# Thèse

présentée pour obtenir le grade de docteur

de l'École Nationale Supérieure
des Télécommunications

Spécialité : Signal et Images

Elisabeth de Carvalho

# Identification de Canal et Égalisation Aveugles et Semi–Aveugles pour les Communications Mobiles

Soutenue le 28 juin 1999 devant le jury composé de

| | |
|---|---|
| Président | Claude Gueguen |
| Rapporteurs | Pierre Comon<br>Phillip Regalia |
| Examinateurs | Antoine Chouly<br>Philippe Loubaton |
| Directeur de Thèse | Dirk Slock |

École Nationale Supérieure des Télécommunications

# ABSTRACT

Most of the present mobile communication standards include a training sequence to estimate the channel. Blind techniques allow the estimation of the channel without requiring training symbols, thus increasing bandwidth efficiency, but lack from robustness. The purpose of semi-blind equalization is to exploit the blind information as well as the information coming from the known symbols. Semi–blind techniques robustify the blind problem and allow the estimation of longer impulse responses than possible with a certain training sequence length; for a desired estimation quality, they also allow the use of shorter training sequences. Furthermore, they offer better performance than blind and training methods.

We present identifiability conditions for semi–blind FIR multichannel estimation: semi–blind methods are able to estimate any channel, even when the position of the known symbols in the burst is arbitrary. Performance bounds for semi–blind multichannel estimation are provided through the analysis of Cramér-Rao bounds and a comparison of semi–blind techniques with blind and training sequence based techniques is done. A study on performance under constraints is proposed to characterize blind performance.

The proposed semi–blind methods are mainly based on Maximum–Likelihood which can incorporate the knowledge of input symbols. For grouped known symbols, suboptimal criteria appear as a linear combination of a training sequence based criterion and the blind ML criterion. In order to build powerful semi–blind ML methods, we also focus on the study of blind ML methods. At last, we present methods that combine a blind criterion with a training sequence based criterion.

Receiver structures are also presented. The structure of the burst mode equalizers are studied and especially the structure of the ISI canceller that we call Non–Causal Decision–Feedback Equalizer (NCDFE): an implementation of the NCDFE is proposed based on soft decisions. At last, performance bounds on Maximum Likelihood Sequence Estimation (MLSE) are given when the channel order is underestimated.

# ACKNOWLEDGEMENTS

I would like to thank Prof. Gueguen, of the GET, for doing me the honour of presiding the jury of this thesis, as well as Prof. Comon of the I3S Laboratory, Prof. Regalia of the Institut National des Télécommunications, Prof. Loubaton of the University of Marne–La–Vallée and Dr. Chouly of the Laboratoires d'Electronique Philips, for accepting to be members of the jury and for their valuable remarks on this work.

I am deeply grateful to my thesis advisor Prof. Dirk Slock: his great avaibility, constant help, encouragements and confidence in me contributed largely to this work.

I would like to thank the Laboratoires d'Electronique Philips, in Limeil–Brevannes, France which financed this thesis under a Cifre convention, and especially Dr. Chouly as well as Dr. Haghiri who launched the semi–blind idea.

I also thank the Institut Eurécom and the Mobile Communications Department for welcoming me and allowing me to conduct this research in excellent conditions.

# CONTENTS

# NOTATIONS

| | |
|---|---|
| $\mathbb{R}$ | Set of real numbers |
| $\mathbb{C}$ | Set of complex numbers |
| $(.)^T$ | Transpose |
| $(.)^H$ | Hermitian transpose |
| $(.)^{-1}$ | Inverse |
| $(.)^+$ | Moore Penrose Pseudo–Inverse |
| $\otimes$ | Kronecker product |
| $\mathrm{E}_X$ | Mathematical Expectation w.r.t. the random quantity $X$ |
| $\mathrm{Re}(.), \mathrm{Im}(.)$ | Real and Imaginary part |
| $I_n$ | Identity matrix of dimension $n$ |
| $I$ | Identity matrix of adequate dimension |
| $0_{p\times q}$ | Matrix with zero entries of dimension $p \times q$ |
| $0$ | Matrix with zero entries of adequate dimension |
| $A(i,j)$ | Element $(i,j)$ of matrix $A$ |
| $\mathrm{rank}(A)$ | Rank of matrix $A$ |
| $\mathrm{range}(A)$ | Column space of $A$ |
| $\mathrm{tr}(A)$ | Trace of square matrix $A$ |
| $\mathrm{vec}(A)$ | $[A_{i,1}^T \ A_{i,2}^T \cdots A_{i,n}^T]^T$ |
| $\mathrm{diag}(A)$ | Diagonal matrix with $\mathrm{diag}(A)(i,i) = A(i,i)$ |
| $P_A$ | Projection Matrix: $P_A = A\left(A^H A\right)^+ A^H$ |
| $P_A^\perp$ | $P_A^\perp = I - P_A$ |
| $\|.\|$ | Euclidean norm of a vector |
| $\|.\|_F$ | Froebenius norm of a matrix |
| $\delta_{ij}$ | Kronecker delta |
| $\hat{\theta}$ | Estimate of parameter $\theta$ |
| $\theta^o$ | True value of parameter $\theta$ |
| $\dfrac{\partial X^T}{\partial \theta}$ | Derivation of a vector w.r.t. to a vector is defined as the derivation of a column vector w.r.t. to a line vector: $\dfrac{\partial X^T}{\partial \theta}(i,j) = \dfrac{\partial X(j)}{\partial \theta(i)}$ |

| | |
|---|---|
| w.r.t. | with respect to |
| w.p. | with probability |
| MSE | Mean Squared Error |
| MMSE | Minimum Mean Squared Error |
| SNR | Signal to Noise Ratio |
| SOS | Second–Order Statistics |
| HOS | Higher–Order Statistics |
| SISO | Single Input Single Output |
| SIMO | Single Input Multiple Output |
| MIMO | Multiple Input Multiple Output |
| ISI | InterSymbol Interference |
| CRB | Cramér–Rao Bound |
| FIM | Fisher Information Matrix |
| ML | Maximum–Likelihood |
| DML | Deterministic Maximum–Likelihood |
| GML | Gaussian Maximum–Likelihood |
| SML | Stochastic Maximum–Likelihood |
| SF | Subspace Fitting |
| SSF | Signal Subspace Fitting |
| NSF | Noise Subspace Fitting |
| CM | Covariance Matching |
| OCM | Optimally weighted Covariance Matching |
| FA | Finite Alphabet |
| TS | Training Sequence |
| GSM | Global System for Mobile communications |
| DECT | Digital European Cordless Telephone |

Chapter 1

# INTRODUCTION

## 1.1 Training Sequence based Methods and Blind Methods

The development of wireless communications has given rise to a host of new research problems in digital communications, but has also refocused attention on some classical problems. Equalization is one of the main signal processing issues in digital communications over channels with InterSymbol Interference (ISI). In mobile communications, the ISI problem, due to multipath propagation, is particularly difficult as the propagation channels characteristics are severely subjected to pathloss and fading, and propagation characteristics change rapidly.

Traditional equalization techniques are based on training. The sender transmits a training sequence (TS) known at the receiver which is used to estimate the channel coefficients or to directly estimate the equalizer. Most of the present mobile communication standards include a training sequence to estimate the channel. In GSM [1], the data is organized and transmitted in bursts. Each normal burst contains a middamble training sequence used to estimate the channel, considered as time–invariant over the duration of a burst. A Viterbi equalizer based on the estimated channel is applied to estimate the transmitted data symbols of the actual burst.

In most cases, training methods appear as robust methods but present some disadvantages. Firstly, bandwidth efficiency decreases bandwidth efficiency as a non–negligible part of the data burst can be occupied: in GSM, for example, 20% of the bits in a burst are used for training. Furthermore, in certain communication systems, training sequences are not available or exploitable, when synchronization between the receiver and the transmitter is not possible.

These reasons motivated the introduction of the blind methods in the 70s with the work of Sato [2]. The idea behind blind equalization techniques is to estimate the channel or the equalizer based only on the received signal without any training symbols.

The first wave of blind techniques was based on the exploitation of the finite alphabet (decision directed, constant modulus algorithms, etc) while the second wave was based on Higher–Order Statistics (HOS) [3]. The HOS techniques use a Single Input Multiple Output (SISO) model, where a single input symbol stream is transmitted through a single linear channel and sampled at the symbol rate: from the second–order moment of the data, only the amplitude of the transfer function of the filter can be determined but not the phase function. Based on a condition of non–gaussianity for the sources (otherwise only first and second–order statistics are available), the higher–order methods can identify the phase function. The major disadvantage of HOS methods is that they often need a large amount of data resulting in a high computational cost.

The introduction of multichannels, or SIMO models where a single input symbol stream is transmitted through multiple linear channels, has given rise to a whole bunch of new blind estimation techniques that do not need higher–order statistics. When the received signal is oversampled at a rate higher than the symbol rate, the resulting sampled signal is cyclo-

stationary. Gardner [4], Tong, Xu and Kailath [5] proved that, due to spectral redundancy properties, both the amplitude and the phase function of the channel can be identified from the Second–Order Statistics (SOS) of the data. This temporally oversampled model was shown to be equivalent to a spatially oversampled model where the signal is received through multiple antennas [6]. As discussed in section 1.4, all the methods based on the multichannel model are not strictly based on second–order statistics of the data but rather on the structural properties of the received signal itself. We will however call these methods SOS methods as they do not use HOS. This thesis focuses on SOS techniques.

As detailed in section 1.4, some SOS estimation techniques suffer from a lack of robustness: channels must verify diversity conditions and the methods can fail when the channel length is overestimated. Furthermore, the blind techniques leave an indeterminacy in the channel or the symbols, a scale or phase factor (possibly discretely valued). This suggests that SOS blind techniques should not be used alone but with some form of additional information. However, the same is true also for training sequence based methods, especially when the sequence is too short to estimate the data. Semi–blind techniques are proposed here to overcome these problems.

## 1.2 The Semi–Blind Principle

In this thesis, we shall focus on blind and semi–blind FIR multichannel estimation that are further used to feed a Viterbi equalizer, or a linear or decision–feedback equalizer.

The data is transmitted by burst and we assume here that known symbols are present in the burst in the form of a training sequence aimed at estimating the channel or simply some known symbols used for synchronization or as guard intervals, like in the GSM or DECT burst. In this case, when using a training or a blind technique to estimate the channel, information gets lost. Training sequence methods base the parameter estimation only on the received signal containing only known symbols, and all the other observations, containing (some) unknown symbols, are ignored. Blind methods are based on the whole received signal, containing known and unknown symbols, possibly using hypotheses on the statistics of the input symbols, like the fact that they are i.i.d. for example, but no use is made of the knowledge of some input symbols. The purpose of semi–blind methods is to combine both training sequence and blind information (see figure 1.1) and exploit the positive aspects of both techniques stated in section 1.1.

Semi–blind techniques, because they incorporate the information of known symbols, avoid the possible pitfalls of blind methods and with only a few known symbols, any channel, single or multiple, becomes identifiable. Furthermore, exploiting the blind information in addition to the known symbols, allows one to estimate longer channel impulse responses than possible with a certain training sequence length, a feature that is of interest for the application of mobile communications in mountainous areas. For methods based on the

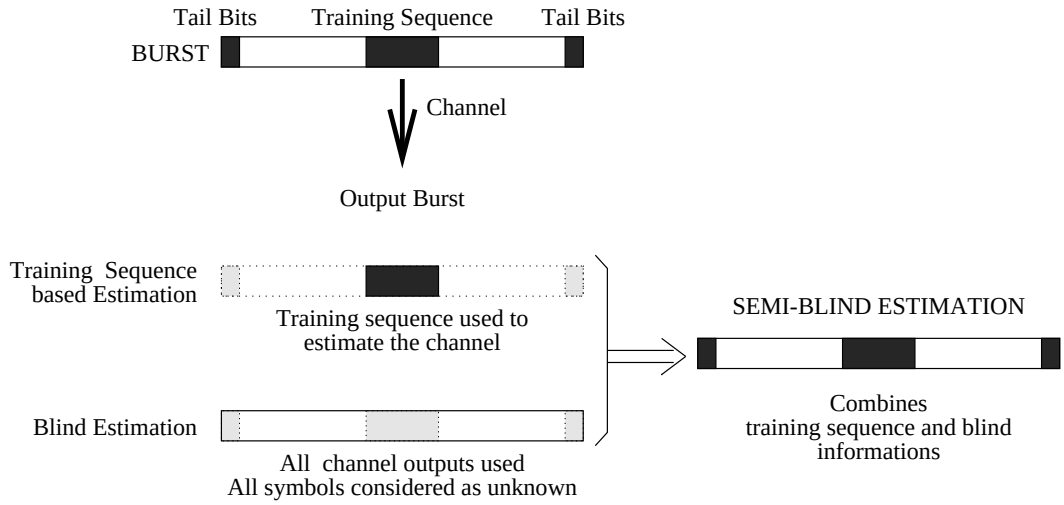



Figure 1.1: Semi–Blind Principle: example of a GSM burst.

second–order moments of the data (which we will call Gaussian methods), one known symbol is sufficient to make any channel identifiable. In addition, semi–blind tachniques allow one to use shorter training sequences for a given channel length and desired estimation quality, compared to a training approach. Apart from these robustness considerations, semi–blind techniques appear also very attractive from a performance point of view, as their performance is superior to that of training sequence or blind techniques separately. Semi–blind techniques are particularly promising when TS and blind methods fail separately: the combination of both can be successful in such cases.

## 1.3 The Multichannel Model

We consider here linear modulation (nonlinear modulations such as GMSK can be linearized with good approximation [7, 8]) over a linear channel with additive noise. The received signal after a linear receiver filter is then the convolution of the transmitted symbols with an overall channel impulse response, which is itself the convolution of the transmit shaping filter, the propagation channel and the receiver filter. The communication system is as figure 1.2.

The overall channel impulse response is modeled as FIR which for multipath propagation in mobile communications appears to be well justified. In mobile communications terminology, this thesis will mostly consider the single-user case; some work has also been done for the multi-user case in which the received signal contains a mixture of multiple users.

We describe the FIR multichannel model used throughout the thesis. This multichannel model applies to different cases (see figure 1.3): oversampling w.r.t. the symbol rate of a single received signal [5, 9, 10] or the separation into the real (in–phase) and imaginary

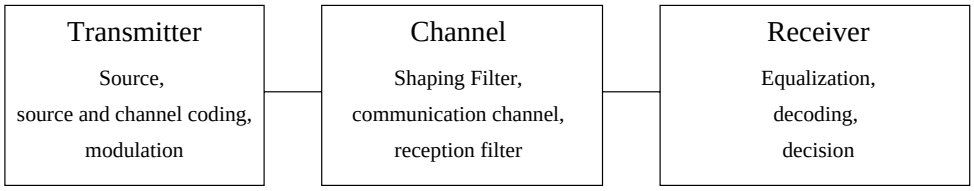



Figure 1.2: Communication system.

(quadrature) component of the demodulated received signal if the symbol constellation is real [11, 12]. In the context of mobile digital communications, a third possibility appears in the form of multiple received signals from an array of sensors (figure 1.3(b)). These three sources for multiple channels can also be combined.

Let us further develop the case of oversampling. The received signal which is cyclostationary [4] can be written as

$$\mathrm{y}(t) \;=\; \sum_{k} \mathrm{h}(t-kT)a(k) \;+\; \mathrm{v}(t) \tag{1.1}$$

where the $a(k)$ are the transmitted symbols, $T$ is the symbol period and $\mathrm{h}(t)$ is the channel impulse response. The FIR channel is assumed of duration $NT$ (approximately). If the received signal is oversampled at the rate $\frac{m}{T}$ (or if $m$ different received signals are captured by $m$ sensors every $T$ seconds, or a combination of both), the discrete input-output relationship can be written as:

$$\begin{aligned} \boldsymbol{y}(k) \;&=\; \sum_{i=0}^{N-1} \boldsymbol{h}(i)a(k-i) + \boldsymbol{v}(k)\;, \\ \boldsymbol{y}(k) = \begin{bmatrix} y_1(k) \\ \vdots \\ y_m(k) \end{bmatrix}, \boldsymbol{v}(k) &= \begin{bmatrix} v_1(k) \\ \vdots \\ v_m(k) \end{bmatrix}, \boldsymbol{h}(k) = \begin{bmatrix} h_1(k) \\ \vdots \\ h_m(k) \end{bmatrix} \end{aligned} \tag{1.2}$$

where the subscript $i$ denotes the $i^{th}$ channel. In the case of oversampling, $y_i(k)\,, i = 1, \ldots, m$ represent the $m$ phases of the polyphase representation of the oversampled signal: $y_i(k) = \mathrm{y}(t_0 + (k + \frac{i}{m})T)$. In this polyphase representation of the oversampled signals, we get a discrete-time circuit in which the sampling rate is the symbol rate.

For real symbols, it will be advantageous to treat the real and imaginary parts of the channel and received signal separately:

$$\begin{bmatrix} \mathrm{Re}(y_l(k)) \\ \mathrm{Im}(y_l(k)) \end{bmatrix} = \sum_{i=0}^{N-1} \begin{bmatrix} \mathrm{Re}(h_l(i)) \\ \mathrm{Im}(h_l(i)) \end{bmatrix} a(k-i) + \begin{bmatrix} \mathrm{Re}(v_l(k)) \\ \mathrm{Im}(v_l(k)) \end{bmatrix}\;,\quad l = 1, \ldots, n \tag{1.3}$$

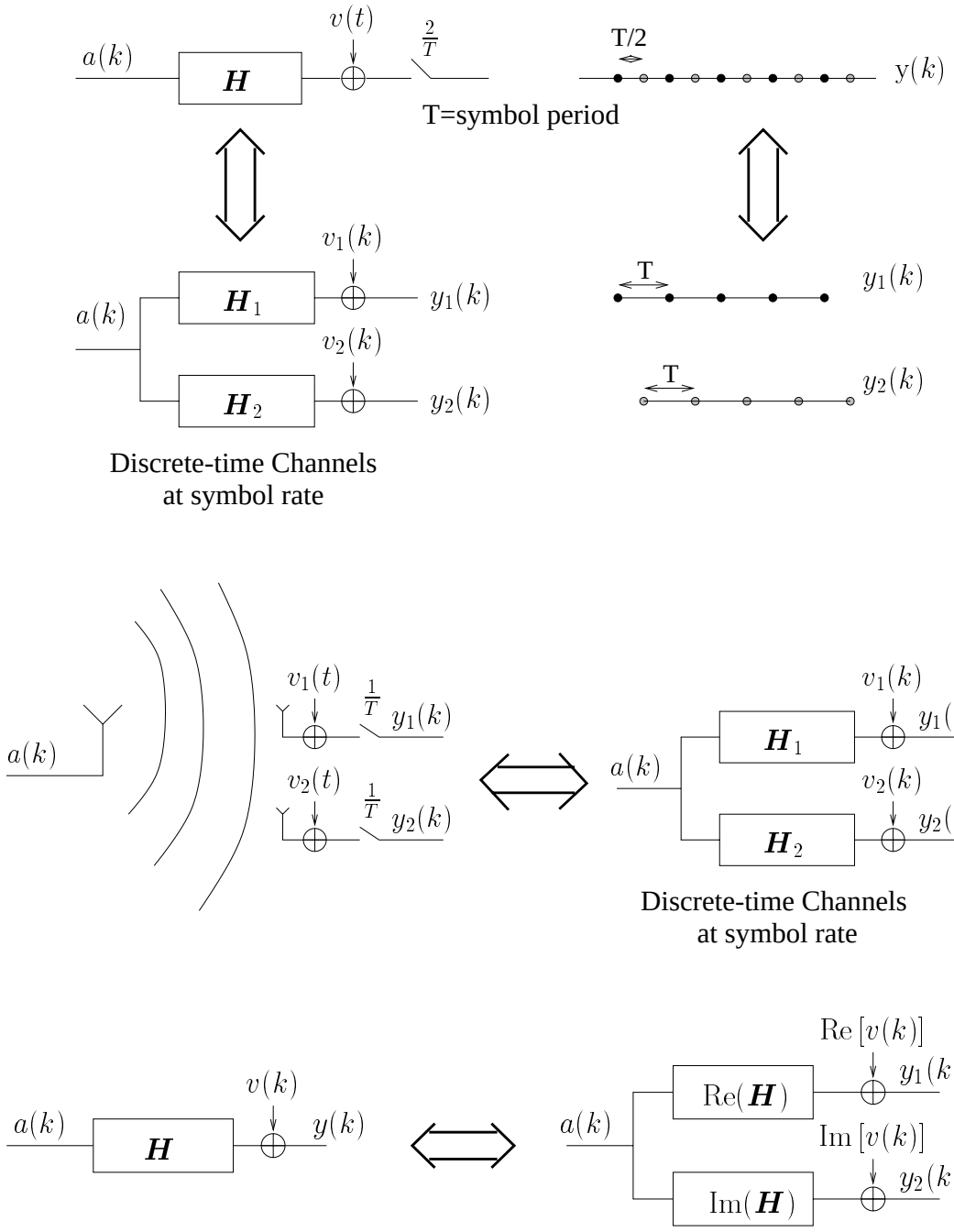



Figure 1.3: Multichannel model: case of oversampling, multiple antennas and separation of inphase and quadrature components when the input symbols are real. Example of a multichannel with 2 subchannels.

where $n$ now denotes the product of the oversampling factor and the number of sensors. The vector signals now become $\boldsymbol{y}(k) = [\mathrm{Re}(y_1(k))\ \mathrm{Im}(y_1(k))\cdots\mathrm{Re}(y_n(k))\ \mathrm{Im}(y_n(k))]^T$ and similarly for $\boldsymbol{h}(k)$ and $\boldsymbol{v}(k)$. This leads to a representation similar to (1.2). However, the number of channels gets doubled: $m = 2n$, which corresponds to an increase in diversity [11, 12]. With this reformulation of the case of real symbols, which we will henceforth assume, all quantities are real when the symbols are real.

In all cases, we can write the input-output relationship as

$$\begin{aligned} \boldsymbol{y}(k) &= \boldsymbol{H}A(k) + \boldsymbol{v}(k)\ , \\ \boldsymbol{H} &= [\boldsymbol{h}(0)\cdots\boldsymbol{h}(N-1)],\ A(k) = [a(k)\cdots a(k-N+1)]^T\ . \end{aligned} \tag{1.4}$$

The output is a vector signal corresponding to a SIMO (Single Input Multiple Output) or vector channel, consisting of $m$ SISO discrete-time channels. Note that monochannels appear as a limiting case of multichannels for which all the zeros are in common (except that in the multichannel case, the white noise variance is identifiable).

Let $\mathbf{H}(z) = \sum_{i=0}^{N-1} \boldsymbol{h}(i)z^{-i} = [\mathrm{H}_1(z)\cdots\mathrm{H}_m(z)]^T$ be the SIMO channel transfer function. Consider additive independent white Gaussian noise $\boldsymbol{v}(k)$ with $r_{\boldsymbol{vv}}(k-i) = \mathrm{E}\,\boldsymbol{v}(k)\boldsymbol{v}^H(i) = \sigma_v^2 I_m\,\delta_{ki}$, and $\mathrm{E}\,\boldsymbol{v}(k)\boldsymbol{v}^T(i) = 0$ in the complex case (circular noise). Assume we receive $M$ samples:

$$\boldsymbol{Y}_M(k) = \mathcal{T}_M(h)\,A_M(k) + \boldsymbol{V}_M(k) \tag{1.5}$$

where $\boldsymbol{Y}_M(k) = [\boldsymbol{y}^T(k)\cdots\boldsymbol{y}^T(k-M+1)]^T$ and similarly for $\boldsymbol{V}_M(k)$. $\mathcal{T}_M(h)$ is a block Toeplitz matrix with $M$ block rows and $[\boldsymbol{H}\ \ 0_{m\times(M-1)}]$ as first block row:

$$\mathcal{T}(h) = \begin{bmatrix} \boldsymbol{h}(0) & \cdots & \boldsymbol{h}(N-1) & 0 & \cdots & 0 \\ 0 & \boldsymbol{h}(0) & \cdots & \boldsymbol{h}(N-1) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \boldsymbol{h}(0) & \cdots & \boldsymbol{h}(N-1) \end{bmatrix} \tag{1.6}$$

and

$$h = \left[\boldsymbol{h}^T(0)\cdots\boldsymbol{h}^T(N-1)\right]^T. \tag{1.7}$$

The channel length is assumed to be $N$ which implies $\boldsymbol{h}(0) \neq 0$ and $\boldsymbol{h}(N-1) \neq 0$ whereas the impulse response is zero outside of the indicated range. We shall simplify the notation in (1.5) with $k = M-1$ to

$$\boldsymbol{Y} = \mathcal{T}(h)A + \boldsymbol{V}. \tag{1.8}$$

**Commutativity of Convolution** We will need the commutativity property of convolution:

$$\mathcal{T}(h)A = \mathcal{A}_m h \tag{1.9}$$

where: $\mathcal{A}_m = \mathcal{A}_1 \otimes I_m$,

$$\mathcal{A}_1 = \begin{bmatrix} a(M-1) & a(M-2) & \cdots & a(M-N) \\ a(M-2) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ a(0) & \cdots & \cdots & a(-N+1) \end{bmatrix}. \tag{1.10}$$

Sometimes, we will simplify $\mathcal{A}_m$ to $\mathcal{A}$.

**Semi–Blind Model** The vector of input symbols can be written as: $A = \mathcal{P} \begin{bmatrix} A_K \\ A_U \end{bmatrix}$ where $A_K$ are the $M_K$ known symbols and $A_U$ the $M_U = M+N-1-M_K$ unknown symbols. The known symbols may be dispersed in the burst and $\mathcal{P}$ designates the appropriate permutation matrix. For blind estimation $A = A_U$, while $A = A_K = A_{TS}$ for TS based estimation. We can split both parts in the channel output as $\mathcal{T}(h)A = \mathcal{T}_K(h)A_K + \mathcal{T}_U(h)A_U$.

**Irreducible, Reducible, Minimum-phase Channels** A channel is called irreducible if its subchannels $\mathrm{H}_i(z)$ have no zeros in common, and reducible otherwise. A reducible channel can be decomposed as:

$$\mathbf{H}(z) = \mathbf{H}_I(z)\mathrm{H}_c(z), \tag{1.11}$$

where $\mathbf{H}_I(z)$ of length $N_I$ is irreducible and $\mathrm{H}_c(z)$ of length $N_c = N - N_I + 1$ is a monochannel for which we assume $\mathrm{H}_c(\infty) = h_c(0) = 1$ (monic). A channel is called minimum–phase if all its zeros lie inside the unit circle. Hence $\mathbf{H}(z)$ is minimum–phase if and only if $\mathrm{H}_c(z)$ is minimum–phase.

**Minimum Zero-Forcing (ZF) Equalizer Length, Effective Number of Channels** The Bezout identity states that for an FIR irreducible channel, FIR ZF equalizers exist [13]. The minimum length for such an FIR ZF equalizer is

$$\underline{M} = \min \{M : \mathcal{T}_M(h) \text{ has full column rank}\} . \tag{1.12}$$

One may note that $\mathcal{T}_M(h)$ has full column rank for $M \geq \underline{M}$. In [14], it is shown that if the $mN$ elements of $\boldsymbol{H}$ are considered random, more precisely independently distributed with a continuous distribution, then

$$\underline{M} = \left\lceil \frac{N-1}{m-1} \right\rceil \quad \text{with probability 1.} \tag{1.13}$$

In this case, the channel is irreducible w.p. 1. One could consider other (perhaps more realistic) channel models. Consider e.g. a multipath channel with $K$ paths in which the

multichannel aspect comes from $m$ antennas. Without elaborating the details, it is possible to introduce an effective number of channels $m_e$ which in this case would equal (w.p. 1)

$$m_e \;=\; \mathrm{rank}(\boldsymbol{H}) \;=\; \min\{m, N, K\}\ . \tag{1.14}$$

With a reduced effective number of channels, the value of $\underline{M}$ increases to $\underline{M} = \left\lceil \frac{N-1}{m_e - 1} \right\rceil$ w.p. 1. Note that in the first probabilistic channel model leading to (1.13), if $m > N$, then in fact $m_e = N$, but this does not change the value of $\underline{M} = 1$. Another type of channel model arises in the case of a hilly terrain. In that case, two or more random non-zero portions of channel impulse response are disconnected by delays. If these delays are substantial, then for the purpose of determining $\underline{M}$, the problem can be approached as a multi–user problem by interpreting the different chunks of the channel as channels corresponding to different users. Multi–user results for $\underline{M}$ [13] could then be applied.

