} | X^{(tr)}, h, \Sigma)}{q(V^{(bl)})} \right) \right] = \mathcal{F} \left( q \left( V^{(bl)} \right), h, \Sigma \right)
$$

(4.37)
for any distribution on the hidden variables $q(V^{(bl)})$, where the notation $E_{V^{(bl)}}(q)$ indicates that the expectation is taken with respect to the distribution $q(\cdot)$ on the hidden variables $V^{(bl)}$.

The maximization of $\mathcal{F}(q(V^{(bl)}), h^{(j)}, \Sigma^{(j)})$ with respect to the distribution of the unknown symbols $q(V^{(bl)})$ during the E-step, given the current estimate of the time-domain channel and of $\Sigma$ at the $j$th iteration of the EM-algorithm, $h^{(j)}$ and $\Sigma^{(j)}$, leads to the following result:

$$q^{(j)}(V^{(bl)}) = p\left( V^{(bl)} \left| Y^{(bl)}, h^{(j)}, \Sigma^{(j)} \right. \right)$$ (4.38)

During the M-step, the lower bound $\mathcal{F}(q(V^{(bl)}), h, \Sigma)$ is maximized with respect to the time-domain channel, $h$, and with respect to $\Sigma$, while keeping fixed the distribution on the unknown symbols $q(V^{(bl)})$. As we did in the previous section, instead of maximizing the lower bound jointly with respect to $h$ and $\Sigma$, we maximize it with respect to one variable while keeping fixed the other.

Using this approach, the $(j+1)$th update of the channel matrix, $h^{(j+1)}$, given $q^{(j)}(V^{(bl)})$ and $\Sigma^{(j)}$ is given by:

$$h^{(j+1)} = \max_h \left\{ \mathcal{F}(q^{(j)}(V^{(bl)}), h, \Sigma^{(j)}) \right\}$$

$$= \max_h \left\{ E_{V^{(bl)}}(q^{(j)}) \left[ \ln \left( \frac{p(Y, V^{(bl)}|X^{(tr)}, h, \Sigma^{(j)})}{q^{(j)}(V^{(bl)})} \right) \right] \right\}$$ (4.39)

This maximization problem was studied in section 3.1.2 when describing the EM-algorithm for determining the ML solution to the Semi-Blind channel estimation approach. In that circumstance we saw that, letting

$$\begin{align*}
\Lambda^{n,j}_{xx} &= E_{V^{(bl)}} \left[ X_n X_n^H \left| Y_n^{(bl)}, h^{(j)}, \Sigma^{(j)} \right. \right] \\
\Lambda^{n,j}_{yx} &= Y_n E_{V^{(bl)}} \left[ X_n^H \left| Y_n^{(bl)}, h^{(j)}, \Sigma^{(j)} \right. \right]
\end{align*}$$ (4.40)

the new channel estimate is given by 2.20, that is

$$h^{(j+1)} = \mathcal{H} \left( \Lambda^{n,j}_{xx}, \Lambda^{n,j}_{yx}, B^{(j)}_{\eta n}, n = 0 \ldots N - 1 \right)$$ (4.41)

The only difference with respect to the M-step of the Semi-Blind channel estimator studied in 3.1.2 resides in the fact that the channel is estimated using the current estimate of the noise precision matrices $B^{(j)}_{\eta n}$, instead of using the true noise covariance matrix.

As regards the update of the positive-definite matrix $\Sigma$, we use the same decomposition used in section 4.2 that is $\Sigma = RR^H$. The maximization of the lower bound is then
performed with respect to $\mathcal{R}$ rather than $\Sigma$, in order to enforce the positive-definite constraint. Therefore, the maximization of the lower bound with respect to $\mathcal{R}$, given the current estimate of the channel $h^{(j+1)}$ and the current distribution on the unknown symbols $q^{(j)}$, leads to the following result:

$$
\mathcal{R}^{(j+1)} = \max_{\mathcal{R}} \left\{ \mathcal{F} \left( q^{(j)} \left( V^{(bl)} \right), h^{(j+1)}, \Sigma \right) \right\} \\
= \max_{\mathcal{R}} \left\{ E_{V^{(bl)}}^{(q^{(j)})} \left[ \ln \left( \frac{p \left( Y, V^{(bl)} \big| X^{(tr)}, h^{(j+1)}, \Sigma \right)}{q^{(j)} \left( V^{(bl)} \right)} \right) \right] \right\} 
$$

and using the fact that $p \left( Y, V^{(bl)} \big| X^{(tr)}, h^{(j+1)}, \Sigma \right) = p \left( Y \big| X, h^{(j+1)}, \Sigma \right) p \left( V^{(bl)} \right)$, and that $p \left( V^{(bl)} \right)$ and $q \left( V^{(bl)} \right)$ are independent from $\mathcal{R}$, we obtain

$$
\mathcal{R}^{(j+1)} = \max_{\mathcal{R}} \left\{ E_{V^{(bl)}}^{(q^{(j)})} \left[ \ln p \left( Y \big| X, h^{(j+1)}, \Sigma \right) \right] \right\} = \\
= \min_{\mathcal{R}} \left\{ -K \sum_{n} \ln \left( \frac{|\mathcal{B}_{\eta_{n}}|}{\pi R} \right) + \sum_{n} \left( \mathcal{B}_{\eta_{n}} S_{n}^{(j)} \right) \right\} 
$$

where the $R \times R$ matrix $S_{n}^{(j)}$ is defined as:

$$
S_{n}^{(j)} = E_{V^{(bl)}}^{(q^{(j)})} \left[ \left( Y_{n} - H_{n}^{(j+1)} X_{n} \right) \left( Y_{n} - H_{n}^{(j+1)} X_{n} \right)^{H} \right] 
$$

This minimization problem was studied in section 4.2, and is equivalent to 4.22 as long as we set $K_{n}^{(tr)} = K$, $H_{n} = H_{n}^{(j+1)}$ and $S_{n}^{(tr)} = S_{n}^{(j)}$. We showed that there is no closed form solution, however the gradient of the cost function with respect to $\mathcal{R}$ can be used in a Gradient Descent algorithm to determine a local minimum of the negative log-likelihood function.

Using the function defined in 4.35, we can write $\Sigma^{(j+1)}$ as

$$
\Sigma^{(j+1)} = G \left( \left\{ \left( S_{n}^{(j)}, K \right), n = 0 \ldots N - 1 \right\}, \Sigma^{(j)} \right) 
$$

Notice that we use the previous estimate of $\Sigma$ as initialization of the Gradient Descent Algorithm. This is a valid initialization, as long as the whole EM-algorithm is initialized with a positive-definite matrix $\Sigma^{(0)}$. In fact, as we showed in section 4.2, function $G(\cdot)$ returns a positive-definite estimate of $\Sigma$, as long as the initialization of the GD algorithm is a positive-definite matrix. Then, if $\Sigma^{(0)}$ is positive definite, $\Sigma^{(1)}$, calculated using 4.45 is positive-definite, and so on up to the $j$th iteration, which returns a positive-definite solution.

Observe that, using the definitions of $\Lambda_{yx}^{(n,j)}$ and $\Lambda_{yx}^{(n,j)}$ in 4.40, $S_{n}^{(j)}$ can be rewritten as:

$$
S_{n}^{(j)} = Y_{n} Y_{n}^{H} + H_{n}^{(j+1)} \Lambda_{yx}^{(n,j)} H_{n}^{(j+1)H} - \Lambda_{yx}^{(n,j)} H_{n}^{(j+1)H} - H_{n}^{(j+1)} \Lambda_{yx}^{(n,j)H} 
$$
Finally, observe that for the calculation of $\Lambda^{(n,j)}_{xx}$ and $\Lambda^{(n,j)}_{yx}$ we need only the first and second order statistics of the unknown symbols with respect to the distribution $q^{(j)}$, which is equivalent to their posterior distribution. In fact, from 4.40 we have

$$\begin{align*}
\Lambda^{(n,j)}_{xx} &= X_n^{(tr)} X_n^{(tr)H} + C \Lambda^{(n,j)}_{xx} C^H \\
\Lambda^{(n,j)}_{yx} &= Y_n^{(tr)} X_n^{(tr)H} + Y_n^{(bl)} Y_n^{(bl)H} \\
&= \begin{cases} 
\Lambda^{(n,j)}_{xx} = X_n^{(tr)} X_n^{(tr)H} + C \Lambda^{(n,j)}_{xx} C^H \\
\Lambda^{(n,j)}_{yx} = Y_n^{(tr)} X_n^{(tr)H} + Y_n^{(bl)} Y_n^{(bl)H} 
\end{cases} (4.47)
\end{align*}$$

where we defined

$$\begin{align*}
\Lambda^{(n,j)}_{vv} &= E_{q^{(j)}} \left[ V_n^{(bl)} V_n^{(bl)H} \right] = E_{V^{(bl)}} \left[ V_n^{(bl)} V_n^{(bl)H} \right] Y_n^{(bl), h^{(j)}} \Sigma^{(j)} \\
\tilde{V}^{(bl)}_n &= E_{V^{(bl)}} \left[ V_n^{(bl)} \right] = E_{V^{(bl)}} \left[ V_n^{(bl)} \right] Y_n^{(bl), h^{(j)}} \Sigma^{(j)} (4.48)
\end{align*}$$

As regards the initialization of the algorithm, we use the same approach described in section 4.3.1 (equation 4.34). Therefore, we can perform an initial channel estimate based only on pilot observations, and assuming white Gaussian noise with variance $\sigma^2_w = 1$ (this estimate is statistically unbiased). We can then use this initial channel estimate to produce an initial estimate of matrix $\Sigma$ based solely on pilot observations, assuming as we did for the pilot based approach that the covariance matrix is the same on all sub-carriers, leading to the following result:

$$\begin{align*}
\Sigma^{(0)} &= I_L \otimes \left( \frac{1}{L N} \sum_n S_n^{(0)} \right) \\
B_{\eta n}^{(0)} &= \left( \frac{1}{N \sigma_w^2} \sum_n S_n^{(0)} \right)^{-1} (4.49)
\end{align*}$$

After this initialization phase, we can start with the Semi-Blind approach described here, by iteratively estimating the posterior first and second order moments of the unknown symbols, the channel and the noise covariance matrices, until convergence, which is determined by comparing the value of the cost function after each iteration of the algorithm. Then, the algorithm is assumed to have converged if the difference between the new cost function and the previous one is smaller than a given threshold $\lambda$.

We summarize here the main points of the EM-algorithm

1. Set $j = -1$, set the threshold $\lambda$

2. Perform an initial channel estimate using 2.20 and assuming white Gaussian noise with variance $\sigma^2_w = 1$:

$$h^{(0)} = H \left( \Lambda^{(n)}_{xx}, \Lambda^{(n)}_{yx}, B_{\eta n} = I_R, n = 0 \ldots N - 1 \right) (4.50)$$
where:

$$
\begin{align*}
\Lambda^{(n)}_{xx} &= X_n^{(tr)} X_n^{(tr)H} \\
\Lambda^{(n)}_{yx} &= Y_n^{(tr)} X_n^{(tr)H}
\end{align*}
$$

(4.51)

3. Perform an initial estimate of \( \Sigma \) and of the noise precision matrices on each sub-carrier using (4.49):

$$
\begin{align*}
\Sigma^{(0)} &= I_L \otimes \left( \frac{1}{L_{nr}} \sum_n S_n^{(0)} \right) \\
B_{\eta}^{(0)} &= \left( \frac{1}{N_{tr}} \sum_n S_n^{(0)} \right)^{-1} \quad \forall \ n = 0 \ldots N - 1
\end{align*}
$$

(4.52)

where:

$$
S_n^{(0)} = Y_n^{(tr)} Y_n^{(tr)H} + H_n^{(0)} \Lambda^{(n)}_{xx} H_n^{(0)H} - \Lambda^{(n)}_{yx} H_n^{(0)H} - H_n^{(0)} \Lambda^{(n)}_{yx} H
$$

(4.53)

4. \( j := j + 1 \)

5. **E-step**: calculate the posterior mean and second order moment of the unknown symbols, using the current estimate of the channel, \( h^{(j)} \), and the current estimate of matrix \( \Sigma \):

$$
\begin{align*}
\Lambda^{(n,j)}_{vv} &= E_{V^{(bl)}_n} \left[ V^{(bl)}_n V^{(bl)H}_n \bigg| Y_n^{(bl)} , h^{(j)} , \Sigma^{(j)} \right] \\
\tilde{V}_n^{(j)} &= E_{V^{(bl)}_n} \left[ V^{(bl)}_n \bigg| Y_n^{(bl)} , h^{(j)} , \Sigma^{(j)} \right] \\
\Lambda^{(n,j)}_{xx} &= X_n^{(tr)} X_n^{(tr)H} + C \Lambda^{(n,j)}_{yy} C^H \\
\Lambda^{(n,j)}_{yx} &= Y_n^{(tr)} X_n^{(tr)H} + Y_n^{(bl)} \tilde{V}_n^{(j)H} C^H
\end{align*}
$$

(4.54)

**M-step**: update the channel matrix as:

$$
h^{(j+1)} = \mathcal{H} \left( \Lambda^{(n,j)}_{xx} , \Lambda^{(n,j)}_{yx} , B_{\eta}^{(j)} , n = 0 \ldots N - 1 \right)
$$

(4.55)

**M-step**: Perform a new estimate of \( \Sigma \) using the current channel estimate \( h^{(j+1)} \) using (4.35) and of the noise precision matrices on each sub-carrier using (4.19):

$$
\begin{align*}
\Sigma^{(j+1)} &= \mathcal{G} \left( \left\{ S_n^{(j+1)} , K_n^{(tr)} , n = 0 \ldots N - 1 \right\} , \Sigma^{(j)} \right) \\
B_{\eta}^{(j+1)} &= \left[ N \left( I_R \otimes \tilde{U}^{(n)}_N \right) \Sigma^{(j+1)} \left( I_R \otimes \tilde{U}^{(n)}_N \right)^H \right]^{-1} \quad \forall \ n = 0 \ldots N - 1
\end{align*}
$$

(4.56)

where:

$$
S_n^{(j+1)} = Y_n^{(tr)} Y_n^{(tr)H} + H_n^{(j+1)} \Lambda^{(n,j)}_{xx} H_n^{(j+1)H} - \Lambda^{(n,j)}_{yx} H_n^{(j+1)H} - H_n^{(j+1)} \Lambda^{(n,j)}_{yx} H
$$

(4.57)
6. Calculate the new cost function $F(q^{(j)}, h^{(j+1)}, \Sigma^{(j+1)})$ and the difference between the new cost-function and the one calculated in the previous iteration, that is:

$$
\Delta^{(j)} = F(h^{(j+1)}, q^{(j)}, \Sigma^{(j+1)}) - F(h^{(j)}, q^{(j-1)}, \Sigma^{(j)})
$$

(4.58)

7. If $\Delta^{(j)} < \lambda$ the algorithm is assumed to have converged and is exited, otherwise another iteration is repeated (from step 4).

Once exited, the algorithm returns the current channel and noise covariance estimates, but also the posterior distribution of the unknown symbols, which can be used in the detection process.

The algorithm defined above can then be applied to any particular case. The choice of the assumption on the distribution of the unknown symbols determines how the posterior first and second order moments of the unknown symbols are calculated during the E-step. These were calculated in the previous chapter in sections 3.2 (true discrete distribution), 3.3 (Gaussian assumption for the unknown symbols) and 3.4 (Constant Modulus assumption), when dealing with Semi-Blind channel estimation. The only difference here consists in the fact that these posterior moments are calculated using the current estimate of the covariance matrix on each sub-carrier, rather than using the true covariance matrix.
Chapter 5

Simulation Results and Discussion

In this chapter we present and discuss some simulation results, and we compare the performance of the Semi-Blind and pilot based estimators described in the previous chapters, for different system setups. The simulations are performed on the LTE system, using the same pilot allocation criterion on the OFDM grid. Before proceeding with the discussion, we briefly describe the structure of the LTE physical frame, which is used for the simulations.

5.1 LTE frame structure

The LTE frame structure is depicted in figure 5.1.

![LTE frame structure diagram](image-url)
As you can see, LTE frames are 10ms in duration. They are divided into 10 sub-frames, each one 1.0ms long. Each sub-frame is further divided into two slots, each of 0.5ms duration.

In turn, each slot can be represented as a rectangular resource grid, of dimension $N \times K$, where $N$ is the number of sub-carriers used for transmission, which depends on the overall bandwidth of the system, and $K$ is the number of OFDM symbols composing each slot, which is equal to 7 in case of Normal Cyclic Prefix, which is the only configuration used in the simulations presented here (the other case is the Extended Cyclic Prefix, with 6 OFDM symbols per slot). If multiple antennas at the transmitter side are used, we can associate a resource grid to each transmitting antenna. The smallest unit composing the resource grid is the resource element, which is identified by two coordinates, sub-carrier number and OFDM symbol number. This corresponds to the signal transmitted by a specific transmitting antenna, on a specific sub-carrier and time. At an higher level, there are the resource blocks (RBs), defined as a grouping of 12 consecutive sub-carriers for the duration of one slot. Finally, a grouping of RBs along the frequency dimension defines one slot.

Figure 5.2: Pilot allocation on one resource block (12 sub-carriers times 7 OFDM symbols) for the cases 1, 2 and 4 transmitting antennas
For the channel estimation task, special reference signals (pilot symbols known at the receiver) are embedded on each resource block. The pattern depends on the number of transmitting antennas, and is depicted in figure 5.2 for the three cases $T = 1$, $T = 2$ and $T = 4$. This pilot allocation criterion will be used also in the simulations.

### 5.2 Simulation setup

In this section, we describe the common simulation parameters used, that is, how the unknown symbols, the pilot sequence, the channel are generated, and the methodology used for performing the simulations.

- **Pilot sequence generation**: the pilots are generated as a random QPSK sequence, and allocated on the OFDM grid according to figure 5.2 depending on the number of transmitting antennas used.

- **Unknown symbols**: the unknown symbols are drawn uniformly from a $M$-QAM constellation, with $M$ in the set $\{4, 16, 64\}$, independently across the sub-carriers and across time. On each sub-carrier, these symbols are mapped into $S$ streams ($S$ is the transmission rank, already used in the previous chapters), which in turn are mapped into the transmitting antennas through the $T \times S$ encoding matrix $C$, whose columns are drawn from an Hadamard sequence, with the property that $C^H C = I_S$. The average transmission power on each sub-carrier is 1, equally distributed across the transmitting antennas. Therefore, the mean power of the $M$-QAM symbols is $\sigma_s^2 = \frac{1}{S}$.

- **Channel**: in the simulations the channel length is known at the receiver and is given by $L = CP + 1$, where $CP$ is the Cyclic Prefix length. This is the maximum channel length supported by the system without generating Inter Symbol Interference. The channel between each transmitting-receiving antenna pairs is generated using the Rayleigh model, with exponential power delay profile, and average unit energy. However, we don’t use this prior knowledge in the estimation process, since we assume the channel is a deterministic unknown.

- **Noise**: at the receiver we assume zero mean Gaussian noise, independent across sub-carriers and across time. The covariance matrix on each sub-carrier is generated according to the model introduced in section 4.1. The SNR of the system is calculated as the ratio between the average transmission power per sub-carrier (which is normalized to 1 as explained in the item *Unknown symbols*), and the
average noise power per sub-carrier per receiving antenna, therefore, using the dB scale, this is defined as

\[
\text{SNR}_{\text{dB}} = -10 \log_{10} \left( \frac{\sum_n \text{trace} \left( \text{Cov} \left( \eta_n \right) \right)}{R_N} \right)
\]

(5.1)

• Each simulation consists of a number of iterations (usually 100, if not otherwise specified). At the beginning of each iteration, a new sequence of unknown symbols and a new MIMO channel are randomly generated, using the model explained above.

5.3 Comparison of Semi-Blind and pilot based approaches for different antenna setups

In this section, we compare the pilot based approach with the Semi-Blind approaches studied in chapter 3, in terms of mean square error of the estimator, and raw bit error probability. For the calculation of the raw Bit Error Probability (BER), an MMSE detector is employed, using the current channel estimate in the detection process.

We compare the performance of the estimators for different antennas setups, namely \(1T \times 1R, 1T \times 2R, 2T \times 1R\) and \(2T \times 2R\) (with the notation \(xT \times yR\) we mean \(x\) transmitting and \(y\) receiving antennas are employed). For all these cases we assume rank one transmission (\(S = 1\)), and only for the setup \(2T \times 2R\) we simulate also transmission rank 2. For the constellation order, we use 4-QAM, so that also the Constant Modulus assumption, which cannot be applied to non constant modulus constellation like 16 or 64-QAM, can be compared.

The common simulation setup used on each scenario consists of \(N = 72\) frequency sub-carriers, which corresponds to 6 resource blocks; the Cyclic prefix is \(CP = 8\), therefore the channel length used is \(L = 9\). One only LTE time slot (7 OFDM symbols) is transmitted and used for the estimate and for calculating the Bit Error Probability (BER). The SNR is let vary between -9dB and 21dB, with steps of 3dB. The random sequences (for generating the channels and for the unknown symbols), are generated using a common seed, so that the simulation results associated to different scenarios are comparable.

In the figures, the solid blue curves represent the MSE or BER of the pilot based approach. The solid red curves are associated to the Semi-Blind approach with the Gaussian approximation for the unknown symbols, whereas the dashed dotted line is the unbiased CRLB for this approach, calculated in Appendix C.3. The green curves
represent the MSE and BER of the Semi-Blind approach with the Constant Modulus assumption for the unknown symbols, whereas the magenta curves with circles are associated to the Semi-Blind approach using the true discrete distribution. The black curves are associated to the *Hard decision feedback* estimator, which was not treated in the thesis. This is a *brute force* estimator, which uses the feedback from the decoder (the decoded symbols) as a pilot sequence: after an initialization of the channel using only the pilot sequence, the two stages process decoding-channel estimation is iterated, feeding the decoded symbols into the channel estimator. This is repeated for a number of iterations (in the simulations we chose 5 iterations). Finally, the dash-dotted blue curve is the unbiased CRLB calculated assuming all the symbols are known at the receiver (all the symbols are pilot), therefore it represents a lower bound to the performance of any Semi-Blind estimation approach.

As regards the BER figure, the first subplot represents the BER associated to the channel estimators, normalized to the BER calculated using the true channel. Therefore, a point in the curve at coordinates \( (SNR = 0, \text{normBER} = 1) \), means that the BER at zero dB is 1.2 times the BER calculated using the true channel. This latter case is plotted in the second subplot, and can be used as a reference (black solid curve with circles). The reason for this choices resides in the fact that the typical representation doesn’t allow a clear comparison of the estimation approaches from a BER perspective.

### 5.3.1 \( 1T \times 1R \) MIMO

We start by considering the case of a simple SISO system \( (T = 1 \text{ and } R = 1) \). Figures 5.3 and 5.4 plot respectively the MSE and the BER of the estimators.

It is clear from the figures that all the Semi-Blind approaches lead to an improvement from both an MSE and a BER perspective. Taking as a reference the 0dB axis, we see that the estimation accuracy achieved by the Semi-Blind estimators in that point is achieved with the pilot based approach at an SNR 4-5dB higher, which is a 5dB improvement. In the next section, when dealing with higher order MIMO systems, we will see that the improvement is even bigger.