In general, for an irreducible channel, $\underline{M} \leq N-1$ [15] in which the upper bound would correspond to $m_e = 2$. Note that $m_e = 1$ corresponds to a reducible channel (in which case $\underline{M} = \infty$).

We summarize here the main notations that will be used in the thesis:

| | | |
|---|---|---|
| $M$ | : | Output Burst Length |
| $M_U$ | : | Number of Unknown Symbols |
| $M_K$ | : | Number of Known Symbols |
| $N$ | : | Channel Length |
| $m$ | : | Number of Subchannels |
| $\mathbf{H}(z) = \sum_{i=0}^{N-1} \boldsymbol{h}(i) z^{-i}$ | : | Transfer Function of the Multichannel |
| $\boldsymbol{H}$ | : | Channel Matrix |
| $\mathcal{T}(h)$ | : | Convolution Matrix |
| $A$ | : | Input Symbol Vector |
| $A_K$ | : | Vector of Known Symbols |
| $A_U$ | : | Vector of Unknown Symbols |
| $\boldsymbol{V}$ | : | Output Noise |
| $\boldsymbol{Y} = \mathcal{T}(h)A + \boldsymbol{V}$ | : | Input–Output Relationship |

## 1.4 Channel Identification Methods

We will see that semi-blind methods are based on TS and blind methods, and consist sometimes simply of a linear combination of a training based criterion and a blind criterion. Semi–blind methods inherit the characteristics of blind methods which is why the study of blind techniques is important. We propose here a brief state of the art of the main blind methods which will help to understand the motivation behind our methodological choices. The reader already familiar with this subject may proceed to section 1.5.

Blind methods can be classified according to the increasing a priori knowledge on the input symbols exploited[1], as follows (see figure 1.4):

1. No information exploited: the deterministic methods.
2. Second–order statistics: the Gaussian methods.
3. Higher–order statistics.
4. Finite symbol constellation alphabet: the Finite Alphabet (FA) methods
5. Complete Symbol Distribution: stochastic methods.

HOS methods are not the subject of this thesis, so we limit the remainder of this discussion to information levels 1, 2, 4 and 5. The proposed review follows this classification.

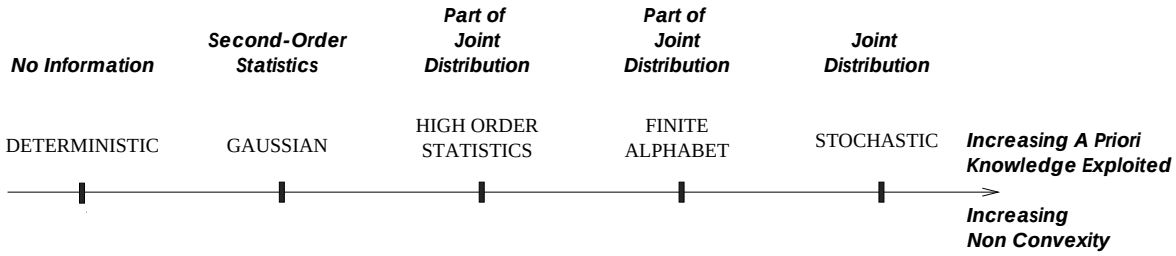



Figure 1.4: Classification of the channel identification methods according to the a priori knowledge about the input symbols exploited.

### 1.4.1 Deterministic Model

In the deterministic model, both input symbols and channel coefficients are assumed to be deterministic quantities. Deterministic methods proceed either to the joint estimation of $h$ and $A$ or to the estimation of $h$ with $A$ considered as a nuisance parameter (the estimation of $h$ and $A$ is decoupled from the estimation of $\sigma_v^2$). The estimation is done directly from

[1] between information level 3 and 4, the order could be inversed

the received signal $\boldsymbol{Y} = \mathcal{T}(h)A + \boldsymbol{V}$, and the channel can be estimated up to a scale factor: deterministic methods are often solved under the constraint $\|h\|^2 = 1$ (although it leaves a phase ambiguity, as explained in Chapter 3).

Consider the noise–free vector of $M$ received samples $\boldsymbol{X}_M = \mathcal{T}_M(h)A_M$. Determining $h$ and $A_M$ from $\boldsymbol{X}_M$ consists in solving a system of equations. As $h$ can be determined up to a scale factor, the system contains $Nm-1$ unknowns for the channel and $M+N-1$ unknowns for the symbols, and there are $Mm$ equations. If $m > 2$, provided that $M \geq N + \left\lceil 2\frac{N-1}{m-1} \right\rceil$, the system contains more equations than unknowns. This simple observation does not provide a proof of identifiability (a rigorous proof is given in Chapter 2), it is just intended to give an insight on how, from a multichannel deterministic point of view, the parameters $A$ and $h$ can be determined. For a monochannel, the number of equations will always be smaller than the number of unknowns, and deterministic methods fail.

Deterministic methods are based on structural properties of the received signal and especially on the low–rank property of $\mathcal{T}(h)$. For an irreducible channel and under certain conditions on the burst length and input symbols, the channel can indeed be determined uniquely (up to a scale factor) from the column space of $\mathcal{T}(h)$ that is called the signal subspace or from its orthogonal complement called the noise subspace. As detailed in Chapter 2, the signal or noise subspace can exactly be obtained from the noise–free signal $\boldsymbol{X}$ and estimated from the noisy data.

**Subspace Fitting Methods** Consider the sample covariance matrix of the received signal $\boldsymbol{Y}_L$ of length $L$ and its expected value (w.r.t. the noise only, as $A$ is deterministic):

$$R_{Y_L Y_L} = \mathcal{T}_L(h) \left[ \sum_{k=0}^{M-L-1} A_L(k) A_L^H(k) \right] \mathcal{T}_L^H(h) + \sigma_v^2 I \tag{1.15}$$

provided some regularity constraints are fulfilled $\sum_{k=0}^{M-L-1} A_L(k)A_L^H(k)$ is a square invertible matrix, so the space spanned by $\mathcal{T}_L(h) \left[ \sum_{k=0}^{M-L-1} A(k)A^H(k) \right] \mathcal{T}_L^H(h)$ is the signal subspace. $R_{Y_L Y_L}$ admits $M+N-1$ (the dimension of the signal subspace) eigenvectors belonging to the signal subspace, and $Mm-(M+N-1)$ eigenvectors belonging to the signal subspace all associated to the eigenvalue $\sigma_v^2$. The eigendecomposition of $R_{Y_L Y_L}$ is:

$$R_{Y_L Y_L} = V_{\mathcal{S}} \Lambda_{\mathcal{S}} V_{\mathcal{S}}^H + V_{\mathcal{N}} \Lambda_{\mathcal{N}} V_{\mathcal{N}}^H \tag{1.16}$$

where the columns of $V_{\mathcal{S}}$ span the signal subspace and the columns of $V_{\mathcal{N}}$ the noise subspace, $\Lambda_{\mathcal{N}} = \sigma_v^2 I$. Let $\widehat{V}_{\mathcal{S}}$ et $\widehat{V}_{\mathcal{N}}$ be estimates of the signal and noise eigenvectors obtained from the sample covariance matrix. The Signal Subspace Fitting (SSF) tries to fit the column space

of $\mathcal{T}(h)$ to its estimates through the quadratic criterion:

$$\min_{\|h\|^2=1} \|P_{\widehat{\mathcal{V}}_N} \mathcal{T}(h)\|^2 . \tag{1.17}$$

Another form of subspace fitting is noise subspace fitting:

$$\max_{\|h\|^2=1} \|P_{\widehat{\mathcal{V}}_S} \mathcal{T}(h)\|^2 . \tag{1.18}$$

The two criteria are quadratic in $h$, but the two methods require an eigendecomposition, which can be costly. A method avoiding the eigendecomposition was proposed in [16].

**Subchannel Response Matching (SRM)** SRM [17], which is also called Cross-Relation (CR) method [18], is based on a linear parametrization of the noise subspace. In the case $m = 2$ where the multichannel $\mathbf{H}(z) = [\mathrm{H}_1^T(z) \;\; \mathrm{H}_2^T(z)]^T$ has 2 subchannels, a parametrization of the noise subspace is $\mathbf{H}^\perp(z) = [-\mathrm{H}_2(z) \;\; \mathrm{H}_1(z)]$: $\mathcal{T}(h^\perp)\mathcal{T}(h) = 0$ where $\mathcal{T}(h^\perp)$ is the convolution matrix built from $\mathbf{H}^\perp(z)$ and spans the entire noise subspace. In the noise free case, $\mathcal{T}(h^\perp)\boldsymbol{Y}$ $(= \mathcal{T}(h^\perp)\mathcal{T}(h)A) = 0$. Using the commutativity of convolution $\mathcal{T}(h^\perp)\boldsymbol{Y} = \mathcal{Y}h$, where $\mathcal{Y}$ is a structured matrix filled out with the elements of $\boldsymbol{Y}$: the channel coefficients can be identified uniquely from this equation as the minimal left eigenvector of $\mathcal{Y}$ [18, 19]. When the received signal is noisy, $h$ is obtained by solving the least–squares quadratic criterion $\min_{\|h\|=1} \|\mathcal{T}(h^\perp)\boldsymbol{Y}\|^2 \Leftrightarrow \min_{\|h\|=1} \|\mathcal{Y}h\|^2$. For more than 2 subchannels, different noise parametrizations are possible [20]. In the case of 2 subchannels, SRM and SF are the same [21].

**Blocking equalizers determined by linear prediction** A minimum parameterization $\overline{\boldsymbol{P}}$ of the noise subspace can be found in terms of prediction quantities [9, 13, 22]: $\overline{\boldsymbol{P}}$ can be obtained from the prediction filters or through the SRM-like criterion $\min_{\overline{\boldsymbol{P}}} \|\mathcal{T}(\overline{\boldsymbol{P}})\boldsymbol{Y}\|^2$ with specific constraints on several coefficients of $\overline{\boldsymbol{P}}$ [23]. The channel is then determined uniquely by the subspace fitting criterion: $\min_{\|h\|=1} \|\overline{\boldsymbol{P}}\mathcal{T}(h)\|^2$. Such a parametrization of the noise subspace offers the advantage to hold in a multi–user context also: this is not true for the parametrization in terms of channel coefficients, which is intended exclusively for single user cases.

**Two–sided Linear Prediction or Least–Squares Smoothing** Recently a certain number of equivalent blind methods have been developed independently. Let $\boldsymbol{X}_N(k)$ be the noise free received signal of length $N$, the channel length, at time $k$.

$$\begin{aligned} \boldsymbol{X}_N(k) = & \underbrace{\bar{h}\, a(k)}_{\text{contribution of the symbol of interest}} + \\ & \underbrace{\bar{h}_{-1}a(k-1) + \cdots + \bar{h}_{-N+1}a(k-N+1)}_{\text{past contributions}} + \underbrace{\bar{h}_1 a(k+1) + \cdots + \bar{h}_{N-1}a(k+N-1)}_{\text{future contributions}} \end{aligned} \tag{1.19}$$

$\bar{h}$ is a block–wise flipped version of the vector $h$, and $\bar{h}_i$ contain a truncated version of $\bar{h}$. What these methods try to do is remove the past and future contributions, so that the resulting signal is $\bar{h}\ a(k)$ from which the channel coefficient can be recovered. $\bar{h}\ a(k)$ can also be seen as the prediction error of the two–sided linear prediction of $\boldsymbol{X}_N(k)$, which is estimated as the minimal eigenvector of a submatrix of $R^{-1}_{Y_L Y_L}$. This last interpretation is given in [24] based on [25], in which the Capon principle for linearly constrained minimum variance beamforming is applied to blind equalization. A Least–Squares smoothing solution is proposed in [26]. Other works give related techniques [27]. A joint order detection and channel estimation [26, 24] can be done, which is a major advantage for a deterministic method. In [24], the link between all these methods is established; it furthermore provides a correction of the methods [25, 26] which consider noisy quantities, which gives biased estimates.

**Deterministic Maximum–Likelihood (DML)** As detailed later in this thesis, the DML criterion is:

$$\min_{A,\|h\|=1} \|\boldsymbol{Y} - \mathcal{T}(h)A\|^2 . \tag{1.20}$$

This criterion can be solved directly in this form by minimizing alternatively w.r.t. $A$ and $h$ [28, 29]. This algorithm possesses some nice properties (see Chapter 7): at each iteration of the alternating minimization, the cost function decreases, and, for an asymptotic number of data, converges to the DML global estimate. It suffers, however, from slow convergence speed.

Another way of solving (1.20) is to eliminate $A$ (by minimizing w.r.t. $A$ and substituting its expression in (1.20)) to get a DML criterion in $h$:

$$\min_h \boldsymbol{Y}^H P^{\perp}_{\mathcal{T}(h)} \boldsymbol{Y} . \tag{1.21}$$

Computationally less intensive solutions to solve this criterion are based on a linear parametrization of the noise subspace. Using the parameterization $\mathbf{H}^{\perp}(z)$:

$$(1.21) \Rightarrow \min_h \boldsymbol{Y}^H \mathcal{T}^H(h^{\perp}) \left[\mathcal{T}(h^{\perp})\mathcal{T}^H(h^{\perp})\right]^+ \mathcal{T}(h^{\perp})\boldsymbol{Y} . \tag{1.22}$$

The Iterative Quadratic Maximum–Likehood (IQML) method was proposed in [19]: at each iteration, the denominator $\mathcal{T}(h^{\perp})\mathcal{T}^H(h^{\perp})$ is considered constant, evaluated from the previous iteration, so that the DML criterion becomes quadratic. In [9, 30], the IQML strategy was also proposed based on the blocking equalizers. At low SNR, IQML is biased and performs poorly: SRM used to initialize IQML in [19] performs in fact better at low SNR conditions. Solutions are proposed in this thesis to remove this bias due to the noise, one of which will be proven to asymptotically attain the DML performance.

DML is the most powerful method among all the deterministic methods. SRM however performs nearby as illustrated in this thesis. The optimally weighted subspace fitting method does not strictly belong to the deterministic category as the optimal weighting assumes the input symbols as i.i.d. and uses their statistics: they are located between the deterministic and Gaussian methods. For some weighting (not the optimal one), subspace fitting will have the same performance as DML asymptotically in the number of data, as in the Direction Of Arrival (DOA) context, but also in the size of the covariance matrix considered.

### 1.4.2 Gaussian Model

In the Gaussian model, the input symbols are considered to be i.i.d. Gaussian random variables with mean 0 and variance $\sigma_a^2$. This model may appear inappropriate as the input symbols are in fact discrete-valued.

The purpose of the Gaussian model is to take into account first and second–order moments of the data, which appear to play a predominant role in the multichannel context. In the blind case, the mean is zero (but it will not be the case for the semi–blind techniques) and the second–order moment is:

$$R_{YY}(\theta) = \sigma_a^2 \mathcal{T}(h)\mathcal{T}^H(h) + \sigma_v^2 I \,. \tag{1.23}$$

Unlike the deterministic case, the input symbols in the Gaussian model are no longer nuisance parameters for the estimation of $h$. The parameters to be jointly estimated are the channel coefficients and the noise variance. The channel is identifiable up to a phase factor and Gaussian methods should be solved using a phase constraint.

Already existing blind methods which base channel estimation on the second–order moments of the data, and in which the input symbols are considered i.i.d. random variables, can be classified into the Gaussian category as the three first methods described below. The Gaussian assumption is intended for the Maximum–Likelihood (ML) approach for which the complete distribution is required. The Gaussian distribution is the simplest distribution, leading to simple derivations and allowing to incorporate the first and second-order moments of the data: $\boldsymbol{Y} \sim \mathcal{N}(m_Y(\theta), R_{YY}(\theta))$; the Gaussian hypothesis for the symbols leads to a Gaussian distribution for $\boldsymbol{Y}$.

**Linear Prediction Approach** Linear prediction based techniques applied to multi–user channel identification were first approached in [31] using HOS. Independently, Slock [9] elaborated this method based on SOS, which was pursued by Abed Meraim [32] and finalized by Gorokhov [33]. Let $\mathbf{P}(z)$ be the MMSE linear prediction filter of the data. In the single channel case, the optimal prediction filter is of infinite length; in the multichannel case however, $\mathbf{P}(z)$ is finite. The fundamental equation is here:

$$\mathbf{P}(z)\mathbf{H}(z) = \boldsymbol{h}(0) \,. \tag{1.24}$$

The prediction error is $\sigma_{\tilde{y}}^2 = \sigma_a^2 \boldsymbol{h}(0)\boldsymbol{h}^H(0)$. $\mathbf{P}(z)$ is estimated from an estimate of the denoised covariance matrix of the data, and $\boldsymbol{h}(0)$ as the maximal eigenvector of the prediction error. $\boldsymbol{H}(z)$ is determined from (1.24) using a least-squares criterion or, giving better performance, by the weighted least–squares criterion [33]. The prediction method can be solved with different levels of complexity: the one based on the Levinson algorithm [34] has the lowest cost (compared to the one which bases the computation of the predictor on the pseudo–inverse of the covariance matrix.)

**The Schur Method** It can be proven that the LDU factorization (done by the Schur algorithm) of the denoised covariance matrix is: $R_{YY} = R_{Y\tilde{Y}} R_{\tilde{Y}\tilde{Y}} R_{\tilde{Y}Y} = LDL^H$, where $\tilde{\boldsymbol{Y}}$ is the prediction error vector. Considering denoised data, the prediction error signal is $\tilde{\boldsymbol{y}}(k) = \boldsymbol{h}(0)a(k)$ (obtained by equation (1.24)) and using $\boldsymbol{y}(k) = \sum_{i=0}^{N-1} \boldsymbol{h}(i)a(k-i)$, the block $(k,k')$ of $R_{Y\tilde{Y}}$ is $R_{Y\tilde{Y}}(k,k') = E(\boldsymbol{y}(k)\tilde{\boldsymbol{y}}^H(k')) = \sigma_a^2 \boldsymbol{h}(k-k')\boldsymbol{h}^H(0)$, so that the channel coefficients can be deduced from the columns of the triangular factor $L$. For more details, see [35, 36]. For the same level of complexity, the Schur method gives better simulated performance than the prediction method [34].

**Covariance Matching (CM) Method** The covariance matching method performs a weighted least-squares fit between the model of the second–order statistics of the received signal and their sample estimate built from the data. Let the vector $r(\theta)$ containing the non–redundant elements of $R_{YY}$ (see Chapter 9) and $\hat{r}$ the corresponding sample estimates. The covariance matching criterion is

$$\min_{h,\sigma_v^2} (r(\theta) - \hat{r})^H \mathcal{W} (r(\theta) - \hat{r}) \tag{1.25}$$

where $\mathcal{W}$ is a weighting matrix. In [37, 38], only the first $N$ non zero correlation coefficients are considered in the CM criterion. The optimal performance is obtained when the number of covariance lags involved tends to infinity, as stated in [39]: covariance matching is then asymptotically the best method exploiting the SOS.

**Gaussian Maximum Likelihood (GML)** As $\boldsymbol{Y} \sim \mathcal{N}(0, R_{YY}(\theta))$, the GML criterion is:

$$\min_{\theta=[h,\sigma_v^2]} \ln\det R_{YY}(\theta) + \boldsymbol{Y}^H R_{YY}^{-1}(\theta)\boldsymbol{Y}\,. \tag{1.26}$$

The Gaussian hypothesis is only used to build the GML criterion, which is solved using the true symbol distribution. A semi–blind ML method based on this model was proposed in [40] and shown to give better performance than ML based on the deterministic model [41]. The Gaussian hypothesis for the sources is also regularly used in direction of arrival finding and the associated ML is proven to give better performance than the deterministic ML methods [42].

The blind optimally weighted CM method [39] based on an asymptotically large covariance matrix will be shown, by numerical evaluation of the theoretical performance expressions, to be equivalent to blind GML. So GML, although based on a inaccurate hypothesis, appears to be the best method using SOS along with the optimally weighted CM. CM and GML are further studied in Chapter 9.

In [33], it is proven that the optimally weighted SSF has the same performance as the optimally weighted CM (computed with the constraint $\|h\|^2 = 1$ and a phase constraint), for a covariance matrix of infinite length and that optimally weighted SSF has the same performance as the weighted least squares based prediction method.

### 1.4.3 Methods Exploiting the Finite Alphabet

These methods are based on ML and exploit the finite alphabet (denoted $\mathcal{A}_p$) constraint of the input symbols:

$$\min_{h, A \in \mathcal{A}_p} \|\boldsymbol{Y} - \mathcal{T}(h)A\|^2 . \tag{1.27}$$

Some FA methods proceed by alternating minimizations between $h$ and $A$, with $A$ constrained to the finite alphabet. Both estimations are done in a least-squares way: the most problematic estimation is that of the symbols because of the FA constraint. The Viterbi algorithm can be used; a trellis search technique was proposed in [43], as well as a reduced–state sequence estimation in [44]. Talwar [45] proposed a much less complex solution: the FA constraint is first ignored and the symbols are estimated by a quadratic least–squares criterion (they are the output of a burst mode MMSE–ZF equalizer), then the estimates are projected onto the nearest discrete value of the finite alphabet. Some methods also exist that give closed form solution [46].

### 1.4.4 Stochastic ML Methods

SML considers the input symbols as random variables. Their true distribution is taken into account: the symbols are assumed zero mean, i.i.d., equiprobable, and with values of the finite alphabet. $f(\boldsymbol{Y}|h) = f(\boldsymbol{Y}|A,h)f(A) = \sum_{A \in \mathcal{A}_p} f(\boldsymbol{Y}|A,h)$, so the SML criterion is:

$$\min_{h, \sigma_v^2} \frac{1}{\sigma_v^2} \sum_{A \in \mathcal{A}_p} \exp\left[ -\frac{1}{\sigma_v^2} \|\boldsymbol{Y} - \mathcal{T}(h)A\| \right]^2 . \tag{1.28}$$

Direct optimization of the SML criterion represents a costly solution. The Expectation–Maximization (EM) [47] algorithm is used to solve SML using the Hidden Markov model (HMM) framework: see [48], for a description of different methods. The EM algorithm will converge to the SML solution given a good initialization. Semi–blind SML is formulated in [49].

### 1.4.5 Further Characterization of the Blind Models

It appears that the more knowledge about the input symbols is incorporated in the model, the better performance one gets. This also comes with an increase in complexity. In the following, we further characterize the different models.

**Identification Indeterminacies** The different methods are also classified according to decreasing severe identification indeterminacies. For example, for complex input constellations, blind deterministic methods can identify the channel up to a complex scale factor, $\hat{h} = \alpha h^o$, with $\alpha \in \mathbb{C}$; in the Gaussian case, the channel can be identified up to a phase factor $\hat{h} = e^{j\varphi} h^o$, with $\varphi \in \mathbb{R}$; FA and stochastic methods can identify the channel up to a discrete–valued phase factor, $\hat{h} = e^{j\varphi} h^o$, with $\varphi$ taking a finite number of discrete values (depending on the symmetry properties of the symbol constellation).

**Robustness to Channel Length Overestimation.** Blind deterministic methods are not robust to channel length overestimation: in general, the different channel lengths have to be tested to detect the right one. The blind Gaussian, FA and stochastic methods will automatically give the right channel order. Note however that the deterministic semi–blind extension will profit from the robustness of TS based methods to channel order overestimation.

**Performance.** The above classification respects the order of increasing performance. The FA methods are particularly powerful: indeed, a performance bound for FA methods corresponds to the case in which all the input symbols would act as training sequence. Computationally less complex methods like [45] are particularly interesting.

In view of the different points mentioned above, one may wonder why we would like to use deterministic methods instead of Gaussian methods and Gaussian methods instead of FA methods. Blind deterministic methods possess the remarkable property of providing, in the noiseless case and with a finite amount of data, the exact channel (apart from indeterminacies). This property is also true for FA methods but not for Gaussian methods, in general (GML is high–SNR consistent). For a finite amount of data, exact second-order statistics cannot be estimated exactly and Gaussian methods will not be able to estimate the channel exactly.

The blind deterministic methods also offer the advantage of allowing closed–form solutions, or convex cost functions, thus avoiding local minima. These methods are one-shot methods (or almost) and so assure a high speed of convergence. For solving blind Gaussian, FA or stochastic techniques, generally iterative and more computationally intensive algorithms need to be used with the risk of falling into local minima if not correctly initialized. This risk is particularly high for the FA techniques: the exploitation of the finite alphabet

leads indeed to highly multimodal cost functions.

## 1.5 Motivation for the Chosen Methods

### 1.5.1 Models

The above discussion was about blind estimation. The associated semi–blind versions inherit from their blind counterpart properties, advantages or disadvantages.

Semi–blind FA methods like [45] are powerful techniques but require a good initialization quality. When the training sequence is too short to give a good channel initialization or when blind methods fail, the initialization may be too bad for the iterative FA methods to work directly. One can instead proceed in smaller steps by first using a semi–blind deterministic method to initialize a semi–blind Gaussian method, which could in turn be used to initialize a semi–blind FA method.

It should be noted that the performance difference for the deterministic and Gaussian models gets smaller as more and more symbols are known. Performance differences are mostly visible in the case of blind methods, especially for ill–conditioned channels.

### 1.5.2 Methods

To develop semi–blind methods, we focused on deterministic and Gaussian ML methods and to a certain extent on FA ML methods for several reasons:

- ML are the most powerful methods.
- They allow to naturally incorporate the knowledge of known input symbols and constitute optimal semi–blind criteria in the sense that no information coming from the known symbols or no blind information is lost.
- When the known symbols are grouped, suboptimal criteria can be found appearing as a linear combination of a training sequence based criterion and a blind criterion. These suboptimal criteria offer the advantage to keep the structure of the blind problem which allows to build fast algorithms.

## 1.6 Thesis Outline and Contributions

The thesis is divided onto three parts. The first one is aimed at determining semi–blind identifiability conditions and performance bounds. It has given rise to a certain number of theoretical studies as detailed below. In the second part, we focus on blind and semi–blind multichannel estimation techniques mainly based on deterministic and Gaussian ML. At last,

in the third part, receiver structures applied to a burst mode transmission of the data are studied. Most of the chapters in this thesis correspond entirely or partly to journal papers in preparation.

### 1.6.1 Part I

**Chapter 2** Blind identifiability conditions are first studied and semi–blind conditions are then derived in terms of channel characteristics, burst length and input symbol characteristics. Identifiability is guaranteed if the parameters can uniquely be determined from the probability density function of the data. To assure semi–blind identifiability, one needs as many known symbols as the number of parameters that blind methods cannot determine. A remarkable property is that this last condition holds even if the known symbols are dispersed all over the burst. For example, a single channel cannot be identified by a blind deterministic method; provided that there are $2N-1$ known symbols present in the burst and not necessarily grouped together, semi–blind techniques will be able to estimate the channel. This result is proved in chapter 5: in fact, we only prove that there is local identifiability and conjecture that there is global identifiability. These results were presented partly in:

> E. de Carvalho and D.T.M. Slock, "Identifiability Conditions for Blind and Semi-Blind Multichannel Estimation," in *European Association for Signal Processing EUSIPCO 98*, Island of Rhodes, Greece, September 1998.

Identifiability conditions for blind and semi-blind multi–user multichannel estimation are give in:

> Luc Deneire, Elisabeth de Carvalho, and Dirk Slock, "Identifiability Conditions for Blind and Semi-Blind Multiuser Multichannel Identification," in *9th IEEE Signal Processing Workshop On Statistical Signal And Array Processing*, Portland, Oregon, USA, September 1998.