Observe that in the SNR range below 0dB the three Semi-Blind estimators perform almost identically, from both an MSE and BER perspective. Conversely, in the high-SNR regime their performance diverges, in particular using the true discrete distribution outperforms the other estimators, based on the CM and Gaussian assumption. The reason resides in the fact that when the noise level is high compared to the signal level, the observations are very noisy, and provide less evidence on the unknown symbols. Therefore, the distribution of the unknown symbols is less relevant in the estimation
process. Moreover, as we anticipated in section 3.3, when the noise level is high, the true distribution of the observations is well approximated with a Gaussian distribution. Conversely, in the high SNR regime, the observations carry mostly information about the transmitted symbols, therefore the prior distribution of the unknown symbols has a greater influence.

![Comparison of pilot based and Semi-Blind approaches (MSE)](image)

**Figure 5.3:** Comparison of pilot based and Semi-Blind approaches (MSE), $1T \times 1R$ MIMO-OFDM, 4-QAM, 72 sub-carriers

From this figure we observe a general pattern, which is also followed in the next simulation results we will present: the closer the assumptions on the distribution of the unknown symbols are to the true discrete distribution, the better is the performance achieved by the Semi-Blind estimator. In fact, using the true distribution leads to the best results (magenta curve with circles). Comparing the Gaussian and the CM approximations, the Gaussian assumption allows two degrees of freedom on the unknown symbols, amplitude (Rayleigh distributed) and phase (uniformly distributed in $[0, 2\pi]$), whereas the CM assumption allows one only degree of freedom, the phase, since the amplitude is fixed and known. Therefore, we can argue that the CM assumption is closer to the true discrete distribution than the Gaussian assumption is. This pattern is clear in the figures, where we observe that the CM assumption leads to a better performance with respect to the Gaussian assumption, from both a BER and MSE perspective. Unfortunately, the CM algorithm we developed in this thesis has scarce applicability, since it is limited to 4-QAM and transmission rank $S = 1$. 
Chapter 5 Simulation Results and Discussion

Through a careful inspection of the BER plot, the BER associated to the Semi-Blind approach with the Gaussian approximation leads to almost 0.5dB improvement with respect to the BER calculated using the pilot based channel estimate. Moreover, observe that the BER curves follow the same pattern as the MSE curve, that is we can observe that a better estimator from an MSE perspective leads to a lower BER. This is a general pattern, which can be observed also in the following simulation results.

5.3.2 $1T \times 2R$ MIMO

An interesting scenario is the $1T \times 2R$ MIMO-OFDM. Again, we plot the MSE and BER plots, with the same notation used for the previous scenario. Moreover, in the MSE plot we add the curves from the previous SISO scenario for comparison (dashed dotted curves with stars, the colors association is the same used before).

As regards the MSE of the estimators (figure 5.5), we notice that the pilot based approach doesn’t lead to any improvement with respect to the SISO case. This was demonstrated in chapter 2 where, for the case of white Gaussian noise and orthogonal pilots, we showed that the variance of the pilot based estimator, averaged over the number of channel entries, is independent of the number of receiving antennas. Conversely, the Semi-Blind approaches lead to a significant improvement with respect to the SISO scenario.
We have intuitively explained the reason in chapter 3 in the introduction to section 3.3 for the case of white Gaussian noise and flat-fading channel, where we showed that the Semi-Blind approach can potentially lead to an improvement in the estimation accuracy of a factor $\frac{2R}{T}$, which is proportional to the number of receiving antennas ([7]).

Alternatively, we can explain it by observing that, augmenting the number of receiving antennas, while keeping fixed the number of transmitting antennas and the transmission rank, more observations are available at the receiver, providing more evidence on the unknown symbols. This redundancy on the observations can be effectively exploited to enhance the estimation accuracy.

![Graph showing simulation results and discussion](image)

**Figure 5.5:** Comparison of pilot based and Semi-Blind approaches (MSE), $1T \times 2R$ MIMO-OFDM, 4-QAM, 72 sub-carriers

Also from a point of view of the BER (figure 5.6), a good improvement is clear with respect to the SISO scenario. In fact, the BER associated to the Semi-Blind estimators is very close to the BER assuming perfect knowledge of the channel, with about 1.2dB improvement with respect to the BER associated to the pilot based approach, using the Gaussian assumption. Observe that in the SNR range around 20dB the Semi-Blind estimators achieve a better BER with respect to using the true channel in the detection process. This fact is not expected. However, this can be easily explained: this simulation consisted of 100 iterations; during each iteration 1 LTE slot was transmitted with 72 sub-carriers, corresponding to $72 \times 7$ symbols (pilots+unknown symbols); taking into account that for each resource block 4 symbols are used for the pilots (see 5.2), and that
each unknown symbol carries 2 bits, on each iteration a total of 960 bits are transmitted. Therefore, during the whole simulation a total of 96000 bits are generated. Now, the BER at 18dB is around $15 \cdot 10^{-5}$, which means that a total of 15 bits are affected by errors. Similarly, at 21dB, where the BER is $2 \cdot 10^{-5}$, the number of bits affected by errors is 2. These numbers are statistically irrelevant, therefore they don’t provide a reliable estimate of the BER.

![Graph](image1.png)

**Figure 5.6:** Comparison of pilot based and Semi-Blind approaches (BER), $1T \times 2R$ MIMO-OFDM, 4-QAM, 72 sub-carriers

**5.3.3 $2T \times 1R$ MIMO**

Another interesting case is $2T \times 1R$ MIMO. In this scenario, the transmission rank is $S = 1$, therefore the unknown symbols, before being transmitted across the antenna array, are encoded through the encoding matrix $C$. For this case, we have two perspectives of the channel: one is the channel between the transmitting-receiving antennas arrays, which we name physical channel and on sub-carrier $n$ is given by $H_n$, the other is the channel between the unknown symbols (before the encoding process) and the receiving antennas, which we name equivalent channel and is the one effectively used in the decoding process, given by $H_n C$ on sub-carrier $n$. The performance of the estimator is therefore measured for both the *physical* and the *equivalent* channels.
Figure 5.7: Comparison of pilot based and Semi-Blind approaches (MSE), $2T \times 1R$ MIMO-OFDM, 4-QAM, 72 sub-carriers

Figure 5.8: Comparison of pilot based and Semi-Blind approaches (MSE), equivalent channel, $2T \times 1R$ MIMO-OFDM, 4-QAM, 72 sub-carriers
In the MSE of the physical channel (figure 5.7) we don’t observe the improvement we got in the previous scenarios. However, if we consider the equivalent channel (figure 5.8), we observe that the MSE is very close to the curves for the SISO scenario (dashed dotted curves with stars). The reason resides in the fact that what we observe between the unknown symbols and the receiver before the encoding process is actually a SISO channel. Moreover, this channel has the same properties as the SISO physical channel. In fact, on each sub-carrier we have:

\[ H_{n}^{(eq)} = H_{n}C = \sum_{l=0}^{L-1} (h_{l}^C) e^{-j2\pi ln/N} \]  

from which it is clear that the equivalent channel \( H^{(eq)} \) is also FIR of length \( L \).

Also from the point of view of the BER, we observe the same performance as the SISO system. The advantage of using two transmitting antennas resides in the fact that, since at the receiver we have two independent realizations of the fading process, the probability of the equivalent channel being in a deep fade is reduced, with respect to the SISO scenario.

![Figure 5.9: Comparison of pilot based and Semi-Blind approaches (BER), 2T × 1R MIMO-OFDM, 4-QAM, 72 sub-carriers](image)
Chapter 5 Simulation Results and Discussion

5.3.4 \(2T \times 2R\) MIMO, transmission rank \(S = 1\)

For the \(2T \times 2R\) MIMO-OFDM setup, we consider the two cases of transmission rank 1 and 2. The first case we consider is \(S = 1\), which means that one information stream is encoded across two antennas.

For the MSE performance, we compare it with the \(1T \times 2R\) MIMO setup. In fact, the equivalent channel can be viewed as a \(1T \times 2R\) SIMO, as we did for the case \(2T \times 1R\), where in that circumstance we compared the equivalent channel with the SISO channel.

For this case we consider only the MSE calculated on the equivalent channel, since the MSE for the physical channel doesn’t highlight the improvement in the estimation accuracy achievable with the Semi-Blind approaches.

![Figure 5.10: Comparison of pilot based and Semi-Blind approaches (MSE), \(2T \times 2R\) MIMO-OFDM, transmission rank 1, 4-QAM, 72 sub-carriers](image)

As we can see from figure 5.10, the MSE of the equivalent channel is very close to the MSE for the \(1T \times 2R\) SIMO scenario described above (dashed dotted curves with stars). The reason is that the equivalent channel behaves like a \(1T \times 2R\) SIMO, and is also FIR of length \(L\).

Observe that, although doubling the number of transmitting antennas, the accuracy of the pilot based estimator is the same as for the \(1T \times 2R\) SIMO scenario. The reason resides in the fact that, in a two transmitting antennas setup, a double number of pilots
is transmitted in the OFDM grid compared to one transmitting antenna setup, as we can see from figure 5.2. Therefore LTE-MIMO systems with higher number of antennas are less bandwidth efficient.

Also from a BER perspective (figure 5.11), we notice the same performance as the $1T \times 2R$ SIMO scenario. Again, the advantage of using two transmitting antennas (transmit diversity) is that the probability of the equivalent channel being in a deep fade is reduced.

Figure 5.11: Comparison of pilot based and Semi-Blind approaches (BER), $2T \times 2R$ MIMO-OFDM, transmission rank 1, 4-QAM, 72 sub-carriers
5.3.5 \( 2T \times 2R \) MIMO, transmission rank \( S = 2 \)

We now consider the second case of rank two transmission, which corresponds to full-rank transmission, the rank of the channel matrix.

Again, the same observations we did in the previous scenarios can be done here for the MSE (figure 5.12) and the BER (figure 5.13). In particular, the MSE is compared with the MSE for the SISO scenario.

Observe that the MSE of the pilot based approach is the same as the one calculated in the SISO scenario. Again, the reason is that the variance of the pilot based estimator is directly proportional to the number of transmitting antennas, and inversely proportional to the number of pilots. Since in a two transmitting antennas system the number of pilots gets doubled with respect to SISO, the estimation accuracy is the same (figure 5.2).

![Figure 5.12: Comparison of pilot based and Semi-Blind approaches (MSE), \( 2T \times 2R \) MIMO-OFDM, transmission rank 2, 4-QAM, 72 sub-carriers](image)

However, for all the Semi-Blind approaches (the curve for the Semi-Blind approach using the CM assumption is not plotted, since the transmission rank \( S > 1 \)) we notice a slightly higher MSE. This can be intuitively explained by observing that the pilot sequence is such that on a specific sub-carrier at a given time, only one antenna transmits a pilot symbol. In this way, the SIMO channel between a specific antenna and the
receiving antenna array can be effectively estimated without interference from the other transmitting antennas. On the contrary, the unknown symbols on a specific sub-carrier at a given time are transmitted at the same time through the antennas array. Therefore, the estimation of each SIMO channel is affected by interference across the transmitting antenna arrays. In other words, while the pilots are orthogonal across the transmitting antennas array, the unknown symbols are not, thus providing inter-antenna interference in the estimation process. Obviously, this problem is not present in a SISO system.

![Comparison of pilot based and Semi-Blind approaches (BER), $2T \times 2R$ MIMO-OFDM, transmission rank 2, 4-QAM, 72 sub-carriers](image)

Figure 5.13: Comparison of pilot based and Semi-Blind approaches (BER), $2T \times 2R$ MIMO-OFDM, transmission rank 2, 4-QAM, 72 sub-carriers
5.4 Estimation accuracy as a function of the sub-carriers

We now study the performance of the channel estimators for MIMO-OFDM systems with different number of sub-carriers, to understand how the performance of Semi-Blind estimators scales to higher order OFDM systems. We do this comparison for a representative case, $1T \times 2R$ MIMO-OFDM, with 4-QAM as modulation format, in order to allow the use of the CM algorithm. The number of sub-carriers compared are $N = 24$ (2 resource blocks), $N = 72$ (6 RBs) and $N = 144$ (12 RBs). The channel length is chosen in such a way that the ratio $\frac{L}{N}$ is a constant. In fact, we know from chapter 2 that the variance of the pilot based estimator is proportionally to this factor, in the case of white Gaussian noise and orthogonal pilots. Therefore, we choose $L = 3$ for $N = 24$, $L = 9$ for $N = 72$, and $L = 18$ for $N = 144.$

From the point of view of the MSE (figure 5.14), we observe that the estimators perform almost identically, independently of the number of sub-carriers used. The reason is that a larger number of sub-carriers, and a proportionally longer channel and larger number of channel parameters to be estimated, are counteracted by a proportionally larger number of pilot and blind observations. Therefore, even if the number of unknown parameters modeling the system (the $2LRT$ real entries of the time-domain channel matrix) grows, the amount of information used in the estimation process grows proportionally.

![Figure 5.14: Comparison of pilot based and Semi-Blind approaches for different number of sub-carriers (MSE). $1T \times 2R$ MIMO-OFDM, 4-QAM](image)
Similarly from the point of the BER we observe the estimators achieve the same performance independently from the number of sub-carriers (figure 5.15).

![Comparison of pilot based and Semi-Blind approaches for different number of sub-carriers (BER), $1T \times 2R$ MIMO-OFDM, 4-QAM](image)

**Figure 5.15:** Comparison of pilot based and Semi-Blind approaches for different number of sub-carriers (BER), $1T \times 2R$ MIMO-OFDM, 4-QAM

### 5.5 Estimation accuracy as a function of the constellation order

Finally, we compare the estimation accuracy of the estimators as a function of the constellation order (4, 16 and 64-QAM), from a point of view of the MSE performance. Again, we consider one representative case, $1T \times 2R$ MIMO-OFDM, with $N=72$ sub-carriers and channel length $L=9$.

The MSE of the estimators is represented in figure 5.16. Notice that the CM approach is not plotted, since it cannot be applied to constellation orders bigger than 4-QAM. Instead, the green curve now is associated to the pilot based approach, whereas the blue curve is associated to the Semi-Blind approach using the true discrete assumption of the unknown symbols. The dashed curves are associated to the 4-QAM case, the curves with the circles and with the stars to 16 and 64 QAM respectively.
Notice that the accuracy of the pilot based approach doesn’t depend on the constellation order. This is obvious since this approach relies solely on the pilot sequence, which is always drawn from a QPSK constellation. Moreover, the accuracy of the Semi-Blind estimator relying on the Gaussian assumption for the unknown symbols is identical, independently of the constellation order. This derives from the fact that this estimator completely discards the discrete nature of the unknown symbols. Conversely, the performance of the Semi-Blind approach using the true discrete distribution gets worse the higher is the constellation order. The reason resides in the fact that the higher is the constellation order, the more uncertainty and degrees of freedom there are on the unknown symbols, which translates into a lower estimation accuracy.

![Figure 5.16: Comparison of pilot based and Semi-Blind approaches for different constellation orders (MSE), $1T \times 2R$ MIMO-OFDM, 72 sub-carriers](image)

Finally, observe that the higher is the constellation order, the more the Semi-Blind estimator using the Gaussian assumption of the unknown symbols approaches the estimator using the true discrete distribution. This is a consequence of the fact that the Gaussian approximation is the more valid the higher is the constellation order, as we explained in the introduction to section 3.3.
5.6 Convergence of the EM-Algorithm, Gaussian approximation

In this section we plot the convergence of the EM-algorithm, both from an MSE and a BER perspective. That is, instead of measuring the convergence of the lower bound to the likelihood function, which is actually the cost function used in the EM-algorithm, we measure the evolution of the MSE and of the BER over the number of iterations. However, this is calculated only for the Semi-Blind approach using the Gaussian assumption for the unknown symbols, which is the most representative case. This is plotted for some representative MIMO setups, since the different scenarios discussed above follow a similar pattern from a convergence point of view.

Figures 5.17, 5.19 and 5.18 represent the evolution of the MSE and BER for the MIMO setups $1T \times 1R$, $1T \times 2R$ and $2T \times 2R$ respectively, for different values of the SNR. To make comparable the MSE and BER at different SNRs, the curves are normalized to the value of the MSE and of the BER calculated with the pilot based estimate. The EM-algorithm is initialized using only the pilot sequence, as we explained in the general treatment in section 3.1.2. We observe that it takes around 5 to 10 iterations for the algorithm to converge to a stable point. This pattern is also followed by the BER plot.

Figure 5.17: Evolution of MSE and BER over the iterations of the EM-algorithm, $1T \times 1R$ MIMO-OFDM, 4-QAM, 72 sub-carriers
Figure 5.18: Evolution of MSE and BER over the iterations of the EM-algorithm, $2T \times 2R$ MIMO-OFDM, transmission rank $S = 2$, 4-QAM, 72 sub-carriers

Figure 5.19: Evolution of MSE and BER over the iterations of the EM-algorithm, $1T \times 2R$ MIMO-OFDM, 4-QAM, 72 sub-carriers
5.7 Joint Estimation of Channel and noise covariance matrix

In this section, we present some results on the joint estimation of the channel and the noise covariance matrix. The algorithm used is described in chapter 4. As in the previous section, only the Semi-Blind channel estimator using the Gaussian assumption for the unknown symbols is considered here.

![Figure 5.20: Joint Estimation of channel and noise covariance matrix, MSE of channel estimator, $2T \times 2R$ MIMO-OFDM, transmission rank 2, 4-QAM, 72 sub-carriers](image)

Let’s consider the MIMO-OFDM setup $2T \times 2R$, transmission rank 2. Figures 5.20 and 5.21 represent respectively the MSE of the channel estimator and the BER. The solid curves with circles are associated to the joint estimate of channel and noise covariance matrix, whereas the blue curves without circles are associated to the channel estimators using the true covariance matrix. The BER for the joint estimate is calculated using the estimated covariance matrix in the detection process, therefore it takes into account the uncertainty on both the channel and the covariance matrix.

We observe that the non perfect knowledge of the covariance matrix represents a performance loss for the estimators, both in the pilot based and in the Semi-Blind approach. However, the estimation accuracy achieved by the Semi-Blind estimator is still better
than the accuracy of the pilot based estimator assuming perfect knowledge of the noise covariance matrix.

The same pattern can be observed from the point of view of the BER. For this case however, we observe that, while the performance loss incurred by the pilot based estimator in terms of MSE is relatively small, the loss in terms of BER is much bigger. The reason can be explained by observing that the pilot based estimator is very robust, even with a non perfect knowledge of the covariance matrices. In fact, in this case, although affected from an higher variance compared to the case of perfect knowledge of the noise covariance matrix, it is an unbiased estimator, as we demonstrated in section 2.1.2.1. Conversely, the estimate of the noise covariance matrices is negatively affected by the uncertainties in the estimation of the channel entries, and suffers from the lack of enough information to estimate a relatively large number of parameters \((2L - 1)R^2\) real parameters, as we showed in section 4.1. This in turn negatively affects the performance of the MMSE detector, hence the performance loss. On the contrary, the Semi-Blind estimator achieves a good performance also from the BER perspective, being close to the case of perfect knowledge of the noise covariance matrices. For this case in fact, the channel and noise covariance matrix estimates benefit from the availability of a larger number of observations, which are exploited to enhance the estimation accuracy.

![Figure 5.21: Joint Estimation of channel and noise covariance matrix (BER), 2T×1R MIMO-OFDM, 4-QAM, 72 sub-carriers](image-url)
Chapter 6

Conclusion

In this thesis we investigated the Semi-Blind approach to channel estimation in a MIMO-OFDM system, and in particular for LTE downlink. In a MIMO system the number of channel parameters is much larger than in a simple SISO, making the channel estimation task particularly critical. This derives from the fact that the MIMO channel can be represented as a set of $RT$ SISO channels, one between each transmitting-receiving antenna pair.

In chapter 2 we saw that, for a given number of pilots allocated on the OFDM grid, the increase in the number of channel parameters translates into a smaller estimation accuracy. Therefore in a MIMO-OFDM system, in order to achieve an acceptable estimation accuracy, more pilot symbols have to be allocated on the OFDM grid compared to a SISO system, thus compromising the bandwidth efficiency.

It is thus clear that in order to enhance the channel estimation accuracy, other approaches exploiting more information at the receiver have to be used.

The Semi-Blind approach studied in this thesis represents a band-efficient solution to channel estimation in MIMO-OFDM systems, since the estimation accuracy is improved by exploiting all the available information at the receiver, pilot symbols plus information symbols, rather than relying solely on the observations corresponding to pilot symbols. We saw that the estimation accuracy depends on the assumption used on the unknown symbols. In the thesis in particular (chapter 3), we considered three assumptions: the true discrete distribution of the unknown symbols, the Gaussian assumption, where the unknown symbols are assumed to be circular Gaussian distributed, and the Constant Modulus assumption, where the unknown symbols are assumed to have constant amplitude and uniform phase. Through simulations, we showed that a greater accuracy is achievable by using the true discrete distribution of the unknown symbols. More in
general, the more closely the assumption on the distribution of the unknown symbols is to the true discrete distribution, the greater is the estimation accuracy in terms of Mean Square Error of the estimate.

However, the increased estimation accuracy achievable with the Semi-Blind approach doesn’t come at no cost. We saw that the Maximum Likelihood estimator has a closed form solution in the case of pilot based estimation, and can be computed with a simple and robust algorithm. This is not the case for the Semi-Blind estimators studied in the thesis, since in general there is no closed form solution to the Likelihood equation (as demonstrated in section 3.1 for the general case). A solution is determined with the use of iterative algorithms, which provide a local solution to the maximization problem.

In general, we saw that the Expectation-Maximization algorithm is an useful framework for treating the Semi-Blind approach, since the unknown symbols can be thought as hidden variables. We showed in 3.1 that this algorithm requires only the computation of the posterior first and second order statistics of the unknown symbols during the Expectation step. The Maximization step can be computed with the same algorithm used to perform the pilot based channel estimation. Therefore, the complexity of this algorithm is determined by the calculation of the posterior first and second order statistics of the unknown symbols, and by the convergence properties (number of iterations required to converge to a local maximum). We saw that, using the true discrete distribution of the unknown symbols is the approach leading to the best performance in terms of Mean Square Error of the estimator, however it is also the most computationally demanding, since it requires the computation of a posterior discrete distribution, which is a combinatorial problem.

To conclude, it is clear that there is a trade-off between estimation accuracy and complexity: the minimum complexity is achieved with the pilot based approach, since the channel estimate can be performed in one single iteration (one shot); however, this approach is also the one achieving the minimum estimation accuracy, since only a small part of the observations is used for the estimate, the other is discarded. On the other hand, the best estimation accuracy is achieved with the Semi-Blind approach using the true discrete distribution of the unknown symbols. However, this is also the most computationally demanding solution.