**Chapter 3** This short chapter gives theoretical elements on the FIMs and CRBs. Their expression is given for Gaussian distribution data. In this case, local identifiability from the density distribution is equivalent to FIM regularity under certain mild conditions. In blind estimation, not all the parameters can be estimated: we have seen for example that a scale or phase factor cannot be estimated in the deterministic or Gaussian model. This results in singularities in the FIM and the CRB (which is the inverse of the FIM) is not defined. In this chapter, we provide a general study for estimation under constraints. These results are applied to the characterization of blind performance in Chapter 4. In particular, we propose a bound, the pseudo–inverse of the FIM, which gives for a minimal number of independent constraints, the lowest CRB.

**Chapter 4 and Chapter 5** One of the objectives of these chapters is to compare semi–blind to blind and training sequence based estimation through their CRBs. We illustrate the superiority of semi–blind techniques already described in section 1.2. This study was initiated in:

> E. de Carvalho and D.T.M. Slock, "Cramér-Rao Bounds for Semi-blind, Blind and Training Sequence based Channel Estimation," in *Proc. SPAWC 97 Conf.*, Paris, France, April 1997.

### 1.6.2 Part II

**Chapter 6** The formulation of blind and semi–blind DML and GML is undertaken and theoretical performance is derived for an asymptotical number of data (unknown and known input symbols); the case of high SNR is also treated. Although DML is a popular method, its performance has never been derived except at high SNR. We prove that DML does not reach the CRB except at high SNR. The CRB for the Gaussian model of Part I is derived assuming the symbols are Gaussian. Here we compute the performance of GML considering the true symbol distribution. Performance is below the CRB. Simulations show that GML performs better than DML. These results are presented in:

> E. de Carvalho and D.T.M. Slock, "Asymptotic Performance of ML Methods for Semi-Blind Channel Estimation," in *Proc. Asilomar Conference on Signals, Systems & Computers*, Pacific Grove, CA, Nov. 1997.

**Chapter 7** We devote this chapter to fast solutions for solving DML. IQML is a popular method to solve DML: it appears that IQML gives biased estimates and performs poorly at low SNR. We propose two solutions to remedy this situation. The first solution removes an estimate of the noise contribution in the IQML criterion. The second one computes the gradient of DML and at each iteration attempts to null it: this algorithm can also be seen as a form of denoised IQML. We compute the asymptotical performance of DIQML and PQML: PQML performs better than DIQML and has the same asymptotic performance as DML. Some properties of the alternating minimization strategy are also stated and the algorithm is compared to DIQML and PQML. PQML appears to be the best method to solve DML. These blind algorithms as well as their semi–blind extensions treated in Chapter 8, were presented in:

> J. Ayadi, E. de Carvalho, and D.T.M. Slock, "Blind and Semi-Blind Maximum Likelihood Methods for FIR Multichannel Identification," in *Proc. ICASSP 98 Conf.*, Seattle, USA, May 1998.

**Chapter 8** This chapter proposes semi–blind DML based algorithms. We consider the known symbols as grouped. Three (slightly) suboptimal semi–blind criteria are derived based on three different training sequence based criteria. The criteria appear as a linear combination of blind DML and a training sequence criterion. The coefficients of this linear combination are optimal in the ML sense. We also derive a semi–blind criterion based on a linear combination of the (denoised) SRM criterion and the TS criterion to initialize the semi–blind DML based algorithms. Some semi–blind algorithms have been proposed that linearly combine a certain blind criterion with a training sequence based criterion. The right weighting seems difficult to find in that case. We propose a solution to this problem and give a first approach to a subspace fitting based semi–blind criterion. This results are given partly in the Icassp 98 paper, as well as in:

> E. de Carvalho and D.T.M. Slock, "Maximum-Likelihood FIR Multi-Channel Estimation with Gaussian Prior for the Symbols," in *Proc. ICASSP 97 Conf.*, Munich, Germany, April 1997.

An extension of the blind and semi–blind PQML to the multi–user case can be found in:

> Elisabeth de Carvalho, Luc Deneire, and Dirk Slock, "Blind and Semi-Blind Maximum Likelihood Techniques for Multiuser Multichannel Identification," in *European Association for Signal Processing EUSIPCO 98*, Island of Rhodes, Greece, September 1998.

**Chapter 9** GML is interpreted as a form of covariance matrix criterion. The performances of GML and of optimally weighted covariance matching are numerically evaluated and compared: they appear to have the same asymptotic performance. The scoring algorithm is used to solve blind and semi–blind GML which is compared to the DML based methods. We furthermore develop two computationally low algorithms based on approximations of the steepest descent algorithm and of the scoring algorithm for blind GML. Part of these results can be found in:

> E. de Carvalho and D.T.M. Slock, "Semi-Blind Maximum-Likelihood Estimation with Gaussian Prior for the Symbols using Soft Decisions," in *48th Annual Vehicular Technology Conference*, Ottawa, Canada, May 1998.

> E. de Carvalho and D.T.M. Slock, "A Fast Gaussian Maximum-Likelihood Method for Blind Multichannel Estimation," in *Signal Processing Advances in Wireless Communications (SPAWC)*, Annapolis, Maryland, USA, May 1999.

**Chapter 10** The soft decision strategy is particularly well suited to the general semi-blind framework. From a semi–blind channel estimate, an equalizer is built. Hard decisions are taken on the more reliable equalizer outputs, *i.e.* the ones that are the closest to the decision

point: those hard decisions are considered as known. The other non–reliable outputs are left undecided and then remain unknown. A new semi–blind criterion can then be derived based on the augmented number of known symbols. This strategy could be seen as an intermediate step between pure semi–blind and the FA method [45] and could prevent the latter to fall into local minima due to errors in the hard decisions. Unfortunately, this strategy introduces correlations between the known/unknown symbols and the noise: as a result, the semi–blind criterion based on the augmented number of known symbols becomes erroneous. We introduce a way of choosing the reliable symbols which allows partly to alleviate the correlation problem. This idea of soft decisions can be found in the VTC 98 paper previously cited.

### 1.6.3 Part III

**Chapter 11** The optimal structure of burst mode equalizers is derived: the structure of the classical equalizers are derived as well as that of the ISI canceller that uses past and future decisions and that we call Non Causal Decision Feedback Equalizer (NCDFE). The performances of the different equalizers are compared. The optimal equalizer filters are time–varying which implies an increasing complexity w.r.t. continuous processing equalizers. By correctly choosing the number and position of some known symbols, (time-invariant) continuous processing filters applied to burst mode can be organized to give sufficiently good performance, so that optimal (time-varying) burst processing implementation can be avoided. The results of this study can be found in:

> D.T.M. Slock and E. de Carvalho, "Unbiased MMSE decision-feedback equalization for packet transmission," in *Proc. EUSIPCO 96 Conf.*, Trieste, Italy, September 1996.
>
> E. de Carvalho and D.T.M. Slock, "Burst Mode Equalization: Optimal Approach and Suboptimal Continuous–Processing Approximation," Submitted, Signal Processing, Special Issue on Signal Processing Technologies for Short–Burst Wireless Communications.

**Chapter 12** When there are no errors in the non–causal feedback, the NCDFE attains the ISI–free situation and appears to be the most powerful equalizer, representing a less complex alternative to the Viterbi equalizer. As for the classical DFE, errors in the non–causal feedback can cause error propagations. Instead of using hard decisions, we use soft decisions which reduce error propagation. The structure of the NCDFE was derived in the following first paper, along with an implementation of MLSE based on the NCDFE; the soft decision strategy was applied in the second paper.

> D.T.M. Slock and E. de Carvalho, "Burst Mode Non-Causal Decision-Feedback Equalization and Blind MLSE," in *Proc. GLOBECOM 96 Conf.*, London, Great Britain, November 1996.

E. de Carvalho and D.T.M. Slock, "Burst Mode Non-Causal Decision-Feedback Equalizer based on Soft Decisions," in *48th Annual Vehicular Technology Conference*, Ottawa, Canada, May 1998.

**Chapter 13** The usual Matched Filter Bound (MFBs) provides the optimal symbol detection performance of receivers *i.e.* when no ISI is present, when the channel is perfectly known. We propose two MFBs to characterize the optimal performance using reduced-order channel models. These bounds are of interest when the physical channel is infinite and needs to be truncated: for the Viterbi equalizer implementation for example, it may be desirable to reduce the channel length in order to lower the complexity. The associated papers are:

E. de Carvalho and D.T.M. Slock, "Maximum-Likelihood Blind Equalization of Multiple FIR Channels," in *Proc. ICASSP 96 Conf.*, Atlanta, USA, May 1996.

D.T.M. Slock and E. de Carvalho, "Matched Filter Bounds for Reduced-Order Multichannel Models ," in *Proc. GLOBECOM 96 Conf.*, London, Great Britain, November 1996.

# Part I

# Identifiability Conditions and Performance Bounds

Chapter 2

# IDENTIFIABILITY CONDITIONS

*The deterministic and Gaussian models for the unknown symbols are considered here. We investigate the identifiability conditions of blind and semi–blind FIR multichannel estimation in terms of channel characteristics, received data length, input symbol excitation modes as well as number of known symbols for semi–blind estimation. Semi–blind methods appear superior to blind and training sequence methods, and allow the estimation of any channel with only a few known symbols. Furthermore, the Gaussian model appears more robust than the deterministic one as it leads to less demanding identifiability conditions.*

## 2.1 Identifiability Definition

Let $\theta$ be the parameter to be estimated and $\boldsymbol{Y}$ the observations. In the regular cases (*i.e.* in the non blind cases), $\theta$ is called identifiable if [50]:

$$\forall\ \boldsymbol{Y},\quad f(\boldsymbol{Y}|\theta) = f(\boldsymbol{Y}|\theta') \quad \Rightarrow \quad \theta = \theta'\,. \tag{2.1}$$

This definition has to be adapted in the blind identification case because blind techniques can at best identify the channel up to a multiplicative factor $\alpha$: $\alpha \in \mathbb{C}$ in the deterministic model and $|\alpha| = 1$ in the Gaussian model. The identifiability condition (2.1) will be for $\theta$ to equal $\theta'$ up to the blind indeterminacy.

For both deterministic and Gaussian models, $f(\boldsymbol{Y}|\theta)$ is a Gaussian distribution: identifiability in this case means identifiability from the mean and the covariance of $\boldsymbol{Y}$.

## 2.2 Identifiability in the Deterministic Model

In the deterministic model, $\boldsymbol{Y} \sim \mathcal{N}(\mathcal{T}(h)A, \sigma_v^2 I)$ and $\theta = [A_U^T\ h^T]^T$. Identifiability of $\theta$ is based on the mean only; the covariance matrix only contains information about $\sigma_v^2$. $A_U$ and $h$ are identifiable if:

$$\mathcal{T}(h)A = \mathcal{T}(h')A' \Rightarrow$$

$$\begin{cases} A_U = A'_U \text{ and } h = h' & \text{for semi–blind and TS based estimation} \\ A = \dfrac{1}{\alpha}A' \text{ and } h = \alpha h' & \text{for blind estimation} \end{cases} \tag{2.2}$$

with $\alpha$ complex, for a complex input constellation, and real, for a real input constellation. Identifiability is then defined from the noise–free data which we shall denote by $\boldsymbol{X} = \mathcal{T}(h)A$.

### 2.2.1 TS Based Channel Identifiability

We recall here the identifiability conditions for TS based channel estimation. From (1.9), $\mathcal{T}(h)A = \mathcal{A}h$: $h$ is determined uniquely if and only if $\mathcal{A}$ has full column rank, which corresponds to conditions $(i)$ – $(ii)$ below.

**Necessary and sufficient conditions [TS]** *The $m$–channel $\boldsymbol{H}(z)$ is identifiable by TS estimation if and only if*

*(i) Burst Length $M \geq N$.*

*(ii) Number of input symbol modes*[1] *$\geq N$.*

Condition $(i)$ is equivalent to: number of known symbols $M_K \geq 2N-1$. The burst length $M$ is the length of $\boldsymbol{Y}$, expressed in symbol periods.

[1] for a definition of the notion of modes, see for example [19, 18]

### 2.2.2 Blind Channel Identifiability

The deterministic blind identifiability definition (2.2) corresponds to what is called strict identifiability in [51]. The authors of [18, 19] define identifiability based on the Cross–Relation (CR) method: a channel is said CR-identifiable if the channel can be identified uniquely (up to a scale factor) by the noise–free CR method. In [19], identifiability is based on the (complex) FIM matrix: a channel is said identifiable if the FIM has exactly one singularity. In [19, 51], those three identifiability forms were found to be equivalent. [18, 19] give sufficient conditions, and necessary conditions separately for the channel, the burst length and the symbol modes for the CR–identifiability (extended to the FIM and strict identifiability in [19, 51]). In [18], necessary and sufficient conditions on the channel and the modes (but not on the burst length though) are also given and a coupled relation between the channel and the input symbols modes appears, which usefulness is not guaranteed.

We give here necessary and then sufficient conditions for deterministic blind identification in terms of channel characteristics, burst length and input symbol modes. Our original objective was to prove that sufficient conditions [DetB] are also necessary conditions. We have not been able to prove it for the moment, but we highly conjecture that this is true.

**Necessary conditions** *In the deterministic model, the m–channel* $\mathbf{H}(z)$ *and the unknown input symbols* $A_U$ *are blindly identifiable only if*

*(i)* $\mathbf{H}(z)$ *is irreducible.*

*(ii) Burst length* $M \geq N + \left\lceil 2\frac{N-1}{m-1} \right\rceil$.

*(iii) Number of input symbol modes* $\geq N+1$.

*Proof:* $(i)$: If the channel is not irreducible, then $\mathcal{T}(h)$ does not have full column rank. If $A$ is in the null space of $\mathcal{T}(h)$, $\boldsymbol{X} = \mathcal{T}(h)A = 0$ and identifiability is not possible: either $A = 0$ and $h$ cannot be identified, or $A \neq 0$ and $A' = 0$ and any $h'$ verifies $\mathcal{T}(h')A' = 0$. If $A$ is not in the null space of $\mathcal{T}(h)$, we can find $A' \neq 0$ verifying $\mathcal{T}(h)A' = 0$ and $A'' = A + A'$ linearly independent from $A$ verifies $\mathcal{T}(h)A'' = \boldsymbol{X}$. The irreducibility condition is also a necessary condition for the subspace fitting method, which, if the channel is reducible, can only identify its irreducible part.

$(ii)$: Condition $(ii)$ says that the number of equations $(= mM)$ should be greater than the number of unknowns: $Nm-1$ unknowns for $\boldsymbol{H}$, $M+N-1$ for the unknown symbols.

$(iii)$: A proof of condition $(iii)$ can be found in [51].

□

**Sufficient conditions [DetB]** *In the deterministic model, the m–channel* $\mathbf{H}(z)$ *and the*

*input symbols A are blindly identifiable if*

*(i)* $\mathbf{H}(z)$ *is irreducible.*

*(ii) Burst length* $M \geq N + 2\underline{M}$.

*(iii) Number of input symbol modes* $\geq N + \underline{M}$.

*Proof:* see Appendix A.

□

These conditions express the fact that one should have enough data with the right properties to be able to completely describe the signal (or noise) subspace. The proof is based on subspace fitting results. An alternative proof based on linear prediction and blocking equalizers has been given in [52].

Note that the sufficient conditions above are sufficient conditions for the subspace fitting method. A priori, sufficient conditions for identifiability as in (2.1) could be weaker than the sufficient conditions for the subspace fitting method. These conditions appear to be sufficient for all the deterministic methods listed in section 1.4 except for SRM [19].

Note that when $2\underline{M} = \left\lceil 2\frac{N-1}{m-1} \right\rceil$ (which happens in the case $m = 2$), the burst length condition is necessary and sufficient.

### 2.2.3 Semi–Blind Channel Identifiability

Consider the general case of a reducible channel: $\mathbf{H}(z) = \mathbf{H}_I(z)\mathrm{H}_c(z)$. We first give necessary and then sufficient conditions for semi–blind identifiability in the case of grouped known symbols. We denote $\underline{M_I}$ as the smallest $M$ for which $\mathcal{T}_M(h_I)$ has full column rank.

**Necessary conditions** *In the deterministic model, the m–channel* $\mathbf{H}(z)$ *and the unknown input symbols* $A_U$ *are semi–blindly identifiable only if*

*(i) Burst length* $M \geq N_I + \left\lceil \frac{2N - M_K - 1}{m-1} \right\rceil$.

*(ii) Number of grouped known symbols* $M_K \geq 2N_c - 1$.

*Proof:* Condition $(i)$ says that the number of equations $(= mM)$ should be larger than the number of unknowns: $N_I m$ unknowns for $\boldsymbol{H}_I$, $N_c - 1$ unknowns for $H_c$ and $M + N - 1 - M_K$ for the unknown symbols. $\mathrm{H}_c(z)$ and the ambiguous scale factor can only be identified thanks to the known symbols: condition $(ii)$ gives the minimal number of grouped known symbols necessary to identify those parameters.

□

**Sufficient conditions [DetSB]** *In the deterministic model, the m–channel* $\mathbf{H}(z)$ *and the unknown input symbols* $A_U$ *are semi–blindly identifiable if*

*(i) Burst length* $M \geq \max(N_I+2\underline{M}_I, N_c-N_I+1)$

*(ii) Number of excitation modes of the input symbols: at least* $N_I+\underline{M}_I$ *that are not zeros of* $\mathbf{H}(z)$ *(and hence* $\mathrm{H}_c(z)$*).*

*(iii) Grouped known symbols: number* $M_K \geq 2N_c-1$*, with number of excitation modes* $\geq N_c$*.*

*Proof:* See Appendix C.

□

For an irreducible channel, 1 known symbol is sufficient. For a monochannel, $2N-1$ grouped known symbol are sufficient. If $2N-1$ grouped known symbols containing $N$ independent modes are available, condition *(ii)* becomes superfluous.

We do not prove identifiability in the case where the known symbols are not grouped. We conjecture however that identifiability is guaranteed with the same number of known symbols even in that case. Indeed, we show in Chapter 4, that FIM regularity holds under conditions almost similar to [DetSB], which implies local identifiability (result of Chapter 3).

In case the known symbols are dispersed and all equal to 0, the sufficient conditions still hold (except that *(iii)* can be relaxed to $M_K \geq 2N_c-2$) but the channel is now identifiable up to a scale factor only. When those zero known symbols are not sufficiently dispersed however so that at least $N_c$ of them are grouped, it is easy to find configurations in which identification cannot be guaranteed, even up to a scale factor.

### 2.2.4 Semi–Blind Robustness to Channel Length Overestimation

A major disadvantage of the deterministic methods is their non robustness to channel length overestimation. Semi–blind methods allow to overcome this problem. We consider again a reducible channel: $\mathbf{H}(z) = \mathbf{H}_I(z)\mathrm{H}_c(z)$.

**Sufficient conditions [DetSBR]** *In the deterministic model, the m–channel* $\mathbf{H}(z)$ *and the unknown input symbols* $A_U$ *are semi–blindly identifiable when the assumed channel length* $N'$ *is overestimated if*

*(i) Burst length* $M \geq \max(N_I+2\underline{M}_I, 2(N'-N_I+1)-N)$*.*

*(ii) Number of input symbol excitation modes: at least* $N_I+\underline{M}_I$ *that are not zeros of* $H_c(z)$*.*

*(iii) Known symbols:* $M_K \geq 2(N'-N_I)+1$*, grouped.*
*Number of known symbol modes* $\geq N'-N_I+1$*.*

*Proof:* See Appendix D.

□

These results are also valid (with probability one), with the same number of known symbols but now arbitrarily distributed.

## 2.3 Identifiability in the Gaussian Model

### 2.3.1 Gaussian Model

The parameters to be estimated are the channel coefficients and the noise variance: $\theta = [h^T \ \sigma_v^2]^T$. Recall that identifiability is identifiability from the mean and covariance matrix, so identifiability in the Gaussian model implies identifiability in any stochastic model, since such a model can be described in terms of the mean and the covariance plus higher–order moments.

### 2.3.2 Blind Channel Identifiability

In the blind case, $m_Y(\theta) = 0$, so identifiability is based on the covariance matrix only. In the Gaussian model, the channel and the noise variance are said identifiable if:

$$C_{YY}(h, \sigma_v^2) = C_{YY}(h', \sigma_v^{2\prime}) \Rightarrow h' = e^{j\varphi} h, \text{ and } \sigma_v^{2\prime} = \sigma_v^2 \,. \tag{2.3}$$

When the signals are real, the phase factor is a sign, when they are complex, it is a complex unitary number.

Blind identifiability conditions based on the second-order statistics of the noise–free outputs of a FIR multichannel driven by a white stationary input sequence were given in [53, 54]. Only conditions on the channel are given: in [53, 54], a channel is said blindly identifiable up to a phase factor if the channel is irreducible. In fact, it is possible to identify blindly the channel based on the second-order moments even for a reducible channel, it is only not possible to determine if the zeros are minimum or maximum–phase. We give conditions on the channel and the correlation sequence length. (The conditions on the input symbols are that they are white).

#### Irreducible Channel

**Sufficient conditions [GaussB1]** *In the Gaussian model, the m–channel* $\mathbf{H}(z)$ *is identifiable blindly up to a phase factor if*

*(i)* $\mathbf{H}(z)$ *is irreducible.*

*(ii) Burst length* $M \geq \underline{M} + 1$

*Proof:* When condition $(ii)$ is verified, $\mathcal{T}_U(h)$ is (strictly) tall and $\sigma_v^2$ can then be uniquely identified as the minimal eigenvalue of $C_{YY}(\theta)$. $\boldsymbol{H}(z)$ can then be identified up to a phase factor from the denoised covariance matrix $C_{YY}(\theta) - \sigma_v^2 I$ by linear prediction [10]: under conditions $(i)$ and $(ii)$, one can find $P(z)$, the multivariate prediction filter of order $\underline{M}$ and

$\boldsymbol{h}(0)$ (the first coefficient of $\boldsymbol{H}$) up to a phase factor from the denoised covariance matrix, and they are related to $\mathbf{H}(z)$ via the relationship:

$$P(z)\mathbf{H}(z) = \boldsymbol{h}(0)\,. \tag{2.4}$$

This relationship allows to recover uniquely $\mathbf{H}(z)$ from $P(z)$ up to a phase factor.

□

If the noise variance was known, condition $(ii)$ would be $M \geq \underline{M}$. These conditions are also sufficient conditions for the covariance matching method and the Gaussian ML method. Note that not all the non–zero correlations (time 0 to $N-1$) are needed for identification but only the first $\underline{M}+1$.

Identifiability could also have been established from a spectral factorization point of view. The spectral factorization of $S_{YY}(z) = \sigma_a^2 \mathbf{H}(z)\mathbf{H}^\dagger(z)$ is unique provided that $\mathbf{H}(z)$ is irreducible and gives $\mathbf{H}(z)$ up to a unitary constant ($\sigma_a^2$ being known). This point of view however requires the knowledge of the whole non-zero correlation sequence.

### Reducible Channel

Let $\mathbf{H}(z)$ be a reducible channel: $\mathbf{H}(z) = \mathbf{H}_I(z)\mathrm{H}_c(z)$.

**Sufficient conditions [GaussB2]** *In the Gaussian model, the m–channel $\boldsymbol{H}$ is identifiable blindly up to a phase factor if*

*(i)* *$H_c(z)$ is minimum–phase.*

*(ii)* $M \geq \max(\underline{M_I}+1, N_c-N_I+1)$.

*Proof:* Under condition $(ii)$, $\mathcal{T}(h_I)$ is strictly tall and $\sigma_v^2$ can be identified as the minimal eigenvalue of $C_{YY}(\theta)$. The irreducible part $\boldsymbol{H}_I$ can be identified up to a scale factor thanks to the deterministic method described in section 2.2.2 [13] provided that $M \geq \underline{M_I}+1$: let $h_I' = \alpha h_I$ be this estimate of $h_I$. $\left(\mathcal{T}^H(h_I')\mathcal{T}(h_I')\right)^{-1}\mathcal{T}^H(h_I')\left[C_{YY}(\theta)-\sigma_v^2 I\right]\mathcal{T}(h_I')\left(\mathcal{T}^H(h_I')\mathcal{T}(h_I')\right)^{-1} = \sigma_a^2\mathcal{T}(\alpha^{-1}h_c)\mathcal{T}^H(\alpha^{-*}h_c)$. $\alpha^{-1}\mathrm{H}_c(z)$ can now be identified up to a phase factor by spectral factorization provided that $\alpha H_c(z)$ or hence $H_c(z)$ is minimum–phase and $\mathcal{T}(h_c)\mathcal{T}^H(h_c)$ contains the $N_c$ non–zero correlations, *i.e.* $M+N_I-1 \geq N_c$ or $M \geq N_c-N_I+1$.

□

### Monochannel Case

In the monochannel case, the noise variance $\sigma_v^2$ cannot be estimated and so neither $h$. However, if we consider $\sigma_v^2$ as known, the channel can be identified by spectral factorization. The sufficient conditions are for the monochannel to be minimum-phase and the burst to be at least of length $N$.

### 2.3.3 Semi–Blind Channel Identifiability

In the semi–blind case, identifiability is based on the mean and the covariance matrix.

#### Identifiability for any Channel

In the semi–blind case, the Gaussian model presents the advantage to allow identification from the mean only. $m_Y(\theta) = \mathcal{T}_K(h)A_K = \mathcal{A}_K h$: if $\mathcal{A}_K$ has full column rank, $h$ can be identified. The difference with the training sequence case is that in the identification of $h$ from $m_Y(\theta) = \mathcal{T}_K(h)A_K$, the zeros due to the mean of $A_K$ also give information, which lowers the requirements on the number of known symbols. For one non-zero known symbol $a(k)$ (with $0 \leq k \leq M-N$, *i.e.* not located at the edges), the non-zero part of $\mathcal{A}_K$ is $a(k)I_{Nm}$. The Gaussian model appears thus more robust than the deterministic model as it allows identification of any channel, reducible or not, multi or monochannel, with only one non-zero known symbol not located at the edges of the input burst.

**Sufficient conditions [GausSB1]** *In the Gaussian model, the m-channel* $\mathbf{H}(z)$ *is semi–blindly identifiable if*

*(i) Burst length* $M \geq N$.

*(ii) At least one non-zero known symbol* $a(k)$ *not located at the edges* ($0 \leq k \leq M-N$).

#### Identifiability for an Irreducible Channel

**Sufficient conditions [GausSB2]** *In the Gaussian model, the m-channel* $\boldsymbol{H}(z)$ *is semi–blindly identifiable if*

*(i)* $\boldsymbol{H}(z)$ *is irreducible.*

*(ii) At least 1 non-zero known symbol (located anywhere) appears.*

*Proof:* Let us assume that $\boldsymbol{Y}$ contains a block of at least $\underline{M}+1$ samples $\boldsymbol{y}(k)$ that contain only unknown symbols (this implies a condition on the burst length which we do not specify above because it depends on the number of known symbols and their position). Then $h$ can be identified blindly up to a unitary constant from the corresponding covariance matrix as indicated in section 2.2.2: $h' = e^{j\varphi}h$. This unitary scale factor can then be identified thanks to the mean $\mathcal{T}_K^+(h')m_Y = e^{-j\varphi}A_K$: one non-zero element of this quantity suffices to identify $\varphi$.

## 2.4 Conclusions

Identifiability conditions for the two main models studied in the thesis were given in terms of channel characteristics, burst length, input symbol excitation modes and number of known symbols for semi–blind estimation and in the case of grouped known symbols. The semi–blind approach appears more robust than blind estimation, as it allows the estimation of any channel with only a few known symbols. In the deterministic case, 1 known symbol is required for an irreducible channel, $2N_c - 1$ for a reducible channel and $2N - 1$ known symbols for a monochannel. We have also proved that semi–blind methods allow to solve the deterministic non robustness to channel length overestimation. The Gaussian model only requires 1 known symbol (not located at the edges of the burst) and is hence more robust than the deterministic model. Identifiability conditions for multi–user multichannel estimation are given in [55].

## A Proof of Sufficient Conditions [DetB]

To show that conditions [DetB] are sufficient, it is sufficient to prove that $h$ and $A$ can be uniquely identified from the mean $\boldsymbol{X} = \mathcal{T}(h)A$ by a blind method: we prove identifiability by the signal subspace fitting approach.