We showed that, apart from these extreme cases, using approximations on the distribution of the unknown symbols represents a good trade-off between estimation accuracy and computational complexity, most of all in the low-SNR regime, where the noisy observations provide less evidence on the unknown symbols: both the Gaussian and the
Constant Modulus assumption outperform the pilot based estimator, within a reasonable complexity, and without incurring in the computational overhead required when treating the true discrete distribution of the unknown symbols.
Appendix A

Complex derivatives

Let $f(\theta)$ be a complex function on the complex parameter $\theta \in \mathbb{C}$. The complex derivative of $f(\theta)$ with respect to its argument $\theta$ is defined as

$$
\frac{\partial f(\theta)}{\partial \theta} = \frac{1}{2} \frac{\partial f(\theta)}{\partial \text{real}(\theta)} - \frac{i}{2} \frac{\partial f(\theta)}{\partial \text{imag}(\theta)}
$$

(A.1)

We now list the expressions for the complex derivatives of functions widely used throughout the thesis:

1. Let $\theta \in \mathbb{C}$. Then we have

$$
\begin{aligned}
\frac{\partial \theta}{\partial \theta} &= 1 \\
\frac{\partial \theta^*}{\partial \theta} &= 0 \\
\frac{\partial |\theta|^2}{\partial \theta} &= \theta \frac{\partial \theta^*}{\partial \theta} + \theta^* \frac{\partial \theta}{\partial \theta} = \theta^*
\end{aligned}
$$

(A.2)

2. Let $A$ be an $N \times N$ matrix with complex entries. Then we have

$$
\begin{aligned}
\frac{\partial |A|}{\partial \theta} &= \text{trace} \left( A^{-1} \frac{\partial A}{\partial \theta} \right) \\
\frac{\partial A^{-1}}{\partial \theta} &= -A^{-1} \frac{\partial A}{\partial \theta} A^{-1}
\end{aligned}
$$

(A.3)
Appendix B

Computation of the posterior mean of constant modulus symbols

Let $V_{nk}$ be the unknown symbol transmitted on sub-carrier $n$ at time $k$, drawn uniformly from an M-PSK constellation $C$. Let $Y_{nk}$ be the corresponding $R \times 1$ observation vector and $H_n$ the $R \times T$ channel matrix. In general $T \geq 1$, in such case the symbol transmitted from the antenna array is given by $X_{nk} = CV_{nk}$, where $C$ is a $T \times 1$ coding matrix.

Continuing from chapter 3.4, equation 3.56, we have the following expression for the posterior probability of the unknown symbol $V_{nk}$:

$$E_{V_{nk}} [V_{nk} | Y_{nk}, h] = \frac{\sum_{\alpha \in C} \alpha \exp \{2\text{real} (\alpha Y_{nk}^H B_{\eta_n} H_n C)\}}{\sum_{\alpha \in C} \exp \{2\text{real} (\alpha Y_{nk}^H B_{\eta_n} H_n C)\}}$$  \hspace{1cm} (B.1)

The approach here proposed consists in carrying out a Taylor series expansion of the terms $\exp \{2\text{real} (\alpha Y_{nk}^H B_{\eta_n} H_n C)\}$, based on the Taylor series expansion of the exponential, and then study the limit case of the constellation order $M$ going to infinity.

Therefore, let

$$f_M (\rho, b) = \frac{1}{M} \sum_{\alpha \in C} \alpha^b \exp \{2\text{real} (\rho^* \alpha)\}$$  \hspace{1cm} (B.2)

where we defined $\rho = C^H H_n^H B_{\eta_n} Y_{nk}$.

Using this notation, we can rewrite the posterior expectation as

$$E [V_{nk} | Y_{nk}, H_n] = \frac{f_M (\rho, 1)}{f_M (\rho, 0)}$$  \hspace{1cm} (B.3)
Using the Taylor series expansion of the exponential: $e^x = \sum_{n=0}^{+\infty} \frac{x^n}{n!}$ and substituting into $f_M (\rho, b)$ we obtain:

$$f_M (\rho, b) = \frac{1}{M} \sum_{\alpha \in \mathcal{C}} \sum_{n=0}^{+\infty} \alpha^b \sum_{n=0}^{+\infty} \frac{[2\text{real}(\rho^*\alpha)]^n}{n!} = \frac{1}{M} \sum_{\alpha \in \mathcal{C}} \sum_{n=0}^{+\infty} \frac{[\rho^*\alpha + \alpha^*\rho]^n}{n!}$$  \hspace{1cm} (B.4)

Now, we can rewrite $(\rho^*\alpha + \alpha^*\rho)^n$ using the binomial theorem $(a + b)^n = \sum_{k=0}^{n} a^k b^{n-k}$, thus obtaining:

$$(\rho^*\alpha + \alpha^*\rho)^n = \sum_{k=0}^{n} \binom{n}{k} (\rho^*\alpha)^k (\alpha^*\rho)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} |\rho|^n e^{i\theta(n-2k)} \alpha^k \alpha^*(n-k)*$$  \hspace{1cm} (B.5)

where in the last passage we have split $\rho$ into its modulo and phase components, $\rho = |\rho| e^{i\theta}$.

Then, substituting into (B.4) we obtain:

$$f_M (\rho, b) = \frac{1}{M} \sum_{\alpha \in \mathcal{C}} \sum_{n=0}^{+\infty} \alpha^b \sum_{k=0}^{n} \frac{1}{k!(n-k)!} |\rho|^n e^{i\theta(n-2k)} \alpha^k \alpha^*(n-k)*$$

$$= \sum_{n=0}^{+\infty} \sum_{k=0}^{n} \frac{1}{k!(n-k)!} |\rho|^n e^{i\theta(n-2k)} \frac{1}{M} \sum_{\alpha \in \mathcal{C}} \alpha^b \alpha^*(n-k)*$$  \hspace{1cm} (B.6)

Now, observe the last sum over the constellation points. Let’s consider a M-PSK or 4-QAM constellation, where all the constellation points have constant amplitude $\sigma_s$ and phase belonging to the set \{\$\theta_0 + \frac{2\pi m}{M}\$, $m = 0 \ldots M - 1\}$, where $\theta_0$ is the phase of the first symbol of the alphabet $\mathcal{C}$ and $\frac{2\pi}{M}$ is the phase spacing between the $M$ constellation points. Then we can write:

$$\sum_{\alpha \in \mathcal{C}} \alpha^b \alpha^*(n-k)* = \sigma_s^{b+n} e^{i\theta_0(b+2k-n)} \sum_{m=0}^{M-1} e^{i\frac{2\pi m}{M}(b+2k-n)}$$

$$= \sigma_s^{b+n} e^{i\theta_0(b+2k-n)} M \chi (b + 2k - n \mod M = 0)$$  \hspace{1cm} (B.7)

where $\chi(\text{prop})$ is the indicating function, which is equal to 1 when the proposition $\text{prop}$ is true, and zero otherwise.

Then, $f_M (\rho, b)$ can be rewritten as:

$$f_M (\rho, b) = \sum_{n=0}^{+\infty} \sum_{k=0}^{n} \frac{|\rho|^n e^{i\theta(n-2k)} \sigma_s^{b+n} e^{i\theta_0(b+2k-n)}}{k!(n-k)!} \chi (b + 2k - n \mod M = 0)$$  \hspace{1cm} (B.8)

Now, notice that $M$ is even, whereas $b$ can assume the values 0 or 1, therefore for $b = 0$ $\chi (b + 2k - n \mod M = 0)$ is always zero for $n$ odd, whereas for $b = 1$ it is always zero for $n$ even, therefore it is possible to restrict the sum over all $n \geq 0$ to a sum over the
Appendix B Computation of the posterior mean of constant modulus symbols

105

even (for \( b = 0 \)), or odd (for \( b = 1 \)) \( n \) only, yielding to:

\[
f_M(\rho, b) = \sum_{n=0}^{+\infty} \sum_{k=0}^{2n+b} \frac{|\rho|^{2n+b} e^{i\theta(2n+b-k)} \sigma_s^{2n+2b} e^{i\theta_0(2k-2n)}}{k! (2n+b-k)!} \chi \left(k - n \mod \frac{M}{2} = 0\right) \quad (B.9)
\]

Now, substituting \( k - n \) with \( k \):

\[
f_M(\rho, b) = \sum_{n=0}^{+\infty} \sum_{k=-n}^{n+b} \frac{|\rho|^{2n+b} e^{i\theta(b-2k)} \sigma_s^{2n+2b} e^{i\theta_0(2k)}}{(k+n)! (n+b-k)!} \chi \left(k \mod \frac{M}{2} = 0\right) \quad (B.10)
\]

Now, substituting \( n \) with \( \frac{M}{2} m + p \), where \( m \) goes from 0 to \( +\infty \) and \( p \) goes from 0 to \( \frac{M}{2} - 1 \), we obtain:

\[
f_M(\rho, b) = \sum_{m=0}^{+\infty} \sum_{p=0}^{\frac{M}{2} m + p} \sum_{k=-\frac{M}{2} m - p}^{\frac{M}{2} m + p+b} \frac{|\rho|^{M m + 2p + b} e^{i\theta(b-2k)} \sigma_s^{M m + 2p + 2b} e^{i\theta_0(2k)}}{(k+\frac{M}{2} m + p)! (\frac{M}{2} m + p+b-k)!} \chi \left(k \mod \frac{M}{2} = 0\right) \quad (B.11)
\]

The sum \( \sum_{k=-\frac{M}{2} m - p}^{\frac{M}{2} m + p+b} \) together with the fact that \( \chi \left(k \mod \frac{M}{2} = 0\right) \) is non null only when \( k \) is multiple of \( \frac{M}{2} \). Therefore the sum, together with the fact that \( p = 0 \ldots M/2 - 1 \), is equivalent to:

\[
\sum_{k=-\frac{M}{2} m - p}^{\frac{M}{2} m + p+b} (\cdot) = \sum_{k=-\frac{M}{2} m}^{\frac{M}{2} m} (\cdot) + \chi (b = 1) \chi \left(p = \frac{M}{2} - 1\right) \chi \left(k = \frac{M}{2} (m + 1)\right) \quad (B.12)
\]

Therefore, substituting this sum into \( f_M(\rho, b) \), and substituting the sum over \( k \) with the sum of \( k \) over the multiples of \( \frac{M}{2} \) yields to:

\[
f_M(\rho, b) = 2^{+\infty} \sum_{m=0}^{\frac{M}{2} - 1} \sum_{p=0}^{m} \sum_{k=-m}^{m} \frac{|\rho|^{M m + 2p + b} e^{i\theta(M k - b)} \sigma_s^{M m + 2p + 2b} e^{i\theta_0(k)}}{\left[\frac{M}{2} (m + k) + p\right]! \left[\frac{M}{2} (m - k) + p + b\right]!} + \chi (b = 1) \sum_{m=0}^{+\infty} \frac{(\rho^*)^{M (m+1) - 1} \sigma_s^{(M (m+1))} e^{i\theta_0 (m+1)}}{[M (m + 1) - 1]!} \quad (B.13)
\]
Finally, splitting the sum over \( k \) into the sums over \( k < 0, k = 0 \) and \( k > 0 \) we have:

\[
f_M (\rho, b) = \sum_{m=0}^{+\infty} \sum_{p=0}^{M-1} \frac{|\rho|^2 e^{i\theta} \sigma_s^{Mm+2p} e^{-i\theta(K-\rho)} \sigma_s^2 e^{iM\theta_0}}{(M/2)^{m+p+b}!} + \sum_{m=1}^{+\infty} \sum_{p=0}^{M-1} \sum_{k=1}^{m} \frac{|\rho|^2 e^{i\theta} \sigma_s^{Mm+2p} e^{-i\theta(K-\rho)} \sigma_s^2 e^{iM\theta_0}}{(M/2)^{m+k+p+b}!} + \sum_{m=1}^{+\infty} \sum_{p=0}^{M-1} \sum_{k=1}^{m} \frac{|\rho|^2 e^{i\theta} \sigma_s^{Mm+2p} e^{-i\theta(K+\rho)} \sigma_s^2 e^{-iM\theta_0}}{(M/2)^{m-k+p+b}!} + \chi (b = 1) \sum_{m=0}^{+\infty} \frac{|\rho|^2 e^{i\theta} \sigma_s^{Mm+2p} e^{-i\theta(M(m+1)-1)} \sigma_s^2 e^{-iM(m+1)\theta_0}}{(M(m+1)-1)!}
\]

(B.14)

Now, observe that, substituting \( \frac{M}{2} m + p \) with \( n \) into the first term, this can be rewritten as:

\[
\sum_{m=0}^{+\infty} \sum_{p=0}^{M-1} \frac{|\rho|^2 e^{i\theta} \sigma_s^{Mm+2p} e^{-i\theta(K-\rho)} \sigma_s^2 e^{iM\theta_0}}{(M/2)^{m+p+b}!} = \sum_{n=0}^{+\infty} \frac{|\rho|^2 e^{i\theta} \sigma_s^{Mm+2p} e^{-i\theta(M(m+1)-1)} \sigma_s^2 e^{iM(m+1)\theta_0}}{n!(n+b)!}
\]

(B.15)

which does not depend on \( M \).