The signal subspace is defined as the column space of $\mathcal{T}(h)$, for $\mathcal{T}(h)$ tall, and the noise subspace as its orthogonal complement. The signal subspace can be formed from $\boldsymbol{X}$. Indeed, let $\mathcal{X}$ of size $m(\underline{M}+1) \times (M-\underline{M})$ and $\mathcal{A}$ of size $(\underline{M}+N) \times (M-\underline{M})$ defined as:

$$\mathcal{X} = \begin{bmatrix} \boldsymbol{x}(M-1) & \cdots & \boldsymbol{x}(\underline{M}) \\ \vdots & \iddots & \vdots \\ \boldsymbol{x}(M-\underline{M}-1) & \cdots & \boldsymbol{x}(0) \end{bmatrix}, \qquad \mathcal{A}_M = \begin{bmatrix} a(M-1) & \cdots & a(\underline{M}) \\ \vdots & \iddots & \vdots \\ a(M-\underline{M}-N) & \cdots & a(-N+1) \end{bmatrix}. \tag{2.5}$$

and related as

$$\mathcal{X} = \mathcal{T}_{\underline{M}+1}(h)\mathcal{A}\,. \tag{2.6}$$

Conditions $(ii)$ and $(iii)$ are necessary and sufficient for $\mathcal{A}$ to have full row rank: $(ii)$ indicates that $\mathcal{A}$ should have at least as many columns as rows and $(iii)$ that the rows are independent. Given that $\mathcal{A}$ has full row rank, the column space of $\mathcal{X}$ equals the column space of $\mathcal{T}_{\underline{M}+1}(h)$, so we can write in particular:

$$P_{\mathcal{X}}^{\perp} = P_{\mathcal{T}_{\underline{M}+1}(h)}^{\perp} \tag{2.7}$$

where $P_{\mathcal{X}} = \mathcal{X}(\mathcal{X}^H\mathcal{X})^+\mathcal{X}^H$ and $P_{\mathcal{X}}^{\perp} = I - P_{\mathcal{X}}$ are the projection operators on the column space of $\mathcal{X}$ and its orthogonal complement. We are searching for a pair $\hat{h}$, $\widehat{A}$ so that $X = \mathcal{T}_M(\hat{h})\widehat{A}$ or $\mathcal{X} = \mathcal{T}_{\underline{M}+1}(\hat{h})\widehat{\mathcal{A}}$. The matrices $\mathcal{T}_{\underline{M}+1}(\hat{h})$ and $\widehat{\mathcal{A}}$ have the same dimensions as $\mathcal{T}_{\underline{M}+1}(h)$ and $\mathcal{A}$. So the rank of $\mathcal{X}$ equals the column dimension of $\mathcal{T}_{\underline{M}+1}(\hat{h})$ and also the row dimension of $\widehat{\mathcal{A}}$ which hence have full column rank and row rank respectively. Hence

$$P_{\mathcal{X}}^{\perp}\mathcal{T}_{\underline{M}+1}(\hat{h})\widehat{\mathcal{A}} = 0 \;\Rightarrow\; P_{\mathcal{T}_{\underline{M}+1}(h)}^{\perp}\mathcal{T}_{\underline{M}+1}(\hat{h}) = 0 \Leftrightarrow \text{range}\left\{\mathcal{T}_{\underline{M}+1}(\hat{h})\right\} \subset \text{range}\left\{\mathcal{T}_{\underline{M}+1}(h)\right\}\,. \tag{2.8}$$

Now, in appendix B[2] (with $M = \underline{M}+1$ here) it is shown that this implies $\hat{h} = \alpha h$ where $\alpha$ is some complex scalar. Now also $A$ can be estimated up to a scale factor: $\widehat{A} = \left(\mathcal{T}^H(\hat{h})\mathcal{T}(\hat{h})\right)^{-1}\mathcal{T}^H(\hat{h})\boldsymbol{X} = A/\alpha$ (*i.e.* the output of the MMSE zero-forcing equalizer built from $\hat{h}$).

## B Channel Identifiability from the Signal Subspace

**Theorem 1 (Subspace Fitting)** *Let $h$ and $h'$ be causal channel impulse responses of length*

[2]The proof in B is a shorter alternative to the proof in Appendix A of [56], generalized to an extended range of signal space dimension $M$.

*$N$ and $N'$ respectively. If $h$ is irreducible, then for $M > \underline{M}$*

$$range\{\mathcal{T}_M(h')\} \subset range\{\mathcal{T}_M(h)\} \;\Rightarrow\; \begin{cases} \mathbf{H}'(z) = \mathbf{H}(z)\,\alpha(z) & ,\ N' \geq N \\ h' = 0 & ,\ N' < N \end{cases} \tag{2.9}$$

*where $\alpha(z)$ is a scalar polynomial of order $N'-N$.*

*Proof:* range $\{\mathcal{T}_M(h')\} \subset$ range $\{\mathcal{T}_M(h)\}$ implies that there exists a transformation matrix $T$ of size $(M+N-1) \times (M+N'-1)$ such that $\mathcal{T}_M(h') = \mathcal{T}_M(h)\,T$. So $\mathcal{T}_M(h)\,T$ is block Toeplitz and hence

$$\mathcal{T}_{M-1}(h)\,T_{1:M+N-2,1:M+N-2} \;=\; \mathcal{T}_{M-1}(h)\,T_{2:M+N-1,2:M+N-1} \tag{2.10}$$

which implies that $T_{1:M+N-2,1:M+N-2} = T_{2:M+N-1,2:M+N-1}$ since $\mathcal{T}_{M-1}(h)$ has full column rank. Hence $T$ is Toeplitz.

Now, $\mathcal{T}_M(h)$ and $\mathcal{T}_M(h')$ are not only block Toeplitz but also banded. So in particular,

$$0 = [\mathcal{T}_M(h')]_{m+1:mM,1} = \mathcal{T}_{M-1}(h)\,T_{2:M+N-1,1} \tag{2.11}$$

which implies $T_{2:M+N-1,1} = 0$ since $\mathcal{T}_{M-1}(h)$ has full column rank, and

$$0 = [\mathcal{T}_M(h')]_{1:m(M-1),M+N'-1} = \mathcal{T}_{M-1}(h)\,T_{1:M+N-2,M+N'-1} \tag{2.12}$$

which implies $T_{1:M+N-2,M+N'-1} = 0$ since again $\mathcal{T}_{M-1}(h)$ has full column rank. Since $T$ is also Toeplitz, this implies that $T$ is zero if $N' < N$ and is banded with $N'-N+1$ nonzero diagonals if $N' \geq N$. Hence in this last case, the coefficients of $T$ specify a scalar polynomial $\alpha(z)$ of order $N'-N$ such that $\mathbf{H}'(z) = \mathbf{H}(z)\,\alpha(z)$.

To summarize the proof in words, a linear transformation that transforms a linear time-invariant (LTI) filter into a LTI filter can only be a LTI filter. If furthermore the filters are FIR and causal, then the transforming filter can only be causal and FIR of order equal to the difference of the orders of the filters.

□

# C Proof of Sufficient Conditions [DetSB]

The semi–blind problem can be decomposed into a blind problem and a TS problem. Conditions for identifying the part of $\mathbf{H}(z)$ that can be identified blindly up to a scale factor, *i.e.* $\mathbf{H}_I(z)$, and then conditions for identifying by TS the rest, *i.e.* the parameters in $\mathrm{H}_c(z)$ and the scale factor, are derived.

Consider the $m(\underline{M}_I+1) \times (M-\underline{M}_I)$ data matrix $\mathcal{X} = \mathcal{T}_{\underline{M}_I+1}(h_I)\mathcal{T}_{\underline{M}_I+N_I}(h_c)\mathcal{A}$. Then $P_{\mathcal{X}} = P_{\mathcal{T}_{\underline{M}_I+1}(h_I)}$ if and only if $\mathcal{T}_{\underline{M}_I+N_I}(h_c)\mathcal{A}$ has full row rank. Condition $(i)$ expresses that the number of columns of this last quantity should be greater than its number of rows, plus

the fact that in general $M \geq M_K - N + 1$, which gets combined with condition $(iii)$. Let $p$ be the number of modes of $A$ (which are assumed to be unrepeated, the extension to the case of higher multiplicity being straightforward [19]): $a(k) = \sum_{i=1}^{p} \alpha_i z_i^k$. It can be shown that $\mathcal{A}$ can be decomposed as

$$\mathcal{A} = \mathcal{M}_1 \mathcal{M}_2 \mathcal{M}_3 =$$

$$\begin{bmatrix} 1 & \cdots & 1 \\ z_1^{-1} & \cdots & z_p^{-1} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ z_1^{-(\underline{M}_I+N-1)} & \cdots & z_p^{-(\underline{M}_I+N-1)} \end{bmatrix} \begin{bmatrix} \alpha_1 & 0 & \cdots & \cdots & 0 \\ 0 & \alpha_2 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \alpha_p \end{bmatrix} \begin{bmatrix} z_1^{M-1} & z_1^{M-2} & \cdots & z_1^{\underline{M}_I} \\ z_2^{M-1} & z_2^{M-2} & \cdots & z_2^{\underline{M}_I} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ z_p^{M-1} & z_p^{M-2} & \cdots & z_p^{\underline{M}_I} \end{bmatrix} \tag{2.13}$$

so that we can write

$$\mathcal{T}(h_c)\mathcal{A} = \mathcal{B}_1 \mathcal{B}_2 \mathcal{M}_2 \mathcal{M}_3 \qquad \text{with} \tag{2.14}$$

$$\mathcal{T}(h_c)\mathcal{M}_1 = \mathcal{B}_1 \mathcal{B}_2 = \begin{bmatrix} 1 & \cdots & 1 \\ z_1^{-1} & \cdots & z_p^{-1} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ z_1^{-\underline{M}_I-N_I+1} & \cdots & z_p^{-\underline{M}_I-N_I+1} \end{bmatrix} \begin{bmatrix} \mathrm{H}_c(z_1) & 0 & \cdots & \cdots & 0 \\ 0 & \mathrm{H}_c(z_2) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \mathrm{H}_c(z_p) \end{bmatrix} \tag{2.15}$$

If $p \geq \underline{M}_I + N_I$, the rank of $\mathcal{T}(h_c)\mathcal{A}$ is determined by the rank of $\mathcal{B}_2$ and has full row rank if $rank(\mathcal{B}_2) \geq \underline{M}_I + N_I$, *i.e.* $A$ has at least $\underline{M}_I + N_I$ modes which are not zeros of $\mathrm{H}_c(z)$. So under conditions $(i)$ and $(ii)$, we can identify $\hat{h}_I = \alpha h_I$ by subspace fitting.

Now $\left(\mathcal{T}^H(\hat{h}_I)\mathcal{T}(\hat{h}_I)\right)^{-1} \mathcal{T}^H(\hat{h}_I)\boldsymbol{X} = \mathcal{T}(h_c)A/\alpha$. Under conditions $(i)$ and $(iii)$ $h_c$ and the scale factor $\alpha$ get identified by TS estimation.

## D Proof of Sufficient Conditions [DetSBR]

Assume a channel $h'$ of length $N'$ and a symbol sequence $A'$ satisfy $\mathcal{T}_M(h)A = \boldsymbol{X} = \mathcal{T}_M(h')A'$. The sequence $A'$ is of length $M+N'-1$, with its training sequence part synchronized to that of $A$ ($A'_K = A_K$). The channel $h'$ may be reducible so that it can be decomposed in general as $\mathbf{H}'(z) = \mathbf{H}'_I(z)\mathrm{H}'_c(z)$ with $N'_I + N'_c - 1 = N'$. To the irreducible $h'_I$ corresponds a minimum ZF equalizer length $\underline{M}'_I$. Consider

$$\mathcal{X} = \mathcal{T}_{\underline{M}'_I+1}(h_I)\mathcal{T}(h_c)\mathcal{A} = \mathcal{T}_{\underline{M}'_I+1}(h'_I)\mathcal{T}(h'_c)\mathcal{A}' \,. \tag{2.16}$$

Assume for a moment that the conditions are satisfied for $\mathcal{T}(h_c)\mathcal{A}$ to have full row rank; we shall see below what this entails. Then (2.16) implies

$$\text{range}\,\{\mathcal{X}\} = \text{range}\left\{\mathcal{T}_{\underline{M}'_I+1}(h_I)\right\} \subset \text{range}\left\{\mathcal{T}_{\underline{M}'_I+1}(h'_I)\right\} \ . \tag{2.17}$$

According to appendix B, this implies

$$\begin{cases} h_I = 0 & , N'_I > N_I \\ h'_I = \alpha\, h_I & , N'_I = N_I \\ h_I \text{ reducible} & , N'_I < N_I \end{cases} \tag{2.18}$$

Hence necessarily $h'_I = \alpha\, h_I$ and $\underline{M}'_I = \underline{M}_I$ so that $\mathcal{T}(h_c)\mathcal{A}$ has full row rank under conditions $(i) - (ii)$. Since $\mathcal{T}_{\underline{M}_I}(h_I)$ has full column rank, (2.16) implies $\mathcal{T}(h_c)\mathcal{A} = \alpha\mathcal{T}(h'_c)\mathcal{A}'$. Let's denote $h_d = \begin{bmatrix} h_c^T & 0\cdots 0 \end{bmatrix}^T$ and $A_d = \begin{bmatrix} A^T & 0\cdots 0 \end{bmatrix}^T$, $h_d$ and $A_d$ being of the same length as $h'_c$ and $A'$ respectively. Then we can also write

$$\mathcal{T}(h_d)A_d = \mathcal{T}(\alpha h'_c)A' \tag{2.19}$$

where the LHS is known. From this we can identify $\alpha \mathrm{H}'_c(z)$ with $2N'_c{-}1 = 2(N'{-}N_I){+}1$ grouped known symbols and we get $\alpha \mathrm{H}'_c(z) = \mathrm{H}_d(z) = \mathrm{H}_c(z)$. We conclude that $\mathbf{H}'(z) = \mathbf{H}(z)$.

Chapter 3

# CRAMÉR–RAO BOUNDS: THEORETICAL ELEMENTS

*In some estimation problems, not all the parameters can be identified, which results in singularity of the Fisher Information Matrix (FIM). The Cramér–Rao Bound, which is the inverse of the FIM, is then not defined. To regularize the estimation problem, one can impose constraints on the parameters and derive the corresponding CRBs. The correspondence between local identifiability and FIM regularity is studied here. Furthermore the number of FIM singularities is shown to be equal to the number of independent constraints necessary to have a well–defined constrained CRB and local identifiability. In general, many sets of constraints can render the parameters identifiable, giving different values for the CRB, that are not always relevant. When the constraints can be chosen, we propose a constrained CRB, the pseudo-inverse of the FIM, which gives, for a minimum number of constraints, the lowest bound on the mean squared estimation error.*

## 3.1 Introduction

The Cramér-Rao Bound (CRB) is a powerful tool in estimation theory as it gives a lower performance bound for parameter estimation problems. It is computed as the inverse of the Fisher Information Matrix (FIM). When the parameters cannot be completely identified, the FIM is singular, and the classical CRB results cannot be directly applied.

The main underlying motivation of this work is the study of the performance of blind (deterministic and Gaussian) channel estimation problems where the parameters can indeed be identified only up to a scale or phase factor. Blind estimation is done under certain parameter constraints to regularize the problem. The performance of blind methods is not correctly evaluated in general or remains somewhat vague. A constraint often used [56] is to consider one coefficient of the channel as known (which is sufficient to render the estimation problem regular): the resulting performance and its bound depend on the choice of this coefficient and appear arbitrary. One of the contributions of this work will be to give a less arbitrary bound and the corresponding set of constraints. Another motivation comes the comparison we will make later between blind and semi–blind methods through the CRBs. To get a significant comparison, semi–blind and blind CRBs have to be computed under the same constraints. For that purpose, this study, which is valid for the regular or the non regular estimation problem, was then necessary.

The first part of this chapter focuses first on the FIMs and especially the correspondence, for a Gaussian data distribution, between the FIM regularity and the parameter identifiability, defined in terms of probability density function. For the blind channel estimation applications considered here, FIM regularity and local identifiability are equivalent.

In a second step, we study the CRBs for estimation under parameter constraints. A similar study was done in [57] for the case where the unconstrained problem is identifiable, *i.e.* the FIM is regular. We adapt here the results to the case where the unconstrained problem leads to nonidentifiability, *i.e.* the FIM is singular. We furthermore outline the correspondence between the number and characteristics of FIM singularities and the number and characteristics of independent constraints needed in order to regularize the estimation problem and to be able to define the constrained CRB. In a last step, assuming that we can choose the set of constraints, we propose a particular CRB for the case of an unidentifiable unconstrained estimation problem: this CRB is the Moore-Penrose pseudo-inverse of the FIM. It corresponds to a minimum number of particular constraints and gives the lowest bound on the mean squared estimation error, *i.e.* $\mathrm{tr}(CRB)$.

## 3.2 CRBs for Real and Complex Parameters

We assume here the FIMs to be regular.

### 3.2.1 CRBs for Real Parameters

Let $\theta$ be a deterministic real parameter vector and $f(\boldsymbol{Y}|\theta)$ the probability density function of the vector of observations $\boldsymbol{Y}$. The FIM associated with $\theta$ is:

$$\mathcal{J}_{\theta\theta} = \mathrm{E}_{Y|\theta}\left(\frac{\partial \ln f(\boldsymbol{Y}|\theta)}{\partial\theta}\right)\left(\frac{\partial \ln f(\boldsymbol{Y}|\theta)}{\partial\theta}\right)^T . \tag{3.1}$$

Let $\hat{\theta}$ be an unbiased estimate of $\theta$ and $\tilde{\theta} = \hat{\theta} - \theta$ the estimation error. Hence $\mathrm{E}\tilde{\theta} = 0$ and $\mathcal{C}_{\tilde{\theta}\tilde{\theta}} = \mathrm{E}\tilde{\theta}\tilde{\theta}^T$ is the error covariance matrix. When $\mathcal{J}_{\theta\theta}$ is nonsingular and under certain regularity conditions [58], $\mathcal{J}_{\theta\theta}^{-1}$ is the Cramér–Rao Bound:

$$\mathcal{C}_{\tilde{\theta}\tilde{\theta}} \geq CRB = \mathcal{J}_{\theta\theta}^{-1} . \tag{3.2}$$

Equality is achieved if and only if:

$$\hat{\theta} - \theta = \mathcal{J}_{\theta\theta}^{-1}\frac{\partial \ln f(\boldsymbol{Y}|\theta)}{\partial\theta} . \tag{3.3}$$

### 3.2.2 CRB for Complex Parameters, Complex CRB.

When $\theta$ is a complex deterministic parameter, the previous results can be applied to $\theta_R = \left[\mathrm{Re}(\theta^T)\ \ \mathrm{Im}(\theta^T)\right]^T$ and $\boldsymbol{Y}_R = \left[\mathrm{Re}(\boldsymbol{Y}^T)\ \ \mathrm{Im}(\boldsymbol{Y}^T)\right]^T$, the associated real parameters and real observations.

It is however possible to define the FIM for $\theta_R$ w.r.t. complex FIM–like matrices. Let $J_{\varphi\psi}$ be defined as:

$$J_{\varphi\psi} = \mathrm{E}_{Y|\theta}\left(\frac{\partial \ln f(\boldsymbol{Y}|\theta)}{\partial\varphi^*}\right)\left(\frac{\partial \ln f(\boldsymbol{Y}|\theta)}{\partial\psi^*}\right)^H \tag{3.4}$$

where $f(\boldsymbol{Y}|\theta) = f(\boldsymbol{Y}_R|\theta) = f(\boldsymbol{Y}_R|\theta_R)$. Derivation w.r.t. the complex vector $\theta = \alpha + j\beta$ is defined as: $\dfrac{\partial}{\partial\theta} = \dfrac{1}{2}\left(\dfrac{\partial}{\partial\alpha} - j\dfrac{\partial}{\partial\beta}\right)$ (see [58] for some properties of complex derivation). Remark that we denote real and complex FIMs by $\mathcal{J}$ and $J$ respectively.

The parametrization in $(\mathrm{Re}(\theta),\ \mathrm{Im}(\theta))$ is equivalent to a parametrization in terms of $(\theta,\ \theta^*)$ via:

$$\theta_R = \begin{bmatrix} \mathrm{Re}(\theta) \\ \mathrm{Im}(\theta) \end{bmatrix} = \mathcal{M}\begin{bmatrix} \theta \\ \theta^* \end{bmatrix}, \quad \mathcal{M} = \frac{1}{2}\begin{bmatrix} I & I \\ -jI & jI \end{bmatrix} \tag{3.5}$$

where $\mathcal{M}$ is non–singular. Knowing that $J_{\theta\theta} = J^*_{\theta^*\theta^*}$ and $J_{\theta\theta^*} = J^*_{\theta^*\theta}$ (true here as $f_{\boldsymbol{Y}|\theta}(\theta,\theta^*) = f_{\boldsymbol{Y}|\theta}(\theta^*,\theta)$), equation (3.5) implies:

$$\mathcal{J}_{\theta_R\theta_R} = \mathcal{M}\begin{bmatrix} J_{\theta\theta} & J_{\theta\theta^*} \\ J^*_{\theta\theta^*} & J^*_{\theta\theta} \end{bmatrix}\mathcal{M}^H . \tag{3.6}$$

$\mathcal{J}_{\theta_R\theta_R}$ can be determined from $J_{\theta\theta}$ and $J_{\theta\theta^*}$ as follows:

$$\mathcal{J}_{\theta_R\theta_R} = 2 \begin{bmatrix} Re(J_{\theta\theta}) & -Im(J_{\theta\theta}) \\ Im(J_{\theta\theta}) & Re(J_{\theta\theta}) \end{bmatrix} + 2 \begin{bmatrix} Re(J_{\theta\theta^*}) & Im(J_{\theta\theta^*}) \\ Im(J_{\theta\theta^*}) & -Re(J_{\theta\theta^*}) \end{bmatrix} . \tag{3.7}$$

We denote $CRB_R = \mathcal{J}^{-1}_{\theta_R\theta_R}$. To quantify the estimation quality, the quantity of interest will be $\mathrm{tr}(CRB_R)$, *i.e.* the lower bound on the mean squared estimation error, which can be expressed directly in terms of the quantities $J_{\theta\theta}$ and $J_{\theta\theta^*}$. Since $\mathcal{M}\mathcal{M}^H = \frac{1}{2}I$, (3.6) implies:

$$\mathcal{J}^{-1}_{\theta_R\theta_R} = 4\,\mathcal{M} \begin{bmatrix} J_{\theta\theta} & J_{\theta\theta^*} \\ J^*_{\theta\theta^*} & J^*_{\theta\theta} \end{bmatrix}^{-1} \mathcal{M}^H . \tag{3.8}$$

Then:

$$\mathrm{tr}(CRB_R) = \mathrm{tr}(\mathcal{J}^{-1}_{\theta_R\theta_R}) = 4\ \mathrm{tr}\left(J_{\theta\theta} - J_{\theta\theta^*}J^{-*}_{\theta\theta}J^*_{\theta\theta^*}\right)^{-1} . \tag{3.9}$$

**Theorem 2** *When $J_{\theta\theta^*} = 0$, $\mathcal{J}_{\theta_R\theta_R}$ is completely determined by $J_{\theta\theta}$. In that case, $J_{\theta\theta}$ can be considered as the complex FIM and the corresponding complex CRB is such that:*

$$C_{\tilde{\theta}\tilde{\theta}} = E\tilde{\theta}\tilde{\theta}^H \geq CRB = J^{-1}_{\theta\theta} . \tag{3.10}$$

*If $J_{\theta\theta^*} \neq 0$, $J^{-1}_{\theta\theta}$ is also a lower bound on $C_{\tilde{\theta}\tilde{\theta}}$, but not as tight as the (real) CRB, $CRB_R$.*

In that case ($J_{\theta\theta^*} = 0$), a single complex singularity of the complex FIM $J_{\theta\theta}$ corresponds to two real singularities since if $\theta_s$ is a singular vector, then so is $j\theta_s$.

## 3.3 CRBs for a Gaussian Data Distribution

### 3.3.1 Real Parameters

The CRB for a Gaussian data distribution:

$$\begin{aligned} \boldsymbol{Y} \sim \mathcal{N}(m_Y(\theta), C_{YY}(\theta)), \qquad & m_Y(\theta) = \mathrm{E}_{Y|\theta}\left(\boldsymbol{Y}\right) \\ & C_{YY}(\theta) = \mathrm{E}_{Y|\theta}\left(\boldsymbol{Y} - m_Y(\theta)\right)\left(\boldsymbol{Y} - m_Y(\theta)\right)^H \end{aligned} \tag{3.11}$$

is [58]:

$$\mathcal{J}_{\theta\theta}(i,j) = \left(\frac{\partial m_Y^T}{\partial\theta_i}\right) C_{YY}^{-1} \left(\frac{\partial m_Y^T}{\partial\theta_j}\right)^T + \frac{1}{2}\mathrm{tr}\left\{C_{YY}^{-1}\left(\frac{\partial C_{YY}}{\partial\theta_i}\right) C_{YY}^{-1}\left(\frac{\partial C_{YY}}{\partial\theta_j}\right)\right\}, \tag{3.12}$$

where, to simplify, the mean and the covariance matrix are denoted $m_Y$ and $C_{YY}$.

The FIM can also be expressed in a closed form. Let's define $\phi$ to be a vector including the elements of the mean and covariance of the data as:

$$\phi = \begin{bmatrix} m_Y \\ \mathrm{vec}\{C_{YY}\} \end{bmatrix} . \tag{3.13}$$

Using the properties: $\mathrm{tr}\{AB\} = \mathrm{vec}^T\{A^T\}\mathrm{vec}\{B\}$ and $\mathrm{vec}\{ABC\} = (C^T \otimes A)\mathrm{vec}\{B\}$, we find:

$$FIM = \left(\frac{\partial m_Y^T}{\partial \theta}\right) C_{YY}^{-1} \left(\frac{\partial m_Y^T}{\partial \theta}\right)^T + \left(\frac{\partial \mathrm{vec}^T\{C_{YY}\}}{\partial \theta}\right) \left(C_{YY}^{-T} \otimes C_{YY}^{-1}\right) \left(\frac{\partial \mathrm{vec}^T\{C_{YY}\}}{\partial \theta}\right)^T \quad (3.14)$$

or

$$\mathcal{J}_{\theta\theta} = \left(\frac{\partial \phi^T}{\partial \theta}\right) \begin{bmatrix} C_{YY}^{-1} & 0 \\ 0 & C_{YY}^{-T} \otimes C_{YY}^{-1} \end{bmatrix} \left(\frac{\partial \phi^T}{\partial \theta}\right)^T . \quad (3.15)$$

From this expression, the following theorem, also given in [59], is deduced:

**Theorem 3** *The FIM for a Gaussian data distribution is regular if and only if* $\left(\frac{\partial \phi^T}{\partial \theta}\right)$ *has full row rank.*

### 3.3.2 Complex Parameters

In a properly formulated blind channel estimation problem, $\boldsymbol{Y}$ and $\theta$ are simultaneously real or complex. If $\boldsymbol{Y}$ is complex, we shall assume it is circular, *i.e.* $\mathrm{E}\boldsymbol{Y}\boldsymbol{Y}^T = 0$. In that case, it is possible to define a complex Gaussian conditional probability density function for $\boldsymbol{Y}$:

$$f(\boldsymbol{Y}|\theta) = \frac{1}{\pi^{Mm} \det C_{YY}(\theta)} \exp\left[-\left(\boldsymbol{Y} - m_Y(\theta)\right)^H C_{YY}^{-1}(\theta) \left(\boldsymbol{Y} - m_Y(\theta)\right)\right] \quad (3.16)$$

where $m_Y(\theta) = \mathrm{E}_{Y|\theta}(\boldsymbol{Y})$ and $C_{YY}(\theta) = \mathrm{E}_{Y|\theta}\left(\boldsymbol{Y} - m_Y(\theta)\right)\left(\boldsymbol{Y} - m_Y(\theta)\right)^H$. Based on the complex probability density function $f(\boldsymbol{Y}|\theta)$, the computation of the complex FIMs $J_{\theta\theta}$ and $J_{\theta\theta^*}$ gives (straightforward extension of [58]):

$$\begin{cases} J_{\theta\theta}(i,j) = \left(\frac{\partial m_Y^H}{\partial \theta_i^*}\right) C_{YY}^{-1} \left(\frac{\partial m_Y^H}{\partial \theta_j^*}\right)^H + \mathrm{tr}\left\{C_{YY}^{-1} \left(\frac{\partial C_{YY}}{\partial \theta_i^*}\right) C_{YY}^{-1} \left(\frac{\partial C_{YY}}{\partial \theta_j^*}\right)^H\right\} \\ J_{\theta\theta^*}(i,j) = \left(\frac{\partial m_Y^H}{\partial \theta_i^*}\right) C_{YY}^{-1} \left(\frac{\partial m_Y^H}{\partial \theta_j^*}\right)^T + \mathrm{tr}\left\{C_{YY}^{-1} \left(\frac{\partial C_{YY}}{\partial \theta_i^*}\right) C_{YY}^{-1} \left(\frac{\partial C_{YY}}{\partial \theta_j^*}\right)\right\} . \end{cases} \quad (3.17)$$

Proceeding as in the real case, the FIM for $\theta_R$ becomes:

$$\mathcal{J}_{\theta_R\theta_R} = 2\mathcal{M} \begin{bmatrix} \frac{\partial \phi^T}{\partial \theta^*} \\ \frac{\partial \phi^T}{\partial \theta} \end{bmatrix} \left(I_2 \otimes \begin{bmatrix} C_{YY}^{-1} & 0 \\ 0 & C_{YY}^{-T} \otimes C_{YY}^{-1} \end{bmatrix}\right) \begin{bmatrix} \frac{\partial \phi^T}{\partial \theta^*} \\ \frac{\partial \phi^T}{\partial \theta} \end{bmatrix}^H \mathcal{M}^H . \quad (3.18)$$

**Theorem 4** *The FIM for a complex Gaussian data distribution is regular if and only if* $\begin{bmatrix} \frac{\partial \phi^T}{\partial \theta^*} \\ \frac{\partial \phi^T}{\partial \theta} \end{bmatrix}$ *has full row rank.*

## 3.4 Correspondence between Identifiability and FIM Regularity for a Gaussian Data Distribution

### 3.4.1 Regular Estimation

Recall that for a Gaussian distribution, identifiability is based on the mean and covariance of the data: $\theta$ is said identifiable if

$$m_Y(\theta) = m_Y(\theta') \text{ and } C_{YY}(\theta) = C_{YY}(\theta') \quad \Rightarrow \quad \theta = \theta' \, . \tag{3.19}$$

We have local identifiability at $\theta$ if identifiability holds for $\theta'$ being restricted to some open neighborhood of $\theta$. In the Gaussian distribution case, there is a correspondence between FIM regularity and local identifiability.