As for the second, the third and the fourth terms, under regularity conditions they converge to zero as \( M \) goes to infinity.

Then, taking the limit for the constellation order \( M \) going to infinity we obtain:

\[
f (\rho, b) = \lim_{M \to +\infty} f_M (\rho, b) = \sum_{n=0}^{+\infty} \frac{1}{n!(n+b)!} |\rho|^{2n+b} e^{i\theta} \sigma_s^{2n+2b} \sigma_s^2 e^{iM\theta_0} \sigma_s^2 e^{iM\theta_0}
\]

(B.16)

Therefore the posterior expectation can be approximated with:

\[
E [V|Y, H] = \frac{f_M (\rho, 1)}{f_M (\rho, 0)} \simeq \sigma_s e^{i\theta} \sum_{n=0}^{+\infty} \frac{1}{n!(n+1)!} \left( |\rho| \sigma_s \right)^{2n+1} \frac{1}{(n+1)!} \left( |\rho| \sigma_s \right)^{2n} = \sigma_s e^{i\theta} g (|\rho| \sigma_s)
\]

(B.17)

where we define the scalar function:

\[
g(x) = \sum_{n=0}^{+\infty} \frac{1}{n!(n+1)!} x^{2n+1} \quad \forall \quad x \geq 0
\]

(B.18)
Appendix C

Cramér–Rao lower bound

C.1 Unbiased Cramér–Rao lower bound for Complex parameters

Let $\bar{\theta} \in \mathbb{C}^{K \times 1}$ be a complex vector of parameters, and $\hat{\theta} \in \mathbb{C}^{K \times 1}$ an unbiased estimator of $\bar{\theta}$.

Making explicit the real and imaginary part of $\bar{\theta}$ and $\hat{\theta}$ as $\bar{\theta} = \bar{\alpha} + i\bar{\beta}$ and $\hat{\theta} = \hat{\alpha} + i\hat{\beta}$, the corresponding real parameter vector and unbiased estimate are given by $\bar{\xi} = [\bar{\alpha}^T, \bar{\beta}^T]^T \in \mathbb{R}^{2K \times 1}$ and $\hat{\xi} = [\hat{\alpha}^T, \hat{\beta}^T]^T \in \mathbb{R}^{2K \times 1}$. Therefore, from the Cramér–Rao lower bound (CRLB) for real parameters ([3]), the following inequality holds:

$$\text{Cov} \left( \hat{\xi} \right) = E \left[ (\hat{\xi} - \bar{\xi}) (\hat{\xi} - \bar{\xi})^T \right] \geq E \left[ \Delta_{\bar{\xi}} \left( \ln p \left( Y | \bar{\xi} \right) \right) \cdot \Delta_{\bar{\xi}} \left( \ln p \left( Y | \bar{\xi} \right) \right)^T \right]^{-1} = \text{CRLB}_{\bar{\xi}} \quad (C.1)$$

where $\ln p \left( Y | \bar{\xi} \right)$ is the log-likelihood of the observations conditioned on the parameter vector $\bar{\xi}$, and $\Delta_R (f(R))$ is a matrix the same dimension of $R$ with elements

$$[\Delta_R (f(R))]_{p,l} = \frac{\partial f (R)}{\partial R(p,l)} \quad (C.2)$$

representing the gradient of the scalar function $f(R)$ with respect to matrix $R$.

The matrix $E \left[ \Delta_{\bar{\xi}} \left( \ln p \left( Y | \bar{\xi} \right) \right) \cdot \Delta_{\bar{\xi}} \left( \ln p \left( Y | \bar{\xi} \right) \right)^T \right]$ in the above expressions represents the Fisher Information Matrix.
The above inequality says that the covariance matrix of the estimation error $\hat{\xi} - \bar{\xi}$ is lower bounded by the inverse of the Fisher Information Matrix, where the inequality is intended as inequality for definite matrices.

Now, we want to determine the lower bound to the covariance matrix of the estimation error for the corresponding set of complex parameters, that is the complex vector $\bar{\gamma} = [\bar{\theta}^T, \bar{\theta}^H]^T \in \mathbb{C}^{2K \times 1}$.

Then, for the covariance matrix we have

$$\text{Cov} \left( \hat{\gamma} \right) = E \left[ (\hat{\gamma} - \bar{\gamma}) (\hat{\gamma} - \bar{\gamma})^H \right] \quad (C.3)$$

Now, expressing the real and imaginary components of the terms $\hat{\theta}$ and $\bar{\theta}$, it is straightforward to show that

$$\hat{\gamma} - \bar{\gamma} = \begin{bmatrix} \hat{\theta} - \bar{\theta} \\ \bar{\theta}^* - \hat{\theta}^* \end{bmatrix} = \left( \begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix} \otimes I_K \right) \cdot (\hat{\xi} - \bar{\xi}) \quad (C.4)$$

and substituting the above expression in (C.3) we obtain

$$\text{Cov} \left( \hat{\gamma} \right) = \left( \begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix} \otimes I_K \right) \text{Cov} \left( \hat{\xi} \right) \left( \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix} \otimes I_K \right) \quad (C.5)$$

Then, from the properties of semidefinite-positive matrices, from (C.1) we have

$$\text{Cov} \left( \hat{\gamma} \right) \geq \left( \begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix} \otimes I_K \right) I_{\bar{\xi}}^{-1} \left( \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix} \otimes I_K \right) \quad (C.6)$$

Now, it can be easily shown that

$$\begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix} = 2 \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix}^{-1} \quad (C.7)$$

therefore we can rewrite the inequality above as

$$\text{Cov} \left( \hat{\gamma} \right) \geq \left\{ \frac{1}{4} \left( \begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix} \otimes I_K \right) I_{\bar{\xi}} \left( \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix} \otimes I_K \right) \right\}^{-1} \quad (C.8)$$
Now, using the definition of Fisher Information Matrix $I_{\bar{\xi}}$ given above in C.1 it is straightforward to show, using the definition of complex derivatives

\[
\frac{1}{4} \left( \left[ \begin{array}{cc} 1 & i \\ 1 & -i \end{array} \right] \otimes I_K \right) I_{\bar{\xi}} \left( \left[ \begin{array}{cc} 1 & 1 \\ -i & i \end{array} \right] \otimes I_K \right)
\]

\[= E \left[ \Delta_{\bar{\gamma}^*} \left( \ln p \left( Y | \bar{\gamma} \right) \right) \Delta_{\bar{\gamma}^*} \left( \ln p \left( Y | \bar{\theta} \right) \right)^H \right] \quad (C.9)\]

Finally, defining the Fisher Information Matrix for complex parameters as

\[I_{\gamma} = E \left[ \Delta_{\gamma^*} \left( \ln p \left( Y | \bar{\theta} \right) \right) \Delta_{\gamma^*} \left( \ln p \left( Y | \bar{\theta} \right) \right)^H \right] \quad (C.10)\]

from C.8 we have the following lower bound to the covariance matrix of the estimation error:

\[Cov \left( \bar{\gamma} \right) \geq I_{\gamma}^{-1} = CRLB_{\gamma} \quad (C.11)\]

This represents the Complex Cramér–Rao lower bound for the estimation of $\gamma$.

Often, it is impractical to use the whole CRLB matrix as a lower bound, and it is instead much more useful to use a scalar as a lower bound. Let’s assume that, instead of the error covariance matrix, we want to measure the average variance of the error on each entry of the complex vector $\bar{\theta}$. This is given by:

\[Var \left( \hat{\theta} \right) = \frac{1}{K} \sum_{k=0}^{K-1} E \left[ |\hat{\theta}_k - \bar{\theta}_k|^2 \right] = \frac{1}{K} \text{trace} \left( Cov \left( \theta \right) \right) \quad (C.12)\]

Now, we want to determine a lower bound to this quantity, as we did with the Cramér–Rao lower bound.

Observe that the trace of the covariance matrix of $\theta$ can be put in relation with the trace of the covariance matrix of $\gamma$ in the following way:

\[\text{trace} \left( Cov \left( \theta \right) \right) = \frac{1}{2} \text{trace} \left( Cov \left( \theta \right) \right) + \frac{1}{2} \text{trace} \left( Cov \left( \theta^* \right) \right) = \frac{1}{2} \text{trace} \left( Cov \left( \gamma \right) \right) \quad (C.13)\]

and using the properties of definite matrices we have:

\[Var \left( \hat{\theta} \right) = \frac{1}{K} \text{trace} \left( Cov \left( \theta \right) \right) \geq \frac{1}{2K} \text{trace} \left( CRLB_{\gamma} \right) \quad (C.14)\]

Moreover, from the properties of definite matrices, the equality between the trace of the CRLB and the trace of the Error Covariance matrix is achieved if and only if $Cov \left( \gamma \right) = CRLB_{\gamma}$, or equivalently if and only if the estimator achieves the CRLB. Therefore, the
lower bound to the mean variance of an unbiased estimator \( \hat{\theta} \) of the complex parameter vector \( \tilde{\theta} \) is given by \[ \text{C.14} \]

### C.2 Unbiased CRLB for pilot based estimator of MIMO-FIR channels

In this section we derive the unbiased Cramér–Rao lower bound for the estimation of the frequency domain MIMO-OFDM FIR Channel of length \( L \) based on pilots alone.