**Theorem 5** *If $\theta$ is not locally identifiable at $\theta$, then the FIM is singular at $\theta$ [59].*

If a parameter can only be identified up a continuous ambiguity, for example a scale factor for the deterministic model or a phase factor for the Gaussian complex model, it cannot be locally identifiable and the corresponding FIM is singular. However, when the parameter is identifiable up to a discrete ambiguity, like, in the Gaussian model, a sign factor in the real case or the inability to determine if a zero is minimum or maximum phase, local identifiability holds, and the FIM can be non–singular. Under weak conditions, local identifiability implies FIM regularity [59]:

**Theorem 6** *Assume that the FIM is of constant rank in the neighborhood of $\theta$. If $\theta$ is locally identifiable, then the FIM is regular at $\theta$.*

And so we have the following theorem:

**Theorem 7** *Assume that the FIM is locally of constant rank at point $\theta$, then $\theta$ is locally identifiable if and only if the FIM is regular at $\theta$.*

For the deterministic and Gaussian models, we shall see (Chapter 5) that this equivalence holds without the local rank assumption for the FIMs.

### 3.4.2 Blind Estimation

In the deterministic and Gaussian input cases, local blind identifiability will be guaranteed if and only if the FIM has as many singularities as the number of continuous blind ambiguities:

| Number of Continuous Ambiguities | Deterministic Input | Gaussian Input |
|---|---|---|
| real | 1 | 0 |
| complex | 2 | 1 |

Furthermore, there will be as many independent constraints needed as the number of singularities to regularize the estimation problem.

## 3.5 CRBs for Estimation with Constraints

In this section, we consider real parameters (hence $\theta$ stands for $\theta_R$ if $\theta$ is complex). When the estimation is (locally) unidentifiable, the FIM is singular and the classical CRB result (3.2) cannot be applied; *e.g.* the channel cannot be estimated by blind estimation and the CRB is then in fact trivially $+\infty$.

In order to characterize the non–regular estimation performance, we define regularized CRBs for which a certain a priori knowledge on the parameter $\theta$, under the form of a certain set of equality constraints, is assumed: this set of constraints should allow to adjust the parameters that cannot be identified and then to regularize the estimation problem. The introduction of a priori information on $\theta$ requires knowledge of $\theta$ in general, which is not available in practice. However, the point here is to evaluate the estimation performance (*e.g.* to compare different estimation algorithms), which implies that we compare $\hat{\theta}$ to the true $\theta$ which hence needs to be available. The sample estimation error covariance matrix will furthermore be compared to the CRB which also depends on $\theta$.

We determine a CRB for estimation under constraints for both cases of regular and singular unconstrained estimation problems. These results are also used in [60], to compare blind and semi–blind channel estimation performance under the same constraints.

CRBs for parametric estimation under constraints were derived in [57] in the case where the unconstrained estimation problem is regular. A simpler derivation of these results was presented in [61]. The main ingredient of this simpler derivation was used in [62] to give an alternative expression for the CRB in the case where the unconstrained problem is unidentifiable. We shall succinctly restate these results, which appeared already in [63] for the case of linear constraints. In these references, and also here, we shall assume that the constraints are sufficient to regularize the estimation problem, *i.e.* to render the CRB finite. So, consider a $k$–fold constraint of the form:

$$\mathcal{K}_\theta = 0 \tag{3.20}$$

where $\mathcal{K}_\theta : \mathbb{R}^n \to \mathbb{R}^k$ is continuously differentiable and $k < n$, $n$ being the number of parameters in the vector $\theta$. A constrained parameter estimator $\hat{\theta}$ is called unbiased if it satisfies the constraints ($\mathcal{K}_{\hat{\theta}} = 0$) and if the parameter estimation bias is zero for all parameter values that satisfy the constraints [61]. The constrained CRB depends on the constraints only through the tangents to the constraint set at the true value of $\theta$:

$$\mathcal{M}_\theta = \left\{ Z \in \mathbb{R}^n \; ; \;\; Z^T \frac{\partial \mathcal{K}_\theta^T}{\partial \theta} = 0 \right\} \; . \tag{3.21}$$

We note that $\dim(\mathcal{M}_\theta)$ can be larger than $n-k$; the results of [61] can be generalized to the case in which the constraints are not independent. We can introduce a matrix $\mathcal{V}_\theta$ of full

column rank such that $\mathcal{V}_\theta = \left(\frac{\partial \mathcal{K}_\theta^T}{\partial \theta}\right)^\perp$, meaning

$$\text{range}\,\{\mathcal{V}_\theta\} = \mathcal{M}_\theta = \left(\text{range}\left\{\frac{\partial \mathcal{K}_\theta^T}{\partial \theta}\right\}\right)^\perp . \tag{3.22}$$

The key step now [61] is that the unbiasedness leads to a particular correlation between the parameter estimation error and the loglikelihood gradient restricted to $\mathcal{M}_\theta$:

**Lemma 1** *Assume the estimator $\hat{\theta}$ and the true parameter $\theta$ satisfy the constraints, then unbiasedness implies*

$$E\,\mathcal{V}_\theta^T \frac{\partial \ln f(\boldsymbol{Y}|\theta)}{\partial \theta}\,(\hat{\theta}-\theta)^T = \mathcal{V}_\theta^T . \tag{3.23}$$

Using this lemma, the CRB derivation becomes an application of the following theorem.

**Theorem 8 (Cauchy-Schwartz inequality for correlation matrices)** *Let $X_1$ and $X_2$ be random vectors with correlation matrices $R_{ij} = EX_iX_j^T$, $i,j = 1,2$. Assume that $R_{11}$ is nonsingular. Then*

$$E(X_2 - R_{21}R_{11}^{-1}X_1)(X_2 - R_{21}R_{11}^{-1}X_1)^T = R_{22} - R_{21}R_{11}^{-1}R_{12} \geq 0 \tag{3.24}$$

*with equality iff $X_2 = R_{21}R_{11}^{-1}X_1$ in m.s.*

With $X_2 = \hat{\theta} - \theta$ and $X_1 = \mathcal{V}_\theta^T \frac{\partial \ln f(\boldsymbol{Y}|\theta)}{\partial \theta}$, this leads immediately to the following main result.

**Theorem 9 (Constrained CRB)** *Assume the constrained estimator $\hat{\theta}$ to be unbiased ($\hat{\theta}$ and $\theta$ satisfy the constraints $\mathcal{K}_\theta = 0$), then*

$$C_{\tilde{\theta}\tilde{\theta}} \geq CRB_C = \mathcal{V}_\theta \left(\mathcal{V}_\theta^T \mathcal{J}_{\theta\theta} \mathcal{V}_\theta\right)^{-1} \mathcal{V}_\theta^T, \tag{3.25}$$

*with equality iff*

$$\hat{\theta} - \theta = CRB_C \frac{\partial \ln f(\boldsymbol{Y}|\theta)}{\partial \theta} \quad \text{in m.s.} \tag{3.26}$$

*A necessary and sufficient condition for the boundedness of $CRB_C$ is the nonsingularity of $\mathcal{V}_\theta^T \mathcal{J}_{\theta\theta} \mathcal{V}_\theta$.*

### 3.5.1 Interpretations and Remarks

The key points to understand how constrained CRBs work are:

- the constrained CRB depends on the constraints only locally (as the CRB is based on derivatives),

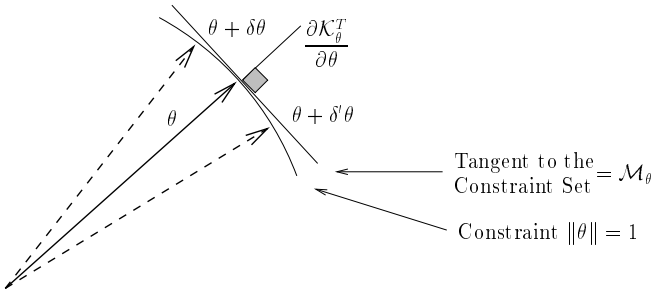



Figure 3.1: Example with constraint $\|\theta\| = 1$.

- locally, the constraint set can be linearized.

To make things clearer, we distinguish between the variable $\theta$ and its true value $\theta^o$. Locally, a point $\theta$ belonging to the constraint set can be approximated as $\theta = \theta^o + \tilde{\theta}$, where $\tilde{\theta}$ belongs to $\mathcal{M}_\theta$, *i.e.* :

$$\theta = \theta^o + \mathcal{V}_{\theta^o} \xi \ . \tag{3.27}$$

In figure 3.1, we show an example with constraint $\|\theta\| = 1$ (for a complex $\theta$ with $n = 1$).

From (3.27) and applying the chain rule, we get

$$\mathcal{J}_{\xi\xi} = \left( \frac{\partial \theta^T}{\partial \xi} \right) \mathcal{J}_{\theta\theta} \left( \frac{\partial \theta^T}{\partial \xi} \right)^T = \mathcal{V}_{\theta^o}^T \mathcal{J}_{\theta\theta} \mathcal{V}_{\theta^o} \ . \tag{3.28}$$

The estimation of $\xi$ is regular provided that $\mathcal{V}_{\theta^o}^T \mathcal{J}_{\theta\theta} \mathcal{V}_{\theta^o}$ is nonsingular. If we now get back to the initial parameter $\theta = \theta^o + \mathcal{V}_{\theta^o} \xi$, using the CRB for a transformation of parameters [58], we find:

$$CRB_{\theta\theta} = \left( \frac{\partial \theta^T}{\partial \xi} \right)^T \mathcal{J}_{\xi\xi}^{-1} \left( \frac{\partial \theta^T}{\partial \xi} \right) = \mathcal{V}_{\theta^o} \left( \mathcal{V}_{\theta^o}^T \mathcal{J}_{\theta\theta} \mathcal{V}_{\theta^o} \right)^{-1} \mathcal{V}_{\theta^o}^T = CRB_C \ . \tag{3.29}$$

We see that the constrained CRB can be interpreted in terms of regular estimation: $\mathcal{V}_\theta^T \mathcal{J}_{\theta\theta} \mathcal{V}_\theta$ can be considered as the FIM for a minimal parameterization $\xi$ of $\theta$, and the results of equivalence mentioned in section 3.4 between FIM and local identifiability could also be applied here.

The $CRB_C$ is independent of the choice of $\mathcal{V}_\theta$ and can in fact also be written as:

$$CRB_C = A_\theta \left( A_\theta^T \mathcal{J}_{\theta\theta} A_\theta \right)^+ A_\theta^T \tag{3.30}$$

where $A_\theta$ is a $n \times q$ matrix, $q \geq \dim(\mathcal{M}_\theta)$, such that $\mathcal{M}_\theta = \text{range}\{A_\theta\}$. In particular, denoting $\frac{\partial \mathcal{K}_\theta^T}{\partial \theta} = \mathcal{K}_\theta'$, we can take $A_\theta = P_{\mathcal{K}_\theta'}^\perp = P_{\mathcal{V}_\theta}$ and obtain

$$CRB_C = P_{\mathcal{V}_\theta} \left( P_{\mathcal{V}_\theta} \mathcal{J}_{\theta\theta} P_{\mathcal{V}_\theta} \right)^+ P_{\mathcal{V}_\theta} = \left( P_{\mathcal{V}_\theta} \mathcal{J}_{\theta\theta} P_{\mathcal{V}_\theta} \right)^+ \ . \tag{3.31}$$

There is a direct correspondence between the number of FIM singularities and the number of constraints necessary to have a finite constrained CRB, which is also the number of constraints necessary to have local identifiability.

**Theorem 10** *For the constrained CRB to be defined, it is necessary and sufficient to fulfill the following conditions.*

- *The number of independent constraints should be at least equal to* $n - r$ *(*$r = rank(\mathcal{J}_{\theta\theta})$*).*
- *At least* $n - r$ *independent columns of* $\dfrac{\partial \mathcal{K}_\theta^T}{\partial \theta}$ *should not be orthogonal to the null space of* $\mathcal{J}_{\theta\theta}$*.*

A constraint of the form $\mathcal{K}_\theta = 0$ has only a local effect and can be locally linearized

**Theorem 11** *The constrained CRB (3.25) can also be interpreted as the CRB under the linear constraint:*

$$\theta^T \left.\frac{\partial \mathcal{K}_\theta^T}{\partial \theta}\right|_{\theta=\theta^o} = \theta^{oT} \left.\frac{\partial \mathcal{K}_\theta^T}{\partial \theta}\right|_{\theta=\theta^o} \tag{3.32}$$

*which means that the components of* $\theta$ *in range* $\left\{ \left.\dfrac{\partial \mathcal{K}_\theta^T}{\partial \theta}\right|_{\theta=\theta^o} \right\}$ *are known.*

### 3.5.2 Minimal Constrained CRB

#### Constraint on the Global Parameter

We assume here that $\mathcal{J}_{\theta\theta}$ is singular. When $\text{range}\{\mathcal{V}_\theta\} = \text{range}\{\mathcal{J}_{\theta\theta}\}$ and since $\mathcal{V}_\theta$ has full column rank, $\mathcal{V}_\theta^T \mathcal{J}_{\theta\theta} \mathcal{V}_\theta$ is regular (minimal number of independent constraints in this case) and the constrained CRB is:

$$CRB_C = \mathcal{J}_{\theta\theta}^+ \, . \tag{3.33}$$

This is a particular constrained CRB: we prove in appendix A that, among all sets of a minimal number of independent constraints, $CRB_C = \mathcal{J}_{\theta\theta}^+$ yields the lowest value for $\text{tr}\{CRB_C\}$. This means that if we want to introduce a priori information in the form of independent constraints, enough to regularize the estimation problem, but not more (minimal number), then all the constraints should concentrate on the unidentifiable part of the parameters ($\text{range}\left\{\dfrac{\partial \mathcal{K}_\theta^T}{\partial \theta}\right\} = \text{null}\{\mathcal{J}_{\theta\theta}\}$) to minimize $\text{tr}\{CRB_C\}$.

#### Constraint on a Parameter Subset

Consider the case of the joint estimation of two parameter vectors $\theta_1$ and $\theta_2$ of length $n_1$ and $n_2$ ($n_1 + n_2 = n$). We are interested in the estimation of $\theta_1$ with $\theta_2$ being nuisance

parameters. The overall parameter vector is $\theta = \left[\theta_1^T \;\; \theta_2^T\right]^T$.

$$\mathcal{J}_{\theta\theta} = \begin{bmatrix} \mathcal{J}_{\theta_1\theta_1} & \mathcal{J}_{\theta_1\theta_2} \\ \mathcal{J}_{\theta_2\theta_1} & \mathcal{J}_{\theta_2\theta_2} \end{bmatrix} . \tag{3.34}$$

We consider the case in which $\mathcal{J}_{\theta\theta}$ is singular but $\mathcal{J}_{\theta_2\theta_2}$ is regular. To regularize the estimation problem, we consider the introduction of (independent) constraints on $\theta_1$ only: $\mathcal{K}_{\theta_1} = 0$, $\mathcal{K}_{\theta_1} : \mathbb{R}^{n_1} \to \mathbb{R}^{k_1}$.

$$\frac{\partial \mathcal{K}_{\theta_1}^T}{\partial \theta} = \begin{bmatrix} \dfrac{\partial \mathcal{K}_{\theta_1}^T}{\partial \theta_1} \\ 0_{n_2,k_1} \end{bmatrix} \tag{3.35}$$

(assumed full column rank). Let $\mathcal{V}_{\theta_1} = \left(\dfrac{\partial \mathcal{K}_{\theta_1}}{\partial \theta_1}\right)^{\perp}$ be a $n_1 \times (n_1 - k_1)$ matrix of full column rank. We can choose

$$\mathcal{V}_{\theta} = \begin{bmatrix} \mathcal{V}_{\theta_1} & 0_{n_1,n_2} \\ 0_{n_2,n_1} & I_{n_2,n_2} \end{bmatrix} . \tag{3.36}$$

The constrained CRB for $\theta$ is:

$$CRB_C = \mathcal{V}_\theta \left(\mathcal{V}_\theta^H \mathcal{J}_{\theta\theta} \mathcal{V}_\theta\right)^{-1} \mathcal{V}_\theta^H = \mathcal{V}_\theta \begin{bmatrix} \mathcal{V}_{\theta_1}^H \mathcal{J}_{\theta_1\theta_1} \mathcal{V}_{\theta_1} & \mathcal{V}_{\theta_1}^H \mathcal{J}_{\theta_1\theta_2} \\ \mathcal{J}_{\theta_2\theta_1} \mathcal{V}_{\theta_1} & \mathcal{J}_{\theta_2\theta_2} \end{bmatrix}^{-1} \mathcal{V}_\theta^H \tag{3.37}$$

and the constrained CRB for $\theta_1$ separately is:

$$CRB_{C,\theta_1} = \mathcal{V}_{\theta_1} \left(\mathcal{V}_{\theta_1}^H \left[\mathcal{J}_{\theta_1\theta_1} - \mathcal{J}_{\theta_1\theta_2} \mathcal{J}_{\theta_2\theta_2}^{-1} \mathcal{J}_{\theta_2\theta_1}\right] \mathcal{V}_{\theta_1}\right)^{-1} \mathcal{V}_{\theta_1}^H = \mathcal{V}_{\theta_1} \left(\mathcal{V}_{\theta_1}^H \mathcal{J}_{\theta_1\theta_1}(\theta) \mathcal{V}_{\theta_1}\right)^{-1} \mathcal{V}_{\theta_1}^H \tag{3.38}$$

where we introduced the notation $\mathcal{J}_{\theta_1\theta_1}(\theta)$ for $\mathcal{J}_{\theta_1\theta_1} - \mathcal{J}_{\theta_1\theta_2} \mathcal{J}_{\theta_2\theta_2}^{-1} \mathcal{J}_{\theta_2\theta_1}$. This notation will be reused below. $\mathcal{J}_{\theta_1\theta_1}^{-1}(\theta)$ would be the unconstrained CRB for $\theta_1$ if $\mathcal{J}_{\theta\theta}$ were regular. Note that with $\mathcal{J}_{\theta_2\theta_2}$ being regular, $\mathcal{J}_{\theta_1\theta_1}(\theta)$ has the same number of singularities as $\mathcal{J}_{\theta\theta}$ in the singular case. Now assume that the constraints are such that range$\{\mathcal{V}_{\theta_1}\}$ = range$\{\mathcal{J}_{\theta_1\theta_1}(\theta)\}$. Then it can be proven that such constraints give the minimal constrained CRB for $\theta_1$,

$$CRB_{C,\theta_1} = \mathcal{J}_{\theta_1\theta_1}^{+}(\theta) \tag{3.39}$$

over all sets of a minimal number of independent constraints on $\theta_1$ only.

## 3.6 Conclusions

This chapter has emphasized on the study of FIMs and CRBs when the estimation problem is not identifiable. There is equivalence between FIM regularity and local identifiability, and there is as many singularities in the FIM as number of continuous ambiguities left in the

estimation of the parameters. The expression for the CRB under constraints has been given and a particular constrained CRB has been derived corresponding to the pseudo–inverse of the FIM. All these results are next applied to the study of blind deterministic and Gaussian models.

# A Minimal CRB

For a minimal number of independent constraints, we prove that $CRB_C = \mathcal{J}_{\theta\theta}^{+}$ is the constrained CRB which gives the lowest value for $\mathrm{tr}\,\{CRB_C\}$ and is attained when $\mathrm{range}\left\{\dfrac{\partial \mathcal{K}_\theta}{\partial \theta}\right\} = \mathrm{null}(\mathcal{J}_{\theta\theta})$.

Let $\mathcal{K}_\theta = 0$ be an arbitrary set of constraints on $\theta$; $\mathcal{V}_\theta = \left(\frac{\partial \mathcal{K}_\theta^T}{\partial \theta}\right)^{\perp}$ has full column rank and we assume that $\mathcal{V}_\theta^T \mathcal{J}_{\theta\theta} \mathcal{V}_\theta$ is invertible. The corresponding constrained CRB is:

$$CRB_C = \mathcal{V}_\theta \left(\mathcal{V}_\theta^T \mathcal{J}_{\theta\theta} \mathcal{V}_\theta\right)^{-1} \mathcal{V}_\theta^T \tag{3.40}$$

Introduce the eigendecomposition of $\mathcal{J}_{\theta\theta} = S_1 \Lambda_1 S_1^T + S_2\, 0\, S_2^T$. In general, $\mathcal{V}_\theta$ has components along $S_1$ and $S_2$: $\mathcal{V}_\theta = S_1 Q_1 + S_2 Q_2$. The fact that the constraints $\mathcal{K}_\theta$ are independent and minimal in number implies that $Q_1$ is square and invertible. Then we obtain

$$\begin{aligned} CRB_C &= \mathcal{V}_\theta \left(\mathcal{V}_\theta^T S_1 \Lambda_1 {S_1}^T \mathcal{V}_\theta\right)^{-1} \mathcal{V}_\theta^T \\ &= \mathcal{V}_\theta \left(Q_1^T \Lambda_1 Q_1\right)^{-1} \mathcal{V}_\theta^T \\ &= \mathcal{V}_\theta Q_1^{-1} \Lambda_1^{-1} Q_1^{-T} \mathcal{V}_\theta^T \\ &= \left(S_1 + S_2 Q_2 Q_1^{-1}\right) \Lambda_1^{-1} \left(S_1 + S_2 Q_2 Q_1^{-1}\right)^T \end{aligned} \tag{3.41}$$

The difference between the $CRB_C$ and $\mathcal{J}_{\theta\theta}^{+} = S_1 \Lambda_1^{-1} {S_1}^H$ may be indefinite in general, however:

$$\mathrm{tr}\,(CRB_C) = \mathrm{tr}\left(\mathcal{J}_{\theta\theta}^{+}\right) + \mathrm{tr}\left(Q_2 Q_1^{-1} \Lambda_1^{-1} Q_1^{-T} Q_2^T\right) \tag{3.42}$$

The second term is non–negative, so $\mathrm{tr}\,(CRB_C) \geq \mathrm{tr}\left(\mathcal{J}_{\theta\theta}^{+}\right)$, with equality iff $Q_2 = 0$, *i.e.* $\mathrm{range}\left\{\dfrac{\partial \mathcal{K}_\theta^T}{\partial \theta}\right\} = \mathrm{null}(\mathcal{J}_{\theta\theta})$ or $\mathrm{range}\,\{\mathcal{V}_\theta\} = \mathrm{range}\,\{\mathcal{J}_{\theta\theta}\}$. In that case $CRB_C = \mathcal{J}_{\theta\theta}^{+}$.

Chapter 4

# CRAMÉR–RAO BOUND FOR BLIND CHANNEL ESTIMATION

*We study here the FIMs and CRBs for blind deterministic and Gaussian channel estimation. We distinguish between the real and complex parameter case since they lead to different FIMs, with different singularities, and require different regularization constraints. The blind deterministic CRB is computed under the commonly used norm constraint which imposes the norm of the channel to be constant. This constraint is sufficient to regularize the problem when the channel is real, but not when it is complex, in which case an additional constraint is required to adjust the phase of the channel. This phase constraint is chosen so that the resulting constrained CRB is the Moore–Penrose pseudo–inverse of the FIM and corresponds to a minimal constrained CRB. When the channel is real the Gaussian FIM is regular, when it is complex however, the FIM is singular: a constraint on the phase is necessary as in the deterministic case and the constrained CRB is again the pseudo–inverse of the FIM.*

## 4.1 Deterministic Model

The deterministic model considers the joint estimation of the unknown input symbols $A$ and the channel coefficients $h$. The parameter vector is $\theta = \left[A^T \; h^T\right]^T$.

### 4.1.1 FIMs

**Circular Complex Input Constellation** As $\boldsymbol{Y}$ is circular, we can work with the complex probability density function of the Gaussian random variable $\boldsymbol{Y} \sim \mathcal{N}(\mathcal{T}(h)A, \sigma_v^2 I)$ (see Chapter 3, section 3.3.2). We denote by $\boldsymbol{X} = \mathcal{T}(h)A$ the signal part of $\boldsymbol{Y}$.

As $J_{\theta\theta^*} = 0$, the complex FIM $J_{\theta\theta}$ is equivalent to the real one $\mathcal{J}_{\theta_R\theta_R}$ and is equal to:

$$J_{\theta\theta} = \frac{1}{\sigma_v^2}\left(\frac{\partial \boldsymbol{X}^H}{\partial \theta^*}\right)\left(\frac{\partial \boldsymbol{X}^H}{\partial \theta^*}\right)^H = \frac{1}{\sigma_v^2}\begin{bmatrix} \mathcal{T}^H(h) \\ \mathcal{A}^H \end{bmatrix}\begin{bmatrix} \mathcal{T}(h) & \mathcal{A} \end{bmatrix} \tag{4.1}$$

because $\dfrac{\partial \boldsymbol{X}^H}{\partial A^*} = \mathcal{T}^H(h)$ and $\dfrac{\partial \boldsymbol{X}^H}{\partial h^*} = \mathcal{A}^H$.

**Real Symbol Constellation** The FIM is the same as in (4.1). This equality of the expressions will allow us to treat the complex and real cases simultaneously.

### 4.1.2 Singularities of the FIMs

Under the blind deterministic identifiability conditions [DetB], $(h, A)$ are identifiable up to a scale factor. This results in one (complex) singularity of the complex FIM (see theorem below). We examine here the singularities in that case. The singularities of the FIM can be considered at the level of $\theta$ (joint estimation of $A$ and $h$) or at the level of $h$ (estimation of $h$ with $A$ considered as nuisance parameter).