Since the channel is FIR of length \( L \), there is dependency of the channel taps in the frequency domain. Therefore the CRLB for the estimation of the channel in the frequency domain is constrained on the channel length \( L \). In order to keep it into account, we first of all determine the Cramér–Rao lower bound for the estimation of the time-domain channel matrix. In fact, letting \( h \) be an \( LRT \)-dimensional column vector with entries \( h(RTl + Tr + t) = h_l(r,t) \), and \( H \) an \( NRT \)-dimensional column vector with entries \( H(RTn + Tr + t) = H_n(r,t) \), \( H \) is a linear function of \( h \) through the Fourier transform

\[
    H = \sqrt{N} \left( \tilde{U}_N \otimes I_{RT} \right) h
\]

Therefore, denoting with \( \text{CRLB}_{h}^{(tr)} \) the complex unbiased Cramér–Rao lower bound for the estimation of \( h \) using the training based approach, the corresponding complex unbiased CRLB for the estimation of \( H \), \( \text{CRLB}_{H}^{(tr)} \) is given by

\[
    \text{CRLB}_{H}^{(tr)} = N \begin{bmatrix} \tilde{U}_N \otimes I_{RT} & 0 \\ 0 & \tilde{U}_N^* \otimes I_{RT} \end{bmatrix} \text{CRLB}_{h}^{(tr)} \begin{bmatrix} \tilde{U}_N \otimes I_{RT} & 0 \\ 0 & \tilde{U}_N^* \otimes I_{RT} \end{bmatrix}^H
\]

Moreover, using the trace of the CRLB instead of the whole CRLB matrix as a performance lower bound, as justified in the introduction to the appendix, we have the following lower bound to the variance of any unbiased estimator of the frequency domain MIMO-OFDM channel \( H \):

\[
    \text{CRLB}_{tr} = \frac{1}{2NR^2} \text{trace} \left( \text{CRLB}_{H}^{(tr)} \right)
\]

and substituting \[ \text{C.16} \] into the above expression we obtain:

\[
    \text{CRLB}_{tr} = \frac{1}{2RT} \text{trace} \left( \text{CRLB}_{h}^{(tr)} \begin{bmatrix} \tilde{U}_N \otimes I_{RT} & 0 \\ 0 & \tilde{U}_N^* \otimes I_{RT} \end{bmatrix} \right)^H
\]
Finally, using the fact that \((\tilde{U}_N \otimes I_{RT})^H (\tilde{U}_N \otimes I_{RT}) = I_{LRT}\) we have

\[
\text{CRLB}_{tr} = \frac{1}{2RT} \text{trace} \left( \text{CRLB}_{h}^{(tr)} \right)
\]  

(C.19)

Therefore, we have the following lower bound to the variance of any unbiased estimator \(\hat{H}\) of \(H\), averaged over the entries of \(\hat{H}\):

\[
\frac{1}{NRT} E \left[ (\hat{H} - H)^H (\hat{H} - H) \right] \geq \frac{1}{2RT} \text{trace} \left( \text{CRLB}_{h}^{(tr)} \right)
\]  

(C.20)

The calculation of \(\text{CRLB}_{h}^{(tr)}\) is performed by first computing the Fisher Information Matrix, which is derived in the following section.

### C.2.1 The Fisher Information Matrix for the estimation of \(h\)

Since there are \(R\) receiving antennas, \(T\) transmitting antennas and the channel length is \(L\), the number of unconstrained complex parameters to estimate is \(RTL\). Let \(h\) be the \(LRT\)-dimensional parameter column vector with entries \(h(RTL + Tr + t) = h_l(r, t)\) where \(h_l(r, t)\) represents the \(l\)th tap of the channel between receiving-transmitting antenna pairs \((r, t)\).

From the definition of Fisher Information Matrix for complex parameters given in C.10, we have the following decomposition:

\[
I_{tr} = \begin{pmatrix}
I_{h^*h} & I_{h^*hH} \\
I_{hh^*} & I_{hhH}
\end{pmatrix}
\]  

(C.21)

The negative log-likelihood of the observations conditioned on the channel matrix \(h\) and on the pilot symbols \(X^{(tr)}\) is given by:

\[
- \ln p \left( Y^{(tr)} | h, X^{(tr)} \right) = - \sum_{n=0}^{N-1} K^{(tr)}_n \ln \left( \frac{|\mathcal{B}_{h_n}|}{\pi^R} \right)
\]

\[+ \sum_{n=0}^{N-1} \text{trace} \left[ \mathcal{B}_{h_n} \left( Y_n^{(tr)} - H_n X_n^{(tr)} \right) \left( Y_n^{(tr)} - H_n X_n^{(tr)} \right)^H \right]
\]  

(C.22)

where \(H_n\) is given by:

\[
H_n = \sum_l h_l e^{-i2\pi ln \frac{R}{N}}
\]  

(C.23)
The derivative of the negative log-likelihood with respect to the channel coefficient $h_l(r, t)^*$ was calculated when deriving the ML estimator, and is given by 2.8:

$$-\frac{\partial \ln p(Y^{(tr)}|h, X^{(tr)})}{\partial h_l(r, t)^*} = \sum_{n=0}^{N-1} \text{trace} \left[ B_{h_n} \left( Y_n^{(tr)} - H_n X_n^{(tr)} \right) X_n^{(tr)H} \delta(t, r) \right] e^{i2\pi \frac{p}{N}}$$  \hspace{1cm} (C.24)

Similarly, for the derivative with respect to $h_l(r, t)$, we have

$$-\frac{\partial \ln p(Y^{(tr)}|h, X^{(tr)})}{\partial h_l(r, t)} = \left( -\frac{\partial \ln p(Y^{(tr)}|h, X^{(tr)})}{\partial h_l(r, t)^*} \right)^*$$  \hspace{1cm} (C.25)

Now for the second derivatives there are four cases, reduced to two from the following relations:

$$\begin{cases} -\frac{\partial^2 \ln p}{\partial h_l(r_1, t_1)\partial h_p(r_2, t_2)^*} = \left( -\frac{\partial^2 \ln p}{\partial h_l(r_1, t_1)^*\partial h_p(r_2, t_2)} \right)^* \\ -\frac{\partial^2 \ln p}{\partial h_l(r_1, t_1)\partial h_p(r_2, t_2)} = \left( -\frac{\partial^2 \ln p}{\partial h_l(r_1, t_1)^*\partial h_p(r_2, t_2)} \right)^* \end{cases}$$  \hspace{1cm} (C.26)

Calculating the first term and taking the expectation we have

$$-E \left[ \frac{\partial^2 \ln p}{\partial h_l(r_1, t_1)\partial h_p(r_2, t_2)^*} \right] = \sum_{n=0}^{N-1} B_{h_n}(r_2, r_1) \left( X_n^{(tr)} X_n^{(tr)H} \right)_{t_1t_2} e^{i2\pi \frac{p-n}{N}} \hspace{1cm} (C.27)$$

$$= \sum_n \left[ B_{h_n} \otimes \left( X_n^{(tr)^*} X_n^{(tr)} \right) \right]_{T_{r_2+t_2}, T_{r_1+t_1}} e^{i2\pi \frac{p-n}{N}}$$

Observe that this is equal to $\Gamma^{(tr)}_{xx}(RTp + Tr_2 + t_2; RTl + Tr_1 + t_1)$, whose entries are defined in 2.12.

Therefore, rewriting the above expression in matrix form, we have

$$\begin{cases} \mathcal{I}_{hh} = \Gamma^{(tr)\ast}_{xx} \\ \mathcal{I}_{h^\ast h^\ast} = \Gamma^{(tr)}_{xx} \end{cases}$$  \hspace{1cm} (C.28)

For the other second derivatives we obtain

$$-\frac{\partial^2 \ln p}{\partial h_l(r_1, t_1)\partial h_p(r_2, t_2)} = -\frac{\partial^2 \ln p}{\partial h_l(r_1, t_1)^*\partial h_p(r_2, t_2)^*} = 0$$  \hspace{1cm} (C.29)

Therefore

$$\mathcal{I}_{hh} = \mathcal{I}_{h^\ast h^\ast} = 0$$  \hspace{1cm} (C.30)
Then, we can write the Fisher Information Matrix as

\[
I_{tr} = \begin{pmatrix}
\Gamma^{(tr)}_{xx} & 0 \\
0 & \Gamma^{(tr)*}_{xx}
\end{pmatrix}
\]  

(C.31)

The complex Cramér–Rao lower bound for the estimation of the time domain-channel matrix is then given by

\[
CRLB_h = I_{tr}^{-1} = \begin{pmatrix}
\Gamma_{xx}^{-(tr)-1} & 0 \\
0 & \Gamma_{xx}^{(tr)-1*}
\end{pmatrix}
\]  

(C.32)

Finally, substituting CRLB\(_h\) into (C.19) we obtain

\[
\frac{1}{NRT} E \left[ \text{trace} \left( (\hat{H} - H)(\hat{H} - H)^H \right) \right] \geq \frac{1}{2RT} \text{trace} (CRLB_h) = \frac{1}{2RT} \text{trace} (I_{tr}^{-1}) = CRLB_{tr}
\]  

(C.33)

Now, observe that \(\Gamma^{(tr)}_{xx}\) is an Hermitian matrix, which implies that its inverse is Hermitian and the diagonal elements are real, therefore from the expression for CRLB\(_h\) in (C.32) we can rewrite the Cramér–Rao lower bound as

\[
CRLB_{tr} = \frac{1}{RT} \text{trace} \left( \Gamma_{xx}^{(tr)-1}\right)
\]  

(C.34)

Observe that the variance of the Maximum Likelihood estimator calculated in section 2.1.2.2 equals the unbiased CRLB. Therefore, in the training based approach the Maximum Likelihood estimator achieves the best performance from the point of view of the MSE among the unbiased estimators of the channel matrix \(H\).

### C.3 Unbiased CRLB for Semi-Blind estimation of MIMO-OFDM FIR Channels

In this section we derive the unbiased Cramér–Rao lower bound for Semi-Blind estimators of MIMO-OFDM FIR channels with the Gaussian assumption for the unknown symbols, denoted by CRLB\(_{sb}\). We will then show that CRLB\(_{sb}\) is lower than CRLB\(_{tr}\), the CRLB calculated in the previous chapter for the training sequence approach (section C.2), demonstrating that the potential estimation accuracy which can be achieved with the Semi-Blind approach is higher than the pilot based approach.

As we demonstrated when computing the CRLB for the estimation of \(H\) for the pilot based approach (section C.2), since the channel is FIR of length \(L\), in order to keep
into account this constraint, we first of all determine the Cramér–Rao lower bound for the estimation of the time-domain channel matrix. Moreover, using the trace of the covariance matrix instead of the whole CRLB matrix as a performance lower bound, as justified in the introduction to the appendix, we have the following lower bound on the variance of any unbiased estimator of the frequency domain channel $H$, as demonstrated in section C.2 (equation C.19)

$$\text{CRLB}_{sb} = \frac{1}{2RT} \text{trace } \left( \text{CRLB}_{h}^{(sb)} \right)$$  \hspace{1cm} (C.35)

where the subscript $sb$ stands for Semi-Blind approach.

For the derivation of the CRLB$_{sb}$, we refer to the system model described in section 1.2 of the introduction to the thesis. Then, using the Gaussian assumption for the unknown symbols, the observations in correspondence of pilot symbols are Gaussian distributed with mean $E \left[ Y_n^{(tr)} \right] = H_nX_n^{(tr)}$ and covariance matrix Cov($\eta_n$) (or equivalently precision matrix $B_{\eta_n}$), whereas the blind observations are Gaussian distributed with zero mean and covariance matrix:

$$\Sigma_{Y_n} = \sigma_s^2 H_n C C^H H_n^H + \text{Cov} (\eta_n)$$  \hspace{1cm} (C.36)

Therefore, the negative log-likelihood function is given by:

$$-\ln p = - \sum_n K_n^{(tr)} \ln \left( \frac{B_{\eta_n}}{\pi R} \right) + \sum_n K_n^{(bl)} \ln \left( \pi R |\Sigma_{Y_n}| \right) +$$

$$+ \sum_n \text{trace} \left[ B_{\eta_n} \left( Y_n^{(tr)} - H_n X_n^{(tr)} \right) \left( Y_n^{(tr)} - H_n X_n^{(tr)} \right)^H \right] +$$

$$+ \sum_n \text{trace} \left( \Sigma_{Y_n}^{-1} Y_n^{(bl)} Y_n^{(bl)H} \right)$$  \hspace{1cm} (C.37)

Observe that the negative log-likelihood C.37 can be split into the sum of the contribution deriving from the pilot symbols and the contribution from the blind observations. Therefore, using the linearity of the derivatives, we can split the Fisher Information Matrix into the Fisher information matrix associated to pilot observations plus the Fisher Information Matrix associated with blind observations, that is:

$$\mathcal{I}_{sb} = \mathcal{I}_{tr} + \mathcal{I}_{bl}$$  \hspace{1cm} (C.38)

$\mathcal{I}_{tr}$ was derived in the previous chapter, when calculating the unbiased Cramér–Rao lower bound for the training sequence channel estimator, and is given by expression C.31.
Therefore we need to calculate only $I_{bl}$. From [C.37] the contribution of the blind observations to the negative log-likelihood function is given by

$$- \ln p \left( Y^{(bl)} | h \right) = \sum_n K_n^{(bl)} \ln \left( \pi R | \Sigma_{Y_n} | \right) + \sum_n \text{trace} \left( \Sigma_{Y_n}^{-1} Y_n^{(bl)} Y_n^{(bl)H} \right)$$  (C.39)

Then the derivative of $- \ln p \left( Y^{(bl)} | h \right)$ with respect to the channel entry $h_l(r, t)$ is given by

$$- \frac{\partial \ln p \left( Y^{(bl)} | h \right)}{\partial h_l(r, t)} = \sum_n \text{trace} \left[ \Sigma_{Y_n}^{-1} \frac{\partial \Sigma_{Y_n}}{\partial h_l(r, t)} \left( K_n^{(bl)} I_R - \Sigma_{Y_n}^{-1} Y_n^{(bl)} Y_n^{(bl)H} \right) \right]$$  (C.40)

The derivative of the covariance matrix of the blind observations with respect to $h_l(r, t)$ is given by

$$\frac{\partial \Sigma_{Y_n}}{\partial h_l(r, t)} = \sigma_s^2 \delta(r, t) C C^H H_n^H e^{-i2\pi \frac{m}{N}}$$  (C.41)

Therefore, substituting into (C.40) we obtain

$$- \frac{\partial \ln p \left( Y^{(bl)} | h \right)}{\partial h_l(r, t)} = \sigma_s^2 \sum_n \left[ C C^H H_n^H \left( K_n^{(bl)} I_R - \Sigma_{Y_n}^{-1} Y_n^{(bl)} Y_n^{(bl)H} \right) \Sigma_{Y_n}^{-1} \right]_{tr} e^{-i2\pi \frac{m}{N}}$$  (C.42)

Similarly we have

$$- \frac{\partial \ln p \left( Y^{(bl)} | h \right)}{\partial h_l(r, t)^*} = \left( - \frac{\partial \ln p \left( Y^{(bl)} | h \right)}{\partial h_l(r, t)} \right)^*$$  (C.43)

For the second derivatives there are four cases, which are reduced to two from the following relations:

$$\begin{align*}
- \frac{\partial^2 \ln p_{bl}}{\partial h_l(r_1, t_1) \partial h_p(r_2, t_2)^*} & = \left( - \frac{\partial^2 \ln p_{bl}}{\partial h_l(r_1, t_1) \partial h_p(r_2, t_2)^*} \right)^* \\
- \frac{\partial^2 \ln p_{bl}}{\partial h_l(r_1, t_1) \partial h_p(r_2, t_2)^*} & = \left( - \frac{\partial^2 \ln p_{bl}}{\partial h_l(r_1, t_1) \partial h_p(r_2, t_2)^*} \right)^* 
\end{align*}$$  (C.44)
Using \( C.42 \) the first term can be written as

\[
- \frac{\partial^2 \ln p \left( Y^{(bl)} | h \right)}{\partial h_l(r_1, t_1) \partial h_p(r_2, t_2)} = \frac{\partial}{\partial h_l(r, t)} \left( - \frac{\partial \ln p \left( Y^{(bl)} | h \right)}{\partial h_l(r, t)} \right) \\
= \sigma_s^2 \sum_n \left[ CCH^H \frac{\partial H_n^H}{\partial h_p(r_2, t_2)} \left( K_n^{(bl)} I_R - \Sigma_n^{-1} Y_n^{(bl)} Y_n^{(bl)H} \Sigma_n^{-1} \right)_{t_1 r_1} e^{-2\pi i n} \\
- \sigma_s^2 \sum_n \left( CCH^H H_n^H \frac{\partial \Sigma_n^{-1}}{\partial h_p(r_2, t_2)} Y_n^{(bl)} Y_n^{(bl)H} \right)_{t_1 r_1} e^{-2\pi i n} \\
+ \sigma_s^2 \sum_n \left[ CCH^H H_n^H \left( K_n^{(bl)} I_R - \Sigma_n^{-1} Y_n^{(bl)} Y_n^{(bl)H} \right) \frac{\partial \Sigma_n^{-1}}{\partial h_p(r_2, t_2)} \right]_{t_1 r_1} e^{-2\pi i n} \right) \\
(C.45)
\]

where in the last equality we used the product rule of derivatives.

Now, taking the expectation with respect to the observations, and using the fact that
\( E[Y_n^{(bl)} Y_n^{(bl)H}] = K_n^{(bl)} \Sigma_n \), we obtain:

\[
- E \left[ \frac{\partial^2 \ln p \left( Y^{(bl)} | h \right)}{\partial h_l(r_1, t_1) \partial h_p(r_2, t_2)} \right] = - \sigma_s^2 \sum_n K_n^{(bl)} \left( CCH^H H_n^H \frac{\partial \Sigma_n^{-1}}{\partial h_p(r_2, t_2)} \right)_{t_1 r_1} e^{-2\pi i n} \\
(C.46)
\]

The derivative of the precision matrix of the blind observations \( \Sigma_n^{-1} \) with respect to the channel entry \( h_p(r_2, t_2)^* \) is given by:

\[
\frac{\partial \Sigma_n^{-1}}{\partial h_p^*(r_2, t_2)} = - \Sigma_n^{-1} \sigma_s^2 H_n CCH^H \delta(t_2, r_2) \Sigma_n^{-1} e^{2\pi i n} \\
(C.47)
\]

Then, substituting this expression into \( C.46 \) we obtain the entries of the Fisher Information Matrix:

\[
- E \left[ \frac{\partial^2 \ln p \left( Y^{(bl)} | h \right)}{\partial h_l(r_1, t_1) \partial h_p(r_2, t_2)} \right] = \sigma_s^4 \sum_n K_n^{(bl)} \left( CCH^H H_n^H \Sigma_n^{-1} H_n CCH^H \right)_{t_1 r_2} e^{i2\pi \frac{(p-1)n}{N}} \\
(C.48)
\]
For the other term in C.44 we have, using C.42

\[- \frac{\partial^2 \ln p(Y^{(bl)}|h)}{\partial h_l(r_1, t_1) \partial h_p(r_2, t_2)} = \frac{\partial}{\partial h_p(r_2, t_2)} \left( - \frac{\partial \ln p(Y^{(bl)}|h)}{\partial h_l(r, t)} \right) =
\]

\[
\sigma_s^2 \sum_n \frac{\partial}{\partial h_p(r_2, t_2)} \left[ CC^H H_n^H K_n^{(bl)} I_R - \Sigma_n^{-1} Y_n^{(bl)} Y_n^{(bl)H} \right] \Sigma_n^{-1} \left| t_{1r_1} \right| e^{-i2\pi \frac{n}{N}}
\]

\[- \sigma_s^2 \sum_n \left( CC^H H_n^H \frac{\partial \Sigma_n^{-1}}{\partial h_p(r_2, t_2)} Y_n^{(bl)} Y_n^{(bl)H} \right) \left| t_{1r_1} \right| e^{-i2\pi \frac{n}{N}}
\]

\[+ \sigma_s^2 \sum_n \left[ CC^H H_n^H \left( K_n^{(bl)} I_R - \Sigma_n^{-1} Y_n^{(bl)} Y_n^{(bl)H} \right) \frac{\partial \Sigma_n^{-1}}{\partial h_p(r_2, t_2)} \right] \left| t_{1r_1} \right| \]  

where in the last equality we used the product rule of derivatives.

Now, taking the expectation with respect to the observations, and using the fact that

\[E[Y_n^{(bl)} Y_n^{(bl)H}] = K_n^{(bl)} \Sigma_n,\]

we obtain:

\[-E \left[ \frac{\partial^2 \ln p(Y^{(bl)}|h)}{\partial h_l(r_1, t_1) \partial h_p(r_2, t_2)} \right] = \sigma_s^2 \sum_n K_n^{(bl)} \left( CC^H H_n^H \Sigma_n^{-1} \frac{\partial \Sigma_n}{\partial h_p(r_2, t_2)} \Sigma_n^{-1} \right) \left| t_{1r_1} \right| e^{-i2\pi \frac{n}{N}}
\]

\[= \sigma_s^4 \sum_n K_n^{(bl)} \left( CC^H H_n^H \Sigma_n^{-1} \partial \Sigma_n \Sigma_n^{-1} \right) \left| t_{1r_1} t_{2r_2} \right| e^{-i2\pi \frac{(l+p)n}{N}}
\]

Finally, substituting C.44 into the above expression we obtain the entries of the Fisher Information Matrix:

\[E \left[ \frac{\partial^2 \ln p(Y^{(bl)}|h)}{\partial h_l(r_1, t_1) \partial h_p(r_2, t_2)} \right] = \sigma_s^4 \sum_n K_n^{(bl)} \left( CC^H H_n^H \Sigma_n^{-1} \partial \Sigma_n \Sigma_n^{-1} \right) \left| t_{1r_1} t_{2r_2} \right| e^{-i2\pi \frac{(l+p)n}{N}}
\]

To summarize, letting \( I_{bd}(\alpha, \beta) = -E \left[ \frac{\partial^2 \ln p(h|\theta)}{\partial \alpha \partial \beta} \right] \), the entries of the Fisher Information Matrix corresponding to the blind observations are given by:

\[
\begin{align*}
I_{bd}(h^*_b(r_1, t_1), h_b(r_2, t_2)) &= \sigma_s^4 \sum_n K_n^{(bl)} \left( \Sigma_n^{-1} \right)_{r_1r_2} \left( CC^H H_n^H \Sigma_n^{-1} H_n CC^H \right)_{t_2t_1} e^{i2\pi \frac{(l+p)n}{N}} \\
I_{bd}(h_b(r_1, t_1), h^*_b(r_2, t_2)) &= \sigma_s^4 \sum_n K_n^{(bl)} \left( \Sigma_n^{-1} \right)^*_{r_1r_2} \left( CC^H H_n^H \Sigma_n^{-1} H_n CC^H \right)^*_{t_2t_1} e^{-i2\pi \frac{(l+p)n}{N}} \\
I_{bd}(h_b(r_1, t_1), h_b(r_2, t_2)) &= \sigma_s^4 \sum_n K_n^{(bl)} \left( \Sigma_n^{-1} H_n CC^H \right)^*_{r_1r_2} \left( \Sigma_n^{-1} H_n CC^H \right)^*_{t_2t_1} e^{-i2\pi \frac{(l+p)n}{N}} \\
I_{bd}(h^*_b(r_1, t_1), h^*_b(r_2, t_2)) &= \sigma_s^4 \sum_n K_n^{(bl)} \left( \Sigma_n^{-1} H_n CC^H \right)_{r_1r_2} \left( \Sigma_n^{-1} H_n CC^H \right)_{t_2t_1} e^{i2\pi \frac{(l+p)n}{N}}
\end{align*}
\]

The unbiased CRLB matrix for the estimation of the time-domain channel matrix \( h \) is therefore given by:

\[\text{CRLB}_{h}^{(sb)} = (I_{tr} + I_{bd})^{-1}\]
with the entries of $I_{bl}$, the FIM associated to the blind observations, given by $C.52$ and $I_{tr}$, the FIM associated to the pilot observations, given by $C.31$.

Therefore have the following lower bound on the variance of any unbiased estimator of the frequency domain channel $H$, (equation $C.35$):

$$
\text{CRLB}_{sb} = \frac{1}{2RT} \text{trace} \left( \text{CRLB}_{h}^{(sb)} \right)
$$

(C.54)
Bibliography