**Singularities of $J_{\theta\theta}$.** $J_{\theta\theta} = \dfrac{1}{\sigma_v^2}\left[\mathcal{T}(h) \;\; \mathcal{A}\right]^H\left[\mathcal{T}(h) \;\; \mathcal{A}\right]$ admits as null vector: $\theta_s = \left[-A^T \;\; h^T\right]^T$. Indeed, $\left[\mathcal{T}(h) \;\; \mathcal{A}\right]\left[-A^T \;\; h^T\right]^T = -\mathcal{T}(h)A + \mathcal{A}h = 0$, by exploiting (1.9). When $\mathcal{T}(h)$ and $\mathcal{A}$ have full column rank, the nullity of $J_{\theta\theta}$ is the dimension of the intersection of the column spaces of $\mathcal{T}(h)$ and $\mathcal{A}$, which is one.

**Singularities of $J_{hh}(\theta) \triangleq \dfrac{1}{\sigma_v^2}\mathcal{A}^H P^{\perp}_{\mathcal{T}(h)}\mathcal{A}$.** If $J_{hh}(\theta)$ were regular, its inverse would be the CRB for $h$. However, $J_{hh}(\theta)$ is singular. Indeed, assume that $h'$ is a singular vector of $J_{hh}(\theta)$: $\mathcal{A}^H P^{\perp}_{\mathcal{T}(h)}\mathcal{A}h' = 0$. Then, as $\mathcal{A}$ has full column rank, this means that $\mathcal{A}h' \in$ range$\{\mathcal{T}(h)\}$: there exists an $A'$ such that $\mathcal{A}h' = \mathcal{T}(h')A = \mathcal{T}(h)A'$. This last relation is satisfied for $(h', A') = (h, A)$. Hence, $J_{hh}(\theta)$ has one singularity with $h$ as its singular vector ($\mathcal{A}^H P^{\perp}_{\mathcal{T}(h)}\mathcal{A}h = \mathcal{A}^H P^{\perp}_{\mathcal{T}(h)}\mathcal{T}(h)A = 0$) and the singularity of $J_{hh}(\theta)$ is due to the same

mechanism that leads to the singularity of the global FIM $J_{\theta\theta}$.
In the complex case, $\mathcal{J}_{\theta_R\theta_R}$ will have 2 singularities spanned by:

$$h_{S_1} = \begin{bmatrix} \mathrm{Re}(h) \\ \mathrm{Im}(h) \end{bmatrix} = h_R \quad \text{and} \quad h_{S_2} = \begin{bmatrix} -\mathrm{Im}(h) \\ \mathrm{Re}(h) \end{bmatrix} = \begin{bmatrix} \mathrm{Re}(jh) \\ \mathrm{Im}(jh) \end{bmatrix} . \tag{4.2}$$

The first null vector can be interpreted as corresponding to the ambiguity in the norm of the channel and the second one to the ambiguity in the phase factor.

### 4.1.3 Equivalence between FIM Regularity and Local Identifiability

The link between blind identifiability and FIM singularities in the specific case of the blind deterministic model was already studied in [19, 51]:

**Theorem 12** *For $M \geq 2(N-1)$, or $M \geq N$ for 2 subchannels ($m = 2$), a channel is blindly identifiable up to a scale factor if and only if the complex FIM $J_{\theta\theta}$ has exactly one singularity.*

*Proof:* see [51].

□

In general, there is a correspondence between local identifiability and FIM regularity.

**Theorem 13** *A channel is locally blindly identifiable up to a scale factor if and only if the complex FIM $J_{\theta\theta}$ has exactly one singularity.*

*Proof:* Assume that the FIM has a singular vector $\theta' = \left[h'^T \;\; A'^T\right]^T$ different from $\left[h^T \;\; -A^T\right]^T$:

$$\mathcal{T}(h)A' + \mathcal{T}(h')A = 0 . \tag{4.3}$$

Then for $\epsilon > 0$ arbitrarily small:

$$\begin{aligned} m_Y(\theta + \epsilon\theta') - m_Y(\theta) &= \mathcal{T}(h + \epsilon h')\left[A + \epsilon A'\right] - \mathcal{T}(h)A \\ &= \epsilon\left[\mathcal{T}(h)A' + \mathcal{T}(h')A\right] + O(\epsilon^2) \\ &= O(\epsilon^2) \end{aligned} \tag{4.4}$$

which implies that $\theta$ is not locally blindly identifiable.

Now assume that $\theta$ is not locally blindly identifiable. Then one can find $\Delta h$ and $\Delta A$, where $\|\Delta h\|$ and $\|\Delta A\|$ are arbitrarily small, and not simultaneously colinear with $h$ and $A$ resp. verifying $\mathcal{T}(h)A = \mathcal{T}(h + \Delta h)(A + \Delta A)$. Then:

$$\begin{aligned} \mathcal{T}(h + \Delta h)(A + \Delta h) - \mathcal{T}(h)A &= \mathcal{A}\Delta h + \mathcal{T}(h)\Delta A + O(\|\Delta h\|\|\Delta A\|) \\ &= 0 . \end{aligned} \tag{4.5}$$

This implies that $\left[\Delta h^T \;\; \Delta A^T\right]^T$ is a singular vector of the FIM, non colinear to $\left[h^T \;\; -A^T\right]^T$. □

Using a similar derivation, we can also show the equivalence between the regularity of $\mathcal{V}_\theta^H J_{\theta\theta} \mathcal{V}_\theta$ and local identifiability under the constraint $\mathcal{K}_\theta$ (with definitions of section 3.5):

**Theorem 14** *The estimation problem under constraint $\mathcal{K}_\theta$ is locally identifiable if and only if the regularized FIM $\mathcal{V}_\theta^H J_{\theta\theta} \mathcal{V}_\theta$ is regular.*

The same theorem will hold for the Gaussian model in section 4.2 also but will not be restated there.

### 4.1.4 Regularized Blind CRB

To define a regularized blind CRB, we assume some a priori knowledge. Different forms of a priori knowledge are possible. A technique often used consists in assuming a coefficient of the channel to be known. This is however not robust as performance depends heavily on the choice of this known coefficient (which can be arbitrarily small). The proposed regularized CRB, the Moore–Penrose pseudo–inverse of the FIM, appears less arbitrary and is directly related to the blind scale factor ambiguity.

Blind methods commonly consider the quadratic constraint: $h^H h = 1$ (see [62]). This constraint does not render the problem identifiable: it leaves a sign ambiguity when $h$ is real and a continuous phase ambiguity when $h$ is complex. In the former case, the computation of mean squared error (MSE) assumes the right sign (the right sign could be taken as the sign giving the smallest error). In the complex case however, which phase factor should be chosen? A frequent choice consists in imposing one element of $h$ to be real and positive; again the resulting CRB depends on the choice of this element.

The blind regularized CRB proposed here is computed under the following constraints:

(1) A quadratic constraint:

$$h^H h = h^{o\,H} h^o \tag{4.6}$$

(equivalent to the usual constraint $h^H h = 1$) which allows one to adjust the norm of the channel.

(2) In the complex case, an additional constraint is necessary to adjust the phase factor:

$$h_{S_2}^{o\,T} h_R = h_{S_2}^{o\,T} h_R^o = 0 \; . \tag{4.7}$$

In both real and complex cases, these constraints leave a sign ambiguity which does not lead to FIM singularity. For MSE evaluation, the ambiguity can be resolved by requiring $h^{o\,T} h > 0$.

**Result 1** *Under constraint (4.6) (and (4.7) for the complex case) the CRB for $h$ is the Moore-Penrose pseudo-inverse of $J_{hh}(\theta)$:*

$$CRB_{C,h} = J^{+}_{hh}(\theta) = \sigma_v^2 \left( \mathcal{A}^H P^{\perp}_{\mathcal{T}(h)} \mathcal{A} \right)^{+} . \tag{4.8}$$

*Proof:* Following the notations of section 3.5, the constraint is:

$$\mathcal{K}_{h_R} = \begin{bmatrix} h_R^T h_R - h^{o\,T}_R h^o_R \\ h^{o\,T}_{S_2} h_R \end{bmatrix} = 0 \tag{4.9}$$

leading to

$$\frac{\partial \mathcal{K}^T_{h_R}}{\partial h_R} = \left[ 2h^o_R \;\; h^o_{S_2} \right] . \tag{4.10}$$

As $h_R$ and $h_{S_2}$ are the singular vectors of $\mathcal{J}_{h_R h_R}(\theta)$ (which corresponds to the previously defined complex $J_{hh}(\theta)$), the orthogonal complement of range $\left\{ \dfrac{\partial \mathcal{K}^T_{\theta_R}}{\partial h_R} \right\}$ equals range $\{\mathcal{J}_{h_R h_R}(\theta)\}$. According to section 3.5.2, the CRB under constraint (4.9) is:

$$CRB_{C_R} = \mathcal{J}^{+}_{h_R h_R}(\theta) \tag{4.11}$$

and the corresponding complex contrained CRB can be proven to be:

$$CRB_C = J^{+}_{hh}(\theta) \tag{4.12}$$

□

The choice of the first constraint is not only motivated by its common use. As mentioned in section 3.5.2, this constraint (possibly combined with the linear constraint on the phase) leads to the minimal constrained CRB.

In Figure 4.1, we illustrate the importance of the choice for a constraint by comparing the proposed CRB $\mathrm{tr}\{J^{+}_{hh}(\theta)\}$ to a constrained CRB for which one of the channel coefficients (of varying position) is assumed known. Two channels are tested: a randomly chosen channel and a channel with slowly decreasing coefficients:

$$\boldsymbol{H}_1 = \begin{bmatrix} 0.9477 & -1.1156 & 1.1748 & 1.6455 \\ -0.5257 & -1.5923 & 0.4851 & -0.4542 \end{bmatrix} \tag{4.13}$$

$$\boldsymbol{H}_2 = \begin{bmatrix} 1.0000 & 0.5000 & -0.1500 & 0.0695 \\ 1.5000 & -0.9500 & 0.3050 & 0.0550 \end{bmatrix} \tag{4.14}$$

One observes that when the channel coefficient which is assumed known is small, the corresponding $CRB_C$ can get quite large (arbitrarily large as the coefficient magnitude shrinks).

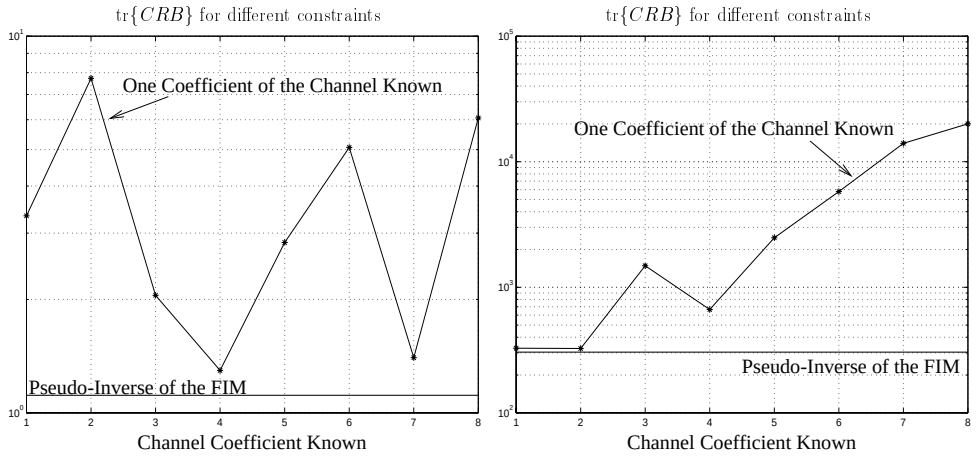



Figure 4.1: Comparison between CRBs with different a priori knowledge. The coefficients designate the coefficients of the vector $h$ for the random channel $\boldsymbol{H}_1$ (left) and the decreasing channel impulse response $\boldsymbol{H}_2$ (right).

**Some Equivalent Constraints** Another choice for the constraint, which leads to the same range $\left\{ \dfrac{\partial \mathcal{K}_{h_R}^T}{\partial h_R} \right\}$ is the following linear constraint:

$$h^{o\,H} h = h^{o\,H} h^o \ . \tag{4.15}$$

This constraint, which leaves no sign ambiguity, corresponds to forcing the components of $h$ in the nullspace of $J_{hh}$ to their true values.

Often, $h$ is estimated under a unit norm constraint $\|\hat{h}\| = 1$, and the scale factor is adjusted in different ways. The following adjustments lead to the same minimal CRB.

- The norm of the channel is adjusted so that $\|\hat{h}\| = \|h^o\|$ and the phase using the phase constraint (4.7). We denote the resulting estimate $\hat{\hat{h}}_{NO}$.
- The scale factor is adjusted in the least–square sense [64, 65]: $\min_\alpha \|h^o - \alpha\hat{h}\|^2$. To be more precise, in this case the trace of the corresponding constrained CRB is $\mathrm{tr}\,\{CRB_C\}$ of equation (4.8).

  *Proof:* The solution of the least-squares problem is $\hat{\hat{h}}_{LS} = \hat{\alpha}\hat{h} = P_{\hat{h}} h^o$. Then, $\Delta\hat{\hat{h}} = P_{\hat{h}} h^o - h^o = -P_{\hat{h}}^\perp h^o$; $C_{\Delta\hat{\hat{h}}\Delta\hat{\hat{h}}} = \mathrm{E}\, P_{\hat{h}}^\perp h^o h^{o\,H} P_{\hat{h}}^\perp$.

$$\begin{aligned}
\mathrm{tr}\left\{C_{\Delta\hat{\hat{h}}\Delta\hat{\hat{h}}}\right\} &= \mathrm{tr}\left\{\mathrm{E}\, P_{\hat{h}}^\perp P_{h^o}\right\}\|h^o\|^2 = \mathrm{tr}\left\{\mathrm{E}\, P_{h^o}^\perp \hat{h}\hat{h}^H\right\}\|h^o\|^2 \\
&= \mathrm{tr}\left\{\mathrm{E}\, P_{h^o}^\perp (\hat{h}\|h^o\|)(\hat{h}\|h^o\|)^H P_{h^o}^\perp\right\} \\
&= \mathrm{tr}\left\{\mathrm{E}\, P_{h^o}^\perp (\Delta\hat{\hat{h}}_{NO})(\Delta\hat{\hat{h}}_{NO})^H P_{h^o}^\perp\right\} \ = \ \mathrm{tr}\left\{P_{h^o}^\perp C_{\Delta\hat{\hat{h}}_{NO}\Delta\hat{\hat{h}}_{NO}} P_{h^o}^\perp\right\} \\
&\geq \mathrm{tr}\left\{P_{h^o}^\perp CRB_{C,h} P_{h^o}^\perp\right\} = \mathrm{tr}\left\{CRB_{C,h}\right\}
\end{aligned}$$

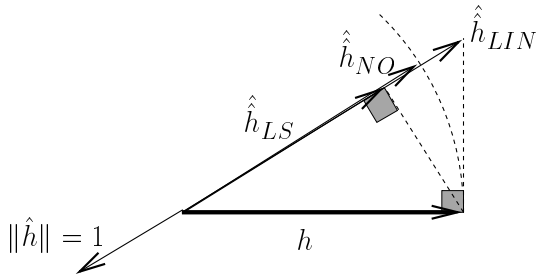



Figure 4.2: Deterministic case: asymptotically equivalent constraints.

□

Another way to adjust the scale factor consists of adjusting $\alpha$ by the following linear constraint $h^{o\,H}\hat{\hat{h}}_{LIN} = h^{o\,H}\alpha\hat{h} = h^{o\,H}h^{o}$, leading to the following channel estimate:

$$\hat{\hat{h}}_{LIN} = \frac{\hat{h}h^{oH}}{h^{o\,H}\hat{h}}h^{o} \ . \tag{4.16}$$

When the estimation of $h$ is consistent, then, asymptotically, the CRB for this constrained channel estimate is the same $CRB_{C,h}$ of (4.8).

In figure 4.2, we show the solutions $\hat{\hat{h}}_{NO}$, $\hat{\hat{h}}_{LS}$, $\hat{\hat{h}}_{LIN}$ for a real channel of length $N = 1$ and with 2 subchannels.

### 4.1.5 Reducible Channel Case

In this case, $\mathbf{H}(z) = \mathbf{H}_I(z)\mathrm{H}_c(z)$ where $\mathrm{H}_c(z)$ is a monic (first coefficient equal to 1) polynomial in $z^{-1}$. In the time domain, this becomes $h = T_c\,h_I$ where $T_c = \mathcal{T}^T_{N_I}(h_c) \otimes I_m$. This decomposition leads us to introduce $A_I = \mathcal{T}(h_c)A$, the input signal filtered by $\mathrm{H}_c(z)$ and we can write the noise-free received signal in the following ways

$$\begin{aligned} \boldsymbol{X} = \mathcal{T}(h)\,A &= \mathcal{T}(h_I)\,\mathcal{T}(h_c)\,A = \mathcal{T}(h_I)\,A_I = \mathcal{A}_I\,h_I \\ &= \mathcal{A}\,h = \mathcal{A}T_c\,h_I = \mathcal{A}_I\,h_I \end{aligned} \tag{4.17}$$

where $\mathcal{A}_I = \mathcal{A}T_c$. Since $\mathcal{T}(h) = \mathcal{T}(h_I)\,\mathcal{T}(h_c)$, we have $P_{\mathcal{T}(h)} = P_{\mathcal{T}(h_I)}$. In the reducible case, the quantities that are blindly identifiable are $h_I$, $A_I$, up to one scalar indeterminacy (assuming certain identifiability conditions for the burst length $M$ and the excitation modes in $A$ in [DetSB]).

To get $h = T_c\,h_I$ from $h_I$, there are $N_c{-}1$ indeterminacies (the coefficients of $h_c$). To get $A$ from $A_I = \mathcal{T}(h_c)A$, there are also $N_c{-}1$ indeterminacies (modes of $A$ that are potentially coinciding with zeros of $\mathrm{H}_c(z)$; alternatively, one needs $N_c{-}1$ "initial conditions" to get $A$ from $A_I$, given $h_c$ (which was already needed to get $h$ from $h_I$)). So there are $2N_c{-}1$ indeterminacies all in all and hence $J_{\theta\theta}$ has $2N_c{-}1$ singularities. Now,

$$J_{hh}(\theta) = \frac{1}{\sigma_v^2}\mathcal{A}^H P^{\perp}_{\mathcal{T}(h)}\mathcal{A} = \frac{1}{\sigma_v^2}\mathcal{A}^H P^{\perp}_{\mathcal{T}(h_I)}\mathcal{A} \tag{4.18}$$

has $N_c$ singularities. Indeed, an alternative decomposition for $h$ is $h = T_I\, h_c$ where $T_I$ is block Toeplitz with $[h_I^T \;\; 0_{1\times(N_c-1)m}]^T$ as first column. Now consider $h' = T_I\, h_c'$ where $h_c'$ is arbitrary (not monic). Then $\mathcal{A}\, h' = \mathcal{T}(h')\, A = \mathcal{T}(h_I)\mathcal{T}(h_c')A$. Hence $J_{hh}(\theta)h' = 0$ and $\text{null}\,\{J_{hh}(\theta)\} = \text{range}\,\{T_I\}$. So we get:

**Result 2** *The CRB for estimating a reducible channel $h$ under the constraint $T_I^{oH}\, h = T_I^{oH}\, h^o$ is*

$$CRB_{C,h} = J_{hh}^{+}(\theta) = \sigma_v^2(\mathcal{A}^H P_{\mathcal{T}(h)}^{\perp}\mathcal{A})^{+}\ . \tag{4.19}$$

Note that this set of constraints implies in particular $h^{oH}h = h^{oH}h^o$. Note also that under these constraints, an estimate $\widehat{h}$ will not necessarily be of the form $\widehat{h} = \widehat{T}_I\,\widehat{h}_c$ with $\widehat{h}_c$ equal to $h_c$ or not: $\widehat{h}$ is not necessarily reducible. Nevertheless, the constraints mentioned are sufficient to make $h$ identifiable. Indeed, identifiability of $h$ with deterministic symbols implies being able to determine $h$ from the noise–free signal. If we do that with for instance the subspace fitting method, then the signal subspace will be $\text{range}\,\{\mathcal{T}(h_I^o)\}$. Subspace fitting will force $\text{range}\left\{\mathcal{T}(\widehat{h})\right\} \subset \text{range}\,\{\mathcal{T}(h_I^o)\}$ which implies $\widehat{\mathbf{H}}(z) = \mathbf{H}_I^o(z)\,\widehat{\mathrm{H}}_c(z)$. The application of the constraints now leads to $\widehat{\mathrm{H}}_c(z) = \mathrm{H}_c^o(z)$ and hence $\widehat{h} = h^o$.

If we want the estimated channel to be reducible, then we have to apply the explicit constraint $h = T_c^o\, h_I$, which can be reformulated as the following implicit constraint: $\mathcal{K}_{\theta_1} = P_{T_c^o}^{\perp}h = 0$ $(\theta_1 = h)$. It turns out that in this case of deterministic input symbols, we can remain working in the complex domain, which we shall do. So we get $\dfrac{\partial \mathcal{K}_{\theta_1}^H}{\partial \theta_1^*} = P_{T_c^o}^{\perp}$ and we can take $\mathcal{V}_{\theta_1} = P_{T_c^o}$. Hence, the constrained CRB for $h$ satisfying the constraints $P_{T_c^o}^{\perp}h = 0$ (compare to (3.31)) and $h^{oH}h = h^{oH}h^o$ is

$$CRB_{C,h} = \sigma_v^2(P_{T_c}\,\mathcal{A}^H P_{\mathcal{T}(h)}^{\perp}\mathcal{A}\,P_{T_c})^{+} = (P_{T_c}\,J_{hh}(\theta)\,P_{T_c})^{+} \tag{4.20}$$

which in general will be smaller than $J_{hh}^{+}(\theta)$ since more a priori information is introduced (in the form of $m(N_c-1)+1$ constraints, compared to the minimal number of $N_c$ constraints to ensure identifiability).

## 4.2 Gaussian Model

In the Gaussian model, the estimation of $h$ is not decoupled from the estimation of $\sigma_v^2$ and the estimation parameter is $\theta = \left[h^T \;\; \sigma_v^2\right]^T$. Unlike in the deterministic model, as $J_{\theta\theta^*} \neq 0$, we cannot treat the complex and real constellations together.

### 4.2.1 FIMs

**Circular Complex Symbol Constellation** When the input constellation is complex, the FIM computation is based on the complex probability density function of $\boldsymbol{Y}$:

$$\boldsymbol{Y} \sim \mathcal{N}(m_Y, C_{YY}), \quad \text{with} \quad C_{YY} = \sigma_a^2 \mathcal{T}(h)\mathcal{T}^H(h) + \sigma_v^2 I, \quad m_Y = 0\,. \tag{4.21}$$

Let $h_R = [\mathrm{Re}(h^T)\ \mathrm{Im}(h^T)]^T$ and $\overline{\theta}_R = [h_R^T\ \sigma_v^2]^T$, the real parameter vector. As $J_{\theta\theta^*}$ is non zero, we cannot consider the complex CRB anymore: the real FIM $\mathcal{J}_{\overline{\theta}_R\overline{\theta}_R}$ is determined via (3.7) thanks to the quantities:

$$J_{\theta\theta}(i,j) = \mathrm{tr}\left\{ C_{YY}^{-1} \left( \frac{\partial C_{YY}}{\partial \theta_i^*} \right) C_{YY}^{-1} \left( \frac{\partial C_{YY}}{\partial \theta_j^*} \right)^H \right\} \tag{4.22}$$

$$J_{\theta\theta^*}(i,j) = \mathrm{tr}\left\{ C_{YY}^{-1} \left( \frac{\partial C_{YY}}{\partial \theta_i^*} \right) C_{YY}^{-1} \left( \frac{\partial C_{YY}}{\partial \theta_j^*} \right) \right\} \tag{4.23}$$

$$\text{where:} \begin{cases} \dfrac{\partial C_{YY}}{\partial h_i^*} = \sigma_a^2 \mathcal{T}(h)\mathcal{T}^H \left( \dfrac{\partial h}{\partial h_i^*} \right) \\ \dfrac{\partial C_{YY}}{\partial \sigma_v^2} = \dfrac{1}{2} I\,. \end{cases} \tag{4.24}$$

**Real Symbol Constellation** When the input constellation is real, the FIM is:

$$\mathcal{J}_{\theta\theta}(i,j) = \frac{1}{2}\mathrm{tr}\left\{ C_{YY}^{-1} \left( \frac{\partial C_{YY}}{\partial \theta_i} \right) C_{YY}^{-1} \left( \frac{\partial C_{YY}}{\partial \theta_j} \right)^T \right\} \tag{4.25}$$

$$\begin{cases} \dfrac{\partial C_{YY}}{\partial h_i} = \sigma_a^2 \mathcal{T}(h)\mathcal{T}^T \left( \dfrac{\partial h}{\partial h_i} \right) + \sigma_a^2 \mathcal{T} \left( \dfrac{\partial h}{\partial h_i} \right) \mathcal{T}^T(h) \\ \dfrac{\partial C_{YY}}{\partial \sigma_v^2} = I\,. \end{cases} \tag{4.26}$$

### 4.2.2 FIM singularities

**Circular Complex Symbol Constellation** Under the Gaussian blind identifiability conditions [GausB], a complex channel $h$ is identifiable up to a phase factor. This corresponds to one singularity of the global FIM $\mathcal{J}_{\overline{\theta}_R\overline{\theta}_R}$:

$$\mathcal{J}_{\overline{\theta}_R\overline{\theta}_R} = \begin{bmatrix} \mathcal{J}_{h_R h_R} & \mathcal{J}_{h_R \sigma_v^2} \\ \mathcal{J}_{\sigma_v^2 h_R} & \mathcal{J}_{\sigma_v^2 \sigma_v^2} \end{bmatrix} \tag{4.27}$$

as well as of:

$$\mathcal{J}_{h_R h_R}(\overline{\theta}_R) = \mathcal{J}_{h_R h_R} - \mathcal{J}_{h_R \sigma_v^2} \left( \mathcal{J}_{\sigma_v^2 \sigma_v^2} \right)^{-1} \mathcal{J}_{\sigma_v^2 h_R}\,. \tag{4.28}$$

$\mathcal{J}^{-1}_{h_R h_R}(\overline{\theta}_R)$ would be the unconstrained CRB for $h$ if its estimation were regular. The null space of $\mathcal{J}_{h_R h_R}(\overline{\theta}_R)$ is spanned by

$$h_S = \left[-\mathrm{Im}(h^T) \;\; \mathrm{Re}(h^T)\right]^T = h_{S_2} \; . \tag{4.29}$$

**Real Symbol Constellation** The real FIM $\mathcal{J}_{\theta\theta}$ is regular under the identifiability conditions, as well as $\mathcal{J}_{hh}(\theta)$.

### 4.2.3 Equivalence between FIM Regularity and Local Identifiability

**Theorem 15** *The (complex or real) FIM is singular if and only if there exist a vector $h'$ (complex or real) and a scalar $\sigma_v^{2\,'}$ such that:*

$$\sigma_a^2 \mathcal{T}(h)\mathcal{T}^H(h') + \sigma_a^2 \mathcal{T}(h')\mathcal{T}^H(h) + \sigma_v^{2\,'} I = 0 \; . \tag{4.30}$$

*Proof:*

**Complex case:** The complex FIM is singular if there exists a $\overline{\theta}'_R = \left[\mathrm{Re}(h'^T) \;\; \mathrm{Im}(h'^T) \;\; \sigma_v^{2\,'}\right]^T$, such that:

$$\mathcal{J}_{\overline{\theta}_R \overline{\theta}_R} \overline{\theta}'_R = 0 \tag{4.31}$$

$$\Leftrightarrow \quad \left[ \left(\frac{\partial \mathrm{vec}^T\{C_{YY}\}}{\partial h^*}\right)^T \quad \left(\frac{\partial \mathrm{vec}^T\{C_{YY}\}}{\partial h}\right)^T \quad \left(\frac{\partial \mathrm{vec}^T\{C_{YY}\}}{\partial \sigma_v^2}\right)^T \right] \begin{bmatrix} h' \\ h'^* \\ \sigma_v^{2\,'} \end{bmatrix} = 0 \tag{4.32}$$

$$\Leftrightarrow \quad \sum_j \left(\frac{\partial C_{YY}}{\partial h_j^*}\right)^H h'_j + \sum_j \left(\frac{\partial C_{YY}}{\partial h_j^*}\right) h'^*_j + \frac{1}{2}\sigma_v^{2\,'} I = 0 \; . \tag{4.33}$$

We have: $\dfrac{\partial C_{YY}}{\partial h_j^*} = \sigma_a^2 \mathcal{T}(h)\mathcal{T}^H\left(\dfrac{\partial h}{\partial h_i}\right)$ and $\displaystyle\sum_j \mathcal{T}\left(\frac{\partial h}{\partial h_i}\right) h'_j = \mathcal{T}(h')$, then:

$$(4.31) \quad \Leftrightarrow \quad \sigma_a^2 \mathcal{T}(h')\mathcal{T}^H(h) + \sigma_a^2 \mathcal{T}(h)\mathcal{T}^H(h') + \frac{1}{2}\sigma_v^{2\,'} I = 0 \; . \tag{4.34}$$

This is equivalent to equation (4.30) (with $\frac{1}{2}\sigma_v^{2\,'} \rightarrow \sigma_v^{2\,'}$). □

**Real case:** The real FIM matrix is singular if there exists a $\theta' = \left[h'^T \;\; \sigma_v^{2\,'}\right]^T$, such that:

$$\mathcal{J}_{\theta\theta}\theta' = 0 \tag{4.35}$$

$$\Leftrightarrow \quad \left[ \left(\frac{\partial \mathrm{vec}^T\{C_{YY}\}}{\partial h}\right)^H \quad \left(\frac{\partial \mathrm{vec}^T\{C_{YY}\}}{\partial \sigma_v^2}\right)^H \right]^H \begin{bmatrix} h' \\ \sigma_v^{2\,'} \end{bmatrix} = 0 \tag{4.36}$$

$$\Leftrightarrow \quad \sum_j \left(\frac{\partial C_{YY}}{\partial h_j}\right)^H h'_j + \sigma_v^{2'} I = 0 \ . \tag{4.37}$$

We have $\dfrac{\partial C_{YY}}{\partial h_j} = \sigma_a^2 \mathcal{T}(h)\mathcal{T}^H\left(\dfrac{\partial h}{\partial h_i}\right) + \sigma_a^2 \mathcal{T}^H\left(\dfrac{\partial h}{\partial h_i}\right)\mathcal{T}^H(h)$. Then:

$$(4.35) \quad \Leftrightarrow \quad \sigma_a^2 \mathcal{T}(h')\mathcal{T}^H(h) + \sigma_a^2 \mathcal{T}(h)\mathcal{T}^H(h') + \sigma_v^{2'} I = 0 \ . \tag{4.38}$$

□

From (4.30), we can deduce the following theorem.

**Theorem 16** *The real/complex channel is locally blindly identifiable if and only if the FIM is regular/1–singular.*

Note that locally a complex channel is identifiable up to a continuous phase factor but a real channel is locally identifiable strictly speaking.

*Proof:* Assume that the FIM has a null vector $\theta' = [h'^T \ \ \sigma_v^{2'}]^T$ which in the complex channel case is non colinear to $h_S$. Then theorem 15 says that $\theta'$ satisfy (4.30). Now, with $\epsilon > 0$ arbitrarily small,

$$\begin{aligned} C_{YY}(\theta+\epsilon\theta') - C_{YY}(\theta) &= \left(\sigma_a^2 \mathcal{T}(h+\epsilon h')\mathcal{T}^H(h+\epsilon h') + \left(\sigma_v^2 + \epsilon\sigma_v^{2'}\right) I\right) - \left(\sigma_a^2 \mathcal{T}(h)\mathcal{T}^H(h) + \sigma_v^2 I\right) \\ &= \sigma_a^2 \mathcal{T}(h)\mathcal{T}^H(\epsilon h') + \sigma_a^2 \mathcal{T}(\epsilon h')\mathcal{T}^H(h) + \epsilon\, \sigma_v^{2'} I + O(\epsilon^2) = O(\epsilon^2) \ . \end{aligned} \tag{4.39}$$

This means that the covariance matrix is locally constant in the direction of $\theta'$ around $\theta$. Similarly to the proof of theorem 13, one can show that if the channel is identifiable, the FIM is regular or 1–singular.

□

In appendix A, we study the conditions on the characteristics of the channel to have local identifiability. The results are contained in the theorem below. The channel is assumed reducible: $\mathbf{H}(z) = \mathbf{H}_I(z)\mathrm{H}_c(z)$.

**Theorem 17** *The Gaussian FIM for a real/complex multichannel is regular/1–singular and the channel is locally blindly identifiable if:*

*(1)* $M \geq \max(\underline{M}_I + 1, N_c - 1)$,

*(2) the channel has no conjugate reciprocal zeros, i.e. there exists no* $z_o \in \mathbb{R}/\mathbb{C}$ *such that* $\mathbf{H}(z_o) = \mathbf{H}(1/z_o^*) = 0$.

*Proof:* appendix A.

□

The no conjugate reciprocal zeros condition was also given in [38], but for the real channel case only, without mentioning that the complex case is singular in any case. Remark that,

in particular, the Gaussian FIM is regular if there are arbitrary zeros (not in conjugate reciprocal pairs) due to the fact that a minimum phase channel is identifiable (example of local identifiability).

The monochannel case is treated in appendix A: the results mentioned above for a multichannel are valid here also except that the noise variance $\sigma_v^2$ cannot be identified, which results in an additional singularity of the FIM when the channel has no conjugate reciprocal zeros (when the channel has conjugate reciprocal zeros, there is no additional singularity).

### 4.2.4 Regularized Blind CRBs

**Complex Symbol Constellation** As in the deterministic case, we need to define a regularized CRB, by introducing some a priori knowledge on the parameters, allowing us to determine the ambiguous phase factor. We assume that the channel is (blindly) identifiable: we do not treat the monochannel or conjugate reciprocal zeros.

The estimation of $h_R$ is considered under the constraint:

$$h_{S_2}^{o\,T} h_R = 0 \tag{4.40}$$

which leads to the constrained CRB for $h_R$:

$$CRB_{C,h_R} = \mathcal{J}^{+}_{h_R h_R}(\theta) \; . \tag{4.41}$$

This linear constraint does not allow to estimate the phase factor completely and a sign ambiguity is left but not reflected in the FIM singularities as it is a discrete ambiguity. For MSE computation purposes, the sign ambiguity can be resolved by requiring $h_R^{o\,T} h_R > 0$, which together with (4.40) can be stated as $h^{o\,H} h > 0$.

**Real Symbol Constellation** No regularization is necessary and the CRB is $\mathcal{J}_{hh}^{-1}(\theta)$. To compare the MSE for an estimator to this CRB, the knowledge of the right sign and right phase of the zeros (*e.g.* minimum phase in the reducible case) should be used.

## 4.3 Conclusions

In this chapter, we have focused on the FIMs and CRBs for blind deterministic and Gaussian estimation. The singularities of the FIM and local identifiability conditions have been studied. For deterministic estimation, a norm constraint on the channel have been imposed. A phase constraint, often ignored, has also been chosen for the deterministic and Gaussian case, such that the resulting CRBs are the pseudo–inverse of the FIMs for the channel and correspond to the minimum CRBs for a minimum number of independent constraints.

# A Local Identifiability Conditions for the Gaussian Model

In this appendix, we study the solutions $(h', \sigma_v^{2'})$ of the equation:

$$\mathcal{T}(h)\mathcal{T}^H(h') + \mathcal{T}(h')\mathcal{T}^H(h) + \sigma_v^{2'} I = 0 \ . \tag{4.42}$$

We first treat the monochannel case for a complex or real channel, which allows us to then treat the multichannel case.

## A.1 Complex Monochannel

We assume that $M \geq N-1$; in this case, equation (4.42) can be written in the $z$–domain as:

$$\sigma_a^2 \mathrm{H}(z)\mathrm{H}'^{\dagger}(z) + \sigma_a^2 \mathrm{H}'(z)\mathrm{H}^{\dagger}(z) + \sigma_v^{2'} = 0 \tag{4.43}$$

where $\mathrm{H}^{\dagger}(z) = \mathrm{H}^H(1/z^*)$. Let's denote $p(z) = \mathrm{H}(z)\mathrm{H}'^{\dagger}(z)$, then:

$$(4.43) \Leftrightarrow p(z) + p^{\dagger}(z) + \sigma_v^{2'} = 0 \ . \tag{4.44}$$

**Solutions of the form $[* \ \ * \ \ 0]^T$: $\sigma_v^{2'} = 0$.**

$$p(z) = \sum_{i=-d_p}^{d'_p} \alpha_i z^{-i} \text{ and } p^{\dagger}(z) = \sum_{i=-d'_p}^{d_p} \alpha^*_{-i} z^{-i} \tag{4.45}$$

$$p(z) + p^{\dagger}(z) = 0 \Rightarrow d'_p = d_p \quad (\text{and } \alpha_i = -\alpha^*_{-i}) \ . \tag{4.46}$$

As $\mathrm{H}(z)$ and $\mathrm{H}^{\dagger}(z)$ are respectively causal and anticausal, $\deg(\mathrm{H}(z)) = \deg(\mathrm{H}'(z)) = N-1 = d_p$. In the following, we assume that $\mathrm{H}(z)$ is monic. Equation (4.44) is also equivalent to:

$$p(z) = -p^{\dagger}(z) \ . \tag{4.47}$$

From this equation, we can deduce that if $z_o$ is a zero of $p(z)$, so is $1/z_o^*$, which implies that $p(z)$ is of the form:

$$p(z) = \alpha \prod_{i=0}^{N_1-1} (1 - z_i z^{-1})(1 - z_i^* z) \ \left[(1 - z^{-1})(1 + z)\right]^{N_2} \ . \tag{4.48}$$

where $N_1 + N_2 = N-1$. We will differentiate the zeros that are equal to 1 or $-1$: $\{z_i\}_{i=1:N_1-1}$ are different from 1 or $-1$. $z$ equals 1 or $-1$.

The number of singularities depends on the characteristics of the channel $\mathrm{H}(z)$ and namely the presence of conjugate reciprocal zeros.

(1) $\mathrm{H}(z)$ has no conjugate reciprocal zeros:

The $N-1$ zeros of $\mathrm{H}(z)$ are among the zeros of $p(z)$, this implies that $p(z)$ has no zeros equal to 1 or $-1$:

$$p(z) = \alpha \prod_{i=0}^{N-2} (1 - z_i z^{-1}) \prod_{i=0}^{N-2} (1 - z_i^* z) = \mathrm{H}(z)\mathrm{H}'^{\dagger}(z) \tag{4.49}$$

furthermore, without loss of generality, we can assume that:

$$\mathrm{H}(z) = \prod_{i=0}^{N-2} (1 - z_i z^{-1}) \ . \tag{4.50}$$

In that case:

$$\mathrm{H}'^{\dagger}(z) = \alpha \prod_{i=0}^{N-1} (1 - z_i^* z) = \alpha \mathrm{H}^{\dagger}(z) \tag{4.51}$$

$$(4.44) \Rightarrow \mathrm{H}'(z) = j\mathrm{H}(z) \ . \tag{4.52}$$

The FIM is 1-singular. Its null space is spanned by $\left[-\mathrm{Im}^T(h) \ \ \mathrm{Re}^T(h)\right]^T$.

(2) $\mathrm{H}(z)$ has 1 pair of conjugate reciprocal zeros: $(z_o, 1/z_o^*)$, $z_o \neq 1$, $z_o \neq -1$.

Again, without loss of generality, we can assume that:

$$\mathrm{H}(z) = (1 - z_o z^{-1})(1 - z_o^* z) z^{-1} z_o^{-*} \underbrace{\prod_{i=1}^{N-3} (1 - z_i z^{-1})}_{\mathrm{H}_1(z)} \ . \tag{4.53}$$

There are 2 degrees of freedom in $\mathrm{H}'(z)$ coming from the fact that $\mathrm{H}'(z)$ can admit 1 and $-1$ as zeros or not. Two possible choices for $\mathrm{H}'(z)$ are then:

$$\begin{cases} \mathrm{H}'(z) = j(1 - z_{N-1} z^{-1})(z^{-1} - z_{N-1}^*) z_o \ \mathrm{H}_1(z) \ , \\ \mathrm{H}'(z) = (1 - z^{-1})(z^{-1} - 1) z_o \ \mathrm{H}_1(z). \end{cases} \tag{4.54}$$

The FIM has 2 singularities coming from the pair of conjugate reciprocal zeros, and 1 singularity corresponding to $j\mathrm{H}(z)$.

(3) $\mathrm{H}(z)$ has one zero equal to 1 or $-1$.

We assume that this zero is equal 1. $\mathrm{H}'(z)$ can be chosen as:

$$\mathrm{H}(z) = (1 - z^{-1}) \underbrace{\prod_{i=0}^{N-3} (1 - z_i z^{-1})}_{\mathrm{H}_1(z)} \ , \ \mathrm{H}'(z) = (1 + z^{-1}) \mathrm{H}_1(z) \ . \tag{4.55}$$

(4) $\mathrm{H}(z)$ has several conjugate reciprocal zeros:

Then, to each pair of conjugate reciprocal zeros different from 1 and $-1$, correspond 2 singularities, and to each zero equal to 1 or $-1$ corresponds 1 singularity.

**Solutions of the form** $[*\ \ *\ \ \sigma_v^{2\prime}]^T$: $\sigma_v^{2\prime} \neq 0$

(1) $\mathrm{H}(z)$ admits conjugate reciprocal zeros $z_o$:

$$\underbrace{p(z_o)}_{=0} + \underbrace{p^\dagger(z_o)}_{=0} + \sigma_v^{2\prime} = 0 \Rightarrow \sigma_v^{2\prime} = 0\ . \tag{4.56}$$

So there is no singular vector of the desired form in this case.

(2) $\mathrm{H}(z)$ has no conjugate reciprocal zeros:

$\mathrm{H}(z)$ is of the form $\mathrm{H}(z) = \prod_{i=0}^{N-2}(1 - z_i z^{-1})$. One can verify that

$$\mathrm{H}'(z) = \prod_{i=0}^{N-2}(1 + z_i z^{-1}) \tag{4.57}$$

is such that:

$$\sigma_a^2 \mathrm{H}(z)\mathrm{H}'(z) + \sigma_a^2 \mathrm{H}'(z)\mathrm{H}(z) = \sigma_a^2 2^{N-1} \prod_{i=0}^{N-2}(1 - \|z_i\|^2)\ . \tag{4.58}$$

And so $H'(z)$ and $\sigma_v^{2\prime} = -\sigma_a^2 2^{N-1} \prod_{i=0}^{N-2}(1 - \|z_i\|^2)$ verify (4.43); and it can also be proved that this is the only singular vector due to the unidentifiability of $\sigma_v^2$. It can also be verified that $\mathrm{H}'(z)$ is not a solution of (4.43) if $\mathrm{H}(z)$ has conjugate reciprocal zeros.

## A.2 Real Monochannel

**Solutions of the from** $[*\ \ *\ \ 0]^T$: $\sigma_v^{2\prime} = 0$.

Similar reasonings apply here.

(1) $\mathrm{H}(z)$ has no pair of conjugate reciprocal zeros:

$p(z) + p^\dagger(z) = 0$ can only be satisfied by $p(z) \equiv 0$. So the FIM is regular.

(2) $\mathrm{H}(z)$ has 1 pair conjugate reciprocal zeros: $(z_o, 1/z_o^*)$, $z_o \neq 1$, $z_o \neq -1$.

$$\mathrm{H}(z) = (1 - z_o z^{-1})(1 - z_o z) z^{-1} z_o^{-1} \underbrace{\prod_{i=1}^{N-3}(1 - z_i z^{-1})(1 - z_i z)}_{\mathrm{H}_1(z)}\ . \tag{4.59}$$

There is now only 1 $\mathrm{H}'(z)$ possible (the first solution in (4.54) is not valid here):

$$\mathrm{H}'(z) = (1 - z^{-1})(1 - z)z^{-1}z_o\mathrm{H}_1(z) \ . \tag{4.60}$$

The FIM has 1 singularity.

(3) $\mathrm{H}(z)$ has one zero equal to 1 or $-1$.

We assume that this zero is 1. $\mathrm{H}'(z)$ can be chosen as:

$$\mathrm{H}(z) = (1 - z^{-1}) \underbrace{\prod_{i=0}^{N-2}(1 - z_i z^{-1})}_{\mathrm{H}_1(z)} \ , \ \mathrm{H}'(z) = (1 + z^{-1})\mathrm{H}_1(z) \ . \tag{4.61}$$

(4) $\mathrm{H}(z)$ has several conjugate reciprocal zeros:

Then, to each pair of conjugate reciprocal zeros different from 1 and $-1$, and to each zero equal to 1 or $-1$ corresponds 1 singularity.

#### Solutions of the form $[* \ * \ \sigma_v^{2\,\prime}]^H$: $\sigma_v^{2\,\prime} \neq 0$

The same singularity as in the complex case, due to the inidentifiability of $\sigma_v^2$, appears (except again if the channel $\mathrm{H}(z)$ has conjugate reciprocal zeros).

### A.3 Multichannel

Assume now that $\mathbf{H}(z)$ is a true multichannel, possibly reducible:

$$\mathbf{H}(z) = \mathbf{H}_I(z)\mathrm{H}_c(z) \ . \tag{4.62}$$

As for the monochannel case, we search first the solutions of the form: $\left[\mathrm{Re}^T(h') \ \ \mathrm{Im}^T(h') \ \ 0\right]^T$. Then $h'$ verifies:

$$\mathcal{T}(h)\mathcal{T}^H(h') + \mathcal{T}(h')\mathcal{T}^H(h) = 0 \ . \tag{4.63}$$

The burst length is assumed to be $M \geq \underline{M} + 1$ which can be lower than $N - 1$ (so the transposition to the z-domain is not as convenient as in the monochannel case). The previous equation implies that $\mathcal{T}(h')$ should have for effect to reduce the previous quantity to at least the same rank as $\mathcal{T}(h)$. So:

$$\mathrm{range}\{\mathcal{T}(h')\} \subset \mathrm{range}\{\mathcal{T}(h)\} \tag{4.64}$$

which implies, using theorem 1:

$$\mathbf{H}'(z) = \mathbf{H}_I(z)\mathrm{H}'_c(z) \tag{4.65}$$

$$\left\{ \begin{array}{rcl} \mathbf{H}'(z) & = & \mathbf{H}_I(z)\mathrm{H}'_c(z) \\ \mathbf{H}(z) & = & \mathbf{H}_I(z)\mathrm{H}_c(z) \end{array} \right. \Rightarrow \mathcal{T}(h_I)\mathcal{T}(h_c)\mathcal{T}^H(h'_c)\mathcal{T}^H(h_I)+\mathcal{T}(h_I)\mathcal{T}(h'_c)\mathcal{T}^H(h_c)\mathcal{T}^H(h_I) = 0 \ . \tag{4.66}$$

As $\mathcal{T}(h_I)$ is full column-rank,

$$(4.66) \Leftrightarrow \mathcal{T}(h_c)\mathcal{T}^H(h'_c) + \mathcal{T}(h'_c)\mathcal{T}^H(h_c) = 0 \ . \tag{4.67}$$

As $\mathcal{T}(h_c)$ is of length at least $N_c-1$ according to the identifiability conditions, (4.67) implies:

$$\mathrm{H}_c(z)\mathrm{H}'^{\dagger}_c(z) + \mathrm{H}'^{\dagger}_c(z)\mathrm{H}_c(z) = 0 \tag{4.68}$$

which leads to the monochannel case treated previously.

As for the solutions of the form $\left[* \ \sigma_v^{2'}\right]^T$, $\sigma_v^{2'} \neq 0$, there are none in this case ($\sigma_v^2$ is identifiable in any case).

Chapter 5

# PERFORMANCE COMPARISON BETWEEN SEMI–BLIND, BLIND AND TS CHANNEL ESTIMATION

*We study the performance of semi–blind FIR multichannel estimation compared to blind and training sequence estimation through the analysis of the associated Cramér–Rao Bounds (CRBs). Deterministic and Gaussian models are considered, but some words will be said about finite alphabet methods. The superiority of semi–blind methods over blind and training sequence methods is demonstrated. Semi–blind estimation allows a significant gain of performance and for a given desired energy allows us to reduce the length of the training sequence. It is also more robust, making possible the estimation of channels that cannot be estimated by training sequence techniques, because the training sequence is too short, or by blind techniques because the channel is ill-conditioned. Furthermore, we show the influence of the number of known symbols on semi–blind performance, and mention some optimization results on the characteristics of the training sequence. Numerical evaluations illustrate all these aspects.*

## 5.1 Introduction

Apart from a performance study, an important issue indirectly treated in this chapter are deterministic semi–blind identifiability conditions for the non-trivial case of arbitrarily distributed known symbols. This case can be solved by examining the Fisher Information Matrix (FIM) regularity as FIM regularity implies local identifiability. If we call $N_c - 1$ the number of zeros of the multichannel, then under certain conditions, $2N_c - 1$ known symbols, with arbitrary positions, are necessary and sufficient to allow identifiability. The case of known symbols all equal to 0 is also treated.

In the performance study of semi–blind estimation, the following points are treated:

- The influence of the number of known symbols on the semi–blind performance is studied. Specifically, we see how the knowledge of only a few symbols allows one to improve the estimation performance significantly.

- The case of monochannels and reducible channels (multichannels with zeros) is studied. We underscore the ability of Gaussian blind methods to estimate (locally) the zeros of a channel in general. In the deterministic case, monochannels can only be estimated by the training symbols, while the blind part brings no information. For reducible channels, blind information for the estimation of the zeros is asymptotically negligible. For the Gaussian methods, blind information is useful in the estimation of monochannels and of the zeros of reducible channels.

- We compare the CRBs for pure training sequence and semi–blind modes and illustrate some of the most interesting properties of semi–blind estimation. The addition of the blind information to the training sequence information results in a significant gain w.r.t. the training sequence mode; also, for a desired performance level, semi–blind allows one to reduce the training sequence length. It is also more robust, allowing the estimation of channels that cannot be estimated by training sequence techniques, if the training sequence is too short.

- We compare the blind and semi–blind modes. Blind methods have to be applied certain constraints: we use the results on CRBs under contraints to compare blind and semi–blind estimation modes under the same constraints. The addition of the training sequence information allows a significant gain of performance w.r.t. blind estimation. This is particularly true for ill-conditioned channels.

- We furthermore compare the deterministic and Gaussian CRBs.

- Some optimization issues are also mentioned: the value of the known symbols, their distribution in the burst and the position of a training sequence in the burst.

## 5.2 Deterministic Model

In the semi–blind case, the estimation parameter is $\theta = [h^T \;\; A_U^T]^T$, where $A_U$ designate the unknown symbols in the burst. As in the blind case, we work with the complex FIM $J_{\theta\theta}$ for circular complex input constellation and treat the real and complex input symbols together.

### 5.2.1 Semi–Blind FIMs

$$J_{\theta\theta} = \frac{1}{\sigma_v^2}\left(\frac{\partial \boldsymbol{X}^H}{\partial \theta^*}\right)\left(\frac{\partial \boldsymbol{X}^H}{\partial \theta^*}\right)^H = \frac{1}{\sigma_v^2}\left[\begin{array}{c} \mathcal{T}_U^H(h) \\ \mathcal{A}^H \end{array}\right]\left[\begin{array}{cc} \mathcal{T}_U(h) & \mathcal{A} \end{array}\right] \tag{5.1}$$

since $\dfrac{\partial \boldsymbol{X}^H}{\partial A_U^*} = \dfrac{\partial \left(\mathcal{T}_K(h)A_K + \mathcal{T}_U(h)A_U\right)^H}{\partial A_U^*} = \mathcal{T}_U^H(h)$ and $\dfrac{\partial \boldsymbol{X}^H}{\partial h^*} = \mathcal{A}^H$.

When the FIM is regular, the semi–blind CRB for complex and real symbols is:

$$CRB_{SB} = \sigma_v^2\left[\mathcal{A}^H P^{\perp}_{\mathcal{T}_U(h)}\mathcal{A}\right]^{-1}. \tag{5.2}$$

### 5.2.2 FIM Regularity

We treat here the general case of a reducible channel $\mathbf{H}(z) = \mathbf{H}_I(z)\mathrm{H}_c(z)$. Sufficient conditions for the semi–blind FIM to be regular are given. The conditions hold for grouped known symbols, as well as arbitrarily dispersed known symbols. They will be very useful as FIM regularity implies local identifiability: for arbitrarily dispersed known symbols, identifiability appears indeed difficult to show directly.

In fact the following general theorem holds for the semi–blind deterministic model.

**Theorem 18** *The channel $h$ is locally identifiable if and only if the semi–blind FIM is regular.*

*Proof:* Assume that the FIM has 1 singularity $\theta' = [A_U'^T \;\; h'^T]^T$:

$$\mathcal{T}_U(h)A_U' + \mathcal{A}h' = 0 \;\;\Leftrightarrow\;\; \mathcal{T}_U(h)A_U' + \mathcal{T}(h')A = 0\,. \tag{5.3}$$

For $\epsilon > 0$ arbitrarily small:

$$\begin{array}{rcl} m_Y(\theta + \epsilon\theta') - m_Y(\theta) & = & \mathcal{T}(h + \epsilon h')\left[A + \epsilon A_U''\right] - \mathcal{T}(h)A \\ & = & \epsilon\left[\mathcal{T}_U(h)A_U' + \mathcal{T}(h')A\right] + O(\epsilon^2) \quad = \quad O(\epsilon^2) \end{array} \tag{5.4}$$

where $A_U''$ has the same length as $A$, is equal to $A_U'$ at the position of the unknown symbols, and has zero entries at the positions of the known symbols: $\mathcal{T}(h)A_U'' = \mathcal{T}_U(h)A_U'$. This implies that $(A_U, h)$ is not locally blindly identifiable.

Now assume that $(A_U, h)$ is not locally blindly identifiable, then one can find $\Delta h$ and $\Delta A''_U$, where $\Delta A''_U$ has zero entries at the positions of the known symbols and $\|\Delta h\|$ and $\|\Delta A''_U\|$ are arbitrarily small verifying $\mathcal{T}(h)A = \mathcal{T}(h+\Delta h)(A+\Delta A_U)$, then:

$$\begin{aligned} \mathcal{T}(h+\Delta h)(A+\Delta A_U) - \mathcal{T}(h)A &= \mathcal{T}_U(h)\Delta A_U + \mathcal{A}\Delta h + O(\Delta h \Delta A_U) \\ &= 0\,. \end{aligned} \tag{5.5}$$

This implies that $\left[\Delta h^T \;\; \Delta {A_U}^T\right]^T$ is a null vector of the FIM.

□

Furthermore, we make the following conjecture:

**Conjecture 1** *In the deterministic model, FIM regularity implies global identifiability.*

In the case of grouped known symbols it can be proved that local identifiability implies global identifiability, so the conjecture can be proved in this case[1].

In appendix A, we examine the singularities of the FIM: the conditions for the FIM to be regular and then for the channel to be locally identifiable are studied. The results can be summarized as follows.

### Non-Zero Known Symbols

**Theorem 19** *The FIM is regular and the channel* $\mathbf{H}(z)$ *is identifiable with probability 1 if*

*(i)* *Burst length* $M \geq \max(N_I + 2\underline{M}_I, N_c - N_I + 1)$.

*(ii)* *Number of excitation modes* $\geq N + \underline{M}_I$.

*(iii)* *Number of known symbols* $\geq 2N_c - 1$, *which is also a* **necessary** *condition.*

*As far as the position of the known symbols in the burst is concerned:*

- *If the known symbols are grouped in a single sequence, with a number of independent input symbol modes* $\geq N_c$, *the channel is identifiable.*
- *If the known symbols are arbitrarily distributed, the channel is identifiable with probability 1.*

The notion of probability 1 here assumes a probability distribution for $h$ with a support that has positive measure (no deterministic relations between the coefficients of $h$ exist). For grouped known symbols, the conditions are the same as the global identifiability conditions [DetSB].

[1] Finding $h$ and $A_U$ from $\mathcal{T}(h)A$ corresponds to solving a set of polynomial equations: there may exist some theoretical results from polynomial algebra which would allow to prove the conjecture for arbitrarily dispersed symbols. We did not pursue this however.

### Zero Known Symbols

When the known symbols are all equal to 0, the channel can at best be identified up to a scale factor. Indeed, $\mathcal{T}(h)A = \mathcal{T}(h')A'$, with $h' = \alpha h$, $A' = A/\alpha$ and $A_K = A'_K = 0$. We can prove the following:

**Theorem 20** *When the known symbols are all equal to 0, the channel is semi–blindly locally identifiable up to a scale factor if and only if the FIM is 1-singular.*

Strictly speaking we can prove only local identifiability, but then apply Conjecture 1.

The position of the known symbols cannot be totally arbitrary. If the known symbols are grouped for example, it can be shown that the FIM has $N_c$ singularities. In Appendix B, we prove the following theorem:

**Theorem 21** *The FIM is 1-singular and the channel* $\mathbf{H}(z)$ *is identifiable with probability 1 up to a scale factor if*

*(i) Burst length* $M \geq N_I + 2\underline{M}_I$.

*(ii) Number of excitation modes* $\geq N + \underline{M}_I$.

*(iii) Number of known symbols* $\geq 2N_c - 2$, *which is also a* **necessary** *condition.*

*(iv) The known symbols are "sufficiently" dispersed: there are at least* $N_c - 1$ *symbols that do not belong to a group of* $N_c$ *or more known symbols.*

If the known symbols are only partially equal to 0, Theorem 19 can be applied: if there are more than $N_c - 1$ known symbols that are zero, then there should have at least $N_c - 1$ zero known symbols that are not in a group of $N_c$ or more zero known symbols.

### 5.2.3 CRB for Training Sequence Based Channel Estimation

To compute the CRB for the TS case, we can use the semi–blind deterministic CRB for the case in which all the input symbols are known:

$$CRB_{TS} = \sigma_v^2 \left[\mathcal{A}^H \mathcal{A}\right]^{-1} . \tag{5.6}$$

The CRB depends on the value of the symbols present in the training sequence. It is minimized for a given training sequence energy when $\mathcal{A}^H\mathcal{A}$ is a multiple of identity [66]. The CRB is then equal to $\frac{\sigma_v^2}{M\sigma_a^2} I$. This indicates a condition on the choice of a deterministic training sequence for good channel estimation. When on the other hand the training sequence is drawn from a sequence of uncorrelated symbols and its length increases, it is interesting to note that $\frac{1}{M}\mathcal{A}^H\mathcal{A}$ tends to $\sigma_a^2 I$ (law of large numbers), which again leads to the minimal value of the CRB.

### 5.2.4 Semi–Blind CRB: Monochannel and Reducible Channel Cases

#### Monochannel

Assume the $M_K$ known symbols are grouped in a single sequence, and, for simplicity reasons, are located at the beginning of the burst. It can be shown that:

$$P^{\perp}_{\mathcal{T}_U(h)} = \begin{bmatrix} 0 & 0 \\ 0 & I_{M_K-N+1} \end{bmatrix} \tag{5.7}$$

so that the FIM is:

$$FIM_{SB} = \frac{1}{\sigma_v^2}\mathcal{A}_{TS}^H\mathcal{A}_{TS} = FIM_{TS}\,. \tag{5.8}$$

$\mathcal{A}_{TS}$ is such that $\boldsymbol{Y}_{TS} = \mathcal{T}_{TS}(h)A_K = \mathcal{T}_{TS}(h)A_{TS} = \mathcal{A}_{TS}h$ where $\boldsymbol{Y}_{TS}$ includes the first $M_K-N+1$ first observations containing only known symbols.

**Result 3** *In the case of grouped known symbols, the blind information is useless in the estimation of a monochannel, which is done by the training sequence only. The semi-blind CRB is:*

$$CRB_{SB} = \sigma_v^2\left(\mathcal{A}_{TS}^H\mathcal{A}_{TS}\right)^{-1} = CRB_{TS}\,. \tag{5.9}$$

When the known symbols are grouped in several training sequences, we observed that blind information plays some transient effect, but the CRB tends to a constant when the number of unknown symbols increases.

#### Reducible Channel

The CRB (5.2) does not exploit the structure of the channel, *i.e.* the fact that the true channel has zeros or not. Here we assume that the channel is reducible and that we have detected the number of zeros. We compute the CRB for the irreducible and the reducible part of the channel. The estimation parameter is: $\theta = \left[A_U^T \;\; h_I^T \;\; \bar{h}_c^T\right]^T$, $\bar{h}_c$ is deduced from $h_c$ by removing the $1^{st}$ (known) coefficient. We have:

$$\frac{\partial \boldsymbol{X}^H}{\partial A_U^*} = \mathcal{T}_U^H(h) \qquad \frac{\partial \boldsymbol{X}^H}{\partial h_I^*} = \mathcal{A}_I^H \qquad \frac{\partial \boldsymbol{X}^H}{\partial h_c^*} = \mathcal{A}_c'^H \tag{5.10}$$

where $\mathcal{A}_I$ is such that $\mathcal{T}(h_I)\left[\mathcal{T}(h_c)A\right] = \mathcal{A}_I h_I$; $\mathcal{T}(h_c)A = \mathcal{A}_c h_c$ and $\mathcal{A}_c' = \mathcal{T}(h_I)\bar{\mathcal{A}}_c$, $\bar{\mathcal{A}}_c$ is deduced from $\mathcal{A}_c$ by removing its $1^{st}$ column. After some calculations, the FIMs for $h_I$ and $\bar{h}_c$ are:

$$(CRB)^{-1}_{h_I h_I} = \sigma_v^2\left(\mathcal{A}_I^H P^{\perp}_{\mathcal{T}_U(h)}\mathcal{A}_I\right) - \sigma_v^2\left(\mathcal{A}_I^H P^{\perp}_{\mathcal{T}_U(h)}\mathcal{A}_c'\right)\left(\mathcal{A}_c'^H P^{\perp}_{\mathcal{T}_U(h)}\mathcal{A}_c'\right)^{-1}\left(\mathcal{A}_c'^H P^{\perp}_{\mathcal{T}_U(h)}\mathcal{A}_I\right) \tag{5.11}$$

$$(CRB)^{-1}_{\bar{h}_c\bar{h}_c} = \sigma_v^2 \left( \mathcal{A}_c'^H P^{\perp}_{\mathcal{T}_U(h)} \mathcal{A}_c' \right) - \sigma_v^2 \left( \mathcal{A}_c'^H P^{\perp}_{\mathcal{T}_U(h)} \mathcal{A}_I \right) \left( \mathcal{A}_I^H P^{\perp}_{\mathcal{T}_U(h)} \mathcal{A}_I \right)^{-1} \left( \mathcal{A}_I^H P^{\perp}_{\mathcal{T}_U(h)} \mathcal{A}_c' \right). \tag{5.12}$$

It can be explicitly shown that for grouped known symbols, the CRB for $h_I$ decreases as in $\frac{1}{M_U}$ and $\frac{1}{M_K}$, as $M_U$ and $M_K$ increase: blind information is useful in the estimation of $h_I$. In general, $M_U \gg M_K$, so that the blind information dominates in the estimation of $h_I$. As the number of unknown symbols grows to infinity, the CRB for $h_c$ becomes constant, so we have the following result.

**Result 4** *In the case of grouped known symbols, the blind information is asymptotically (in the number of unknown symbols) useless in the estimation of the zeros of the channel, which is asymptotically done by the training sequence only.*

For dispersed known symbols, the same behavior was observed. A similar study (as for the estimation of the zeros) can be done for the estimation of the ambiguous scale factor: the blind information plays asymptotically no role either in the estimation of this factor.

**Remark** For a number of grouped known symbols at the beginning of the burst of at least $2N_c - 1$, it can be verified that $P_{\mathcal{T}_U(h)} = P_{\mathcal{T}'_U(h_I)}$, where $\mathcal{T}'_U(h_I)$ is $\mathcal{T}(h_I)$ truncated of the first $M'_K = M_K - N_c + 1$ columns. The semi–blind CRB for $h$ is then:

$$CRB_{SB} = \sigma_v^2 \left[ \mathcal{A}^H P^{\perp}_{\mathcal{T}'_U(h_I)} \mathcal{A} \right]^{-1}. \tag{5.13}$$

The CRB does not depend on the values of the zeros (if any) but only on their number.

### 5.2.5 Semi–blind CRB with constraints

The regularized blind CRB in (4.8) corresponds to the estimation of the channel but with constraints (4.6), (4.7). The regular semi–blind CRB (5.2) does not use these constraints. This is why a direct comparison between blind and semi–blind modes through these CRBs is not possible. To allow a comparison, we use a semi–blind CRB computed under the blind constraints:

$$CRB_{SB,C} = \sigma_v^2 \left( P_h^{\perp} \mathcal{A}^H P^{\perp}_{\mathcal{T}(h)} \mathcal{A} P_h^{\perp} \right)^{+}. \tag{5.14}$$

### 5.2.6 Comparisons and Numerical Evaluations

We compare here the different estimation modes through their CRBs. These comparisons are illustrated by curves showing the trace of the CRBs w.r.t. the number of known (or unknown) symbols in the input burst for a complex input constellation, QPSK, and a real one, BPSK. In the case of a BPSK, the number of channels gets doubled as in (1.2). The known input symbols are randomly chosen and grouped at the beginning of the burst. The SNR, defined

as $\frac{\sigma_a^2 \|h\|^2}{m\sigma_v^2}$ (average SNR per subchannel), is 10dB; $M = 100$. The different channels tested were chosen randomly, our purpose being not to study specific channel cases but rather to see the general mechanism of semi–blind estimation. Four different types of channels were tested: an irreducible channel $\boldsymbol{H}_{well}$, an ill–conditioned channel with a nearly common zero $\boldsymbol{H}_{ill}$, a monochannel $\boldsymbol{H}_{mono}$ and a reducible channel with irreducible part $\boldsymbol{H}_I$ and reducible part $H_c$. The different channels are given in Appendix D.

### Semi–Blind

Figure 5.2 shows the semi–blind CRB w.r.t. the number of known symbols using a BPSK, for the well–conditioned channel $H_{well}$ (left) and the ill–conditioned channel $H_{ill}$ (right). When very few symbols are known, performance is bad: this is due to the difficulty of estimating the scale factor of the channel or the nearly common zero with few known symbols. However, we observe that after the introduction of (very) few more known symbols, performance increases dramatically, especially in the case of the ill–conditioned channel. After this threshold of improvement, the estimation of the channel being already sufficiently good, it is necessary to introduce a large number of known symbols to get a significant further improvement. These numerical evaluations indicate that semi–blind techniques could improve performance drastically w.r.t. blind techniques with only few known symbols.

Figure 5.11 (left) shows the CRBs using a QPSK for the monochannel $\boldsymbol{H}_{mono}$. We plot the CRB for a fixed number of 10 known symbols grouped at the beginning of the burst w.r.t. the number of unknown symbols in the burst: it can be seen that the blind part brings strictly no information to the estimation of the monochannel. We present also the case of a reducible channel $\mathbf{H}(z) = \mathbf{H}_I(z)\mathrm{H}_c(z)$ in Figure 5.11 (right). The difference in the slope between the CRBs for $h_I$ and $h_c$ is visible as $\frac{M_K}{M_U} \to 0$ and it can be seen that the CRB for $h_c$ becomes constant.

### Semi–Blind vs Training Sequence (TS)

The known symbols used in the TS mode are the same (same symbols and same number) as the symbols known in the semi–blind mode, as indicated in Figure 5.4. A comparison between the FIMs gives:

$$\mathcal{A}_{TS}^H \mathcal{A}_{TS} \leq \mathcal{A}^H P_{\mathcal{T}_U(h)}^{\perp} \mathcal{A}\,. \tag{5.15}$$

Hence, the CRB for the TS mode is greater than that for the semi–blind mode.

Figure 5.5 is one of the most important figures of this study. We can see that semi–blind estimation represents an important gain w.r.t. the TS mode, especially when few symbols are known. In Figure 5.5 (left), we have a gain of factor 20 for 10 known symbols and of 3 for 25 symbols. Besides, for the same performance, fewer known symbols for semi–blind

estimation are needed compared to TS based estimation. To get the performance of semi–blind estimation with 10 or 25 known symbols, one needs respectively 50 and 70 TS symbols. The CRB when all the $M+N-1$ input symbols are known is given as a reference.

#### Semi–Blind vs Blind

In this comparison, the input bursts are the same (same symbols and same length) but part of the symbols is known in the semi–blind mode, as indicated in Figure 5.7. We use the constrained CRBs (4.8) for blind estimation and (5.14) for semi–blind estimation. Figure 5.8 shows the CRBs ($M_K = 0$ corresponds to the blind case) for $\boldsymbol{H}_{well}$ and $\boldsymbol{H}_{ill}$. The introduction of very few known symbols is sufficient to improve performance significantly w.r.t. blind estimation, again especially in the case of the ill-conditioned channel.

## 5.3 Gaussian Input Model

### 5.3.1 Semi–Blind FIMs

**Circular Complex Symbol Constellation** The FIM computation is based on the complex density probability function of $\boldsymbol{Y}$:

$$\boldsymbol{Y} \sim \mathcal{N}(m_Y, C_{YY}) \quad \text{with} \quad C_{YY} = \sigma_a^2 \mathcal{T}_U(h)\mathcal{T}_U^H(h) + \sigma_v^2 I, \quad m_Y = \mathcal{T}_K(h)A_K \,. \tag{5.16}$$

Let $h_R = [\mathrm{Re}^T(h)\ \mathrm{Im}^T(h)]^T$ and $\overline{\theta}_R = [h_R^T\ \sigma_v^2]^T$, the real parameter vector. $J_{\theta\theta^*}$ being non zero, we cannot consider the complex CRB anymore: the real CRB $\mathcal{J}_{\overline{\theta}_R\overline{\theta}_R}$ is determined via (3.7) thanks to the quantities:

$$J_{\theta\theta}(i,j) = \left(\mathcal{A}_K^H C_{YY}^{-1}\mathcal{A}_K\right)(i,j) + \mathrm{tr}\left\{C_{YY}^{-1}\left(\frac{\partial C_{YY}}{\partial \theta_i^*}\right)C_{YY}^{-1}\left(\frac{\partial C_{YY}}{\partial \theta_j^*}\right)^H\right\} \tag{5.17}$$

$$J_{\theta\theta^*}(i,j) = \mathrm{tr}\left\{C_{YY}^{-1}\left(\frac{\partial C_{YY}}{\partial \theta_i^*}\right)C_{YY}^{-1}\left(\frac{\partial C_{YY}}{\partial \theta_j^*}\right)\right\} \tag{5.18}$$

$$\text{where:} \quad \begin{cases} \dfrac{\partial C_{YY}}{\partial h_i^*} = \sigma_a^2 \mathcal{T}_U(h)\mathcal{T}_U^H\left(\dfrac{\partial h}{\partial h_i^*}\right) \\ \dfrac{\partial C_{YY}}{\partial \sigma_v^2} = \dfrac{1}{2}I \,. \end{cases} \tag{5.19}$$

We have introduced $\mathcal{A}_K$ from $\mathcal{T}_K(h)A_K = \mathcal{A}_K h$.

**Real Symbol Constellation** When the input constellation is real, the FIM is:

$$\mathcal{J}_{\theta\theta}(i,j) = \left(\mathcal{A}_K^T C_{YY}^{-1} \mathcal{A}_K\right)(i,j) + \frac{1}{2}\mathrm{tr}\left\{C_{YY}^{-1}\left(\frac{\partial C_{YY}}{\partial \theta_i}\right) C_{YY}^{-1}\left(\frac{\partial C_{YY}}{\partial \theta_j}\right)^T\right\} \tag{5.20}$$

$$\begin{cases} \dfrac{\partial C_{YY}}{\partial h_i} = \sigma_a^2 \mathcal{T}_U(h)\mathcal{T}_U^T\left(\dfrac{\partial h}{\partial h_i}\right) + \sigma_a^2 \mathcal{T}_U\left(\dfrac{\partial h}{\partial h_i}\right)\mathcal{T}_U^T(h) \\ \dfrac{\partial C_{YY}}{\partial \sigma_v^2} = I\,. \end{cases} \tag{5.21}$$

In (5.17) and (5.20), $\left(\mathcal{A}_K^T C_{YY}^{-1}\mathcal{A}_K\right)(i,j) = 0$ when $i$ and/or $j$ equal $mN{+}1$ (corresponding to $\sigma_v^2$).

Here also, we have the equivalence between local semi–blind identifiability and FIM regularity:

**Theorem 22** *The channel $h$ is locally identifiable if and only if the semi–blind FIM is regular.*

*Proof:* It can be shown that the semi–blind FIM is singular if and only if there exits a vector $\left[h'^T \;\; \sigma_v^{2'}\right]^T$ such that:

$$\mathcal{A}_K h' = 0 \text{ and } \sigma_a^2 \mathcal{T}_U(h)\mathcal{T}_U^H(h') + \mathcal{T}_U(h')\mathcal{T}_U^H(h) + \sigma_v^{2'} I = 0\,. \tag{5.22}$$

The equivalence now follows using the same method as for theorem 18.

□

As derived in [67], for 1 non–zero known symbol not located at the edges of the burst, we have (global) identifiability and then also FIM regularity:

### 5.3.2 Semi–Blind CRBs: Reducible Channel

We now treat the case of a reducible channel: $\theta = \left[h_I^T \;\; \bar{h}_c^T \;\; \sigma_v^2\right]^T$.

**Circular Complex Symbol Constellation** Let $h_R = [\mathrm{Re}(h_I^T)\ \mathrm{Im}(h_I^T)\ \mathrm{Re}(\bar{h}_c^T)\ \mathrm{Im}(\bar{h}_c^T)]^T$ and $\overline{\theta}_R = [h_R^T\, \sigma_v^2]^T$ be the parameter vector. The quantities of interest are:

$$J_{\theta\theta}(i,j) = \left(\left[\mathcal{A}_{I_K}\ \bar{\mathcal{A}}_{c_K}\right]^H C_{YY}^{-1}\left[\mathcal{A}_{I_K}\ \bar{\mathcal{A}}_{c_K}\right]\right)(i,j) + \mathrm{tr}\left\{C_{YY}^{-1}\left(\frac{\partial C_{YY}}{\partial \theta_i^*}\right) C_{YY}^{-1}\left(\frac{\partial C_{YY}}{\partial \theta_j^*}\right)^H\right\} \tag{5.23}$$

$$J_{\theta\theta^*}(i,j) = \mathrm{tr}\left\{C_{YY}^{-1}\left(\frac{\partial C_{YY}}{\partial \theta_i^*}\right) C_{YY}^{-1}\left(\frac{\partial C_{YY}}{\partial \theta_j^*}\right)\right\} \tag{5.24}$$

$$\text{with:} \begin{cases} \dfrac{\partial C_{YY}}{\partial h^*_{I_i}} = \sigma_a^2 \mathcal{T}(h)\mathcal{T}^H(h_c)\mathcal{T}^H\left(\dfrac{\partial h_I}{\partial h_{I_i}}\right) \\ \dfrac{\partial C_{YY}}{\partial \bar{h}^*_{c_i}} = \sigma_a^2 \mathcal{T}(h)\mathcal{T}^H\left(\dfrac{\partial h_c}{\partial \bar{h}_{c_i}}\right)\mathcal{T}^H(h_I) \\ \dfrac{\partial C_{YY}}{\partial \sigma_v^2} = \dfrac{1}{2} I \ . \end{cases} \tag{5.25}$$

$\mathcal{A}_{I_K}$ and $\mathcal{A}_{c_K}$ are such that:

$$\begin{cases} \mathcal{T}(h_I)\left[\mathcal{T}_K(h_c)A_K\right] = \mathcal{A}_{I_K} h_I \ , \\ \mathcal{T}_K(h_c)A_K = \mathcal{A}_{c_K} h_c \ . \end{cases} \tag{5.26}$$

$\bar{\mathcal{A}}_{c_K}$ is $\mathcal{A}_{c_K}$ truncated of its first column.

**Real Symbol Constellation** When the input constellation is real, the FIM is:

$$\mathcal{J}_{\theta\theta}(i,j) = \left( \left[\mathcal{A}_{I_K} \ \bar{\mathcal{A}}_{c_K}\right]^T C_{YY}^{-1} \left[\mathcal{A}_{I_K} \ \bar{\mathcal{A}}_{c_K}\right] \right)(i,j) + \frac{1}{2}\mathrm{tr}\left\{ C_{YY}^{-1}\left(\frac{\partial C_{YY}}{\partial \theta_i}\right) C_{YY}^{-1}\left(\frac{\partial C_{YY}}{\partial \theta_j}\right)^T \right\} \tag{5.27}$$

$$\begin{cases} \dfrac{\partial C_{YY}}{\partial h^*_{I_i}} = \sigma_a^2 \mathcal{T}(h)\mathcal{T}^T(h_c)\mathcal{T}^T\left(\dfrac{\partial h_I}{\partial h_{I_i}}\right) + \sigma_a^2 \mathcal{T}\left(\dfrac{\partial h_I}{\partial h_{I_i}}\right)\mathcal{T}(h_c)\mathcal{T}^T(h) \\ \dfrac{\partial C_{YY}}{\partial \bar{h}_{c_i}} = \sigma_a^2 \mathcal{T}(h)\mathcal{T}^T\left(\dfrac{\partial h_c}{\partial \bar{h}_{c_i}}\right)\mathcal{T}^T(h_I) + \sigma_a^2 \mathcal{T}(h_I)\mathcal{T}\left(\dfrac{\partial h_c}{\partial \bar{h}_{c_i}}\right)\mathcal{T}^T(h) \\ \dfrac{\partial C_{YY}}{\partial \sigma_v^2} = I \ . \end{cases} \tag{5.28}$$

The blind and training sequence information are both useful to the estimation of both $h_I$ and $\bar{h}_c$. The CRBs for $h_I$ and $h_c$ evolve as $\dfrac{1}{M_U}$ when $M_U \to \infty$ and as $\dfrac{1}{M_K}$ when $M_K \to \infty$. Some numerical evaluations of the corresponding CRBs will be given in the next section.

**Result 5** *Unlike in the deterministic model, in the Gaussian model, blind information is useful in the estimation of the zeros of the channel.*

Semi–Blind CRBs for a monochannel can also be derived in the same way.

## 5.4 Comparison between Deterministic and Gaussian Models

### 5.4.1 Comparison between CRBs

This comparison is not obvious in general. If we consider the simple case of an instantaneous channel, $N = 1$: for the real and complex constellation cases, it can be verified that the difference between the Gaussian and the deterministic FIM is indefinite and that the sign of the difference between the trace of the FIMs depends on the value of the parameters. The results here are then different from those obtained in DOA [42] for which blind deterministic CRBs are below blind Gaussian CRBs.

### 5.4.2 High SNR

We assume that $\frac{\sigma_v^2}{\sigma_a^2} \to 0$. For simplicity reasons, we will consider the irreducible channel case only. For the Gaussian model, at high SNR and large $M_U$, the terms $J_{\sigma_v^2\sigma_v^2}$, $J_{h\sigma_v^2}$ and $J_{hh^*}$ are of order $\frac{M_U}{\sigma_v^4}$; the term $J_{hh}$ is of order $\frac{M_U}{\sigma_v^2}$. So, at high SNR, the influence of the estimation $\sigma_v^2$ on the estimation of $h$ becomes negligible, and the term $J_{\theta\theta^*}$ can be neglected. The FIM for $h$ is then the complex *FIM* $J_{hh}(\theta)$ (as in the deterministic case). In Appendix C, we prove the following result:

**Result 6** *Asymptotically in the number of unknown symbols and in SNR, the deterministic and Gaussian semi–blind FIMs are equal.*

### 5.4.3 Comparisons and Numerical Evaluations

The parameter values are the same as in the deterministic case. In our numerical evaluations, the estimation of $\sigma_v^2$ had nearly no influence on the estimation of $h$. The semi–blind curve (Figure 5.3) shows again a significant improvement when more and more symbols in the burst are known, especially for few known symbols. The CRB for the monochannel in (5.45) for 10 known symbols with a variable number of unknown symbols is plotted in Figure 5.11: it can be seen that the blind part brings information to the estimation of the monochannel. The case of a reducible channel (see (5.46)) is shown in Figure 5.12 for QPSK: we remark the same asymptotic behavior of the CRBs for $h_I$ and $\bar{h}_c$ w.r.t. the number of unknown symbols.

In Figure 5.6 semi–blind appears again better than the training sequence mode. Direct comparison between blind and semi–blind estimation is possible when the input constellation is real because the FIM is invertible: see Figure 5.6. We do not show here the complex case.

In Figure 5.3, both deterministic and Gaussian semi–blind curves can be compared. In the various examples we evaluated we observed that the Gaussian model allows better performance than the deterministic model.

## 5.5 Methods Exploiting the Finite Alphabet of the Input Symbols

In [67], a classification of blind methods in terms of the a priori knowledge of the input symbols exploited was proposed. The methods exploiting the Finite Alphabet (FA) nature of the input symbols perform joint symbol sequence detection and channel estimation and seem particularly interesting from a performance point of view. What is the CRB for the channel in this case? Assume the symbol detections to have a low probability of error. Then all symbols present in the problem act as training sequence for the channel estimation (with possibly some erroneous symbols). Hence the CRB for (error-free) training-sequence based channel estimation (with the training sequence being all input symbols) constitutes a lower bound

for the channel estimation error covariance matrix. The tightness of this bound depends on the probability of error. Nevertheless the error covariance matrix can be expected to be very small.

The disadvantage of the FA algorithms is that their cost function is highly multimodal: they require a very good initialization which could be provided by semi–blind deterministic or Gaussian algorithms [67].

## 5.6 Optimization Issues

**Values of the Known Symbols** (Deterministically) white input symbol sequences, in the sense that $\mathcal{A}^H\mathcal{A} = M\sigma_a^2 I$, optimizes the performance of training sequence based estimation. Optimization of the semi–blind CRB w.r.t. the known symbols depends on the channel; we expect however that such white sequences, even if they do not strictly optimize the semi–blind performance, would be among the best choices.

**Distribution of the Known Symbols over the Burst** Should the known symbols be grouped or separated? The answer seems again to depend on the channel. In this section we will call "minimum–phase" multichannel, a multichannel for which all the subchannels are minimum–phase, the energy is then concentrated in the first coefficients of the multichannel; a "maximum" phase multichannel will have maximum–phase subchannels.

We did some test to compare the deterministic CRBs for a minimum and maximum phase channel $\boldsymbol{H}_{min}$ and $\boldsymbol{H}_{max}$ and a randomly chosen channel $\boldsymbol{H}_{rand}$. In the tables below, we show the trace of the CRBs for the three channels for a fixed sequence of 10 known symbols, randomly chosen from a QPSK constellation (top table) or equal to 0 (buttom table), grouped in the middle of the burst or uniformly dispersed all over the burst. The burst length is $M = 100$. The CRBs are averaged over 1000 realizations of the unknown symbols in the case of QPSK.

| Known Symbols | $\boldsymbol{H}_{min}$ | $\boldsymbol{H}_{max}$ | $\boldsymbol{H}_{rand}$ |
|---|---|---|---|
| grouped | 0.36 | 0.79 | 0.22 |
| separated | 1.33 | 2.38 | 0.24 |

| Known Symbols | $\boldsymbol{H}_{min}$ | $\boldsymbol{H}_{max}$ | $\boldsymbol{H}_{rand}$ |
|---|---|---|---|
| grouped | 3.23 | 9.99 | 0.78 |
| separated | 0.53 | 1.36 | 0.16 |

When the known symbols are chosen randomly, for the minimum and maximum–phase channels, performances are better when the known symbols are grouped than uniformly separated in the burst. For the random channel, both choices seem equivalent. When the

known symbols are all equal to 0, it is however better to have them dispersed all over the burst in all cases.

**Position of the Training Sequence in the Burst** Again, the answer depends on the characteristics of the channel. What could be done is study the CRBs w.r.t. the position of the training sequence for a stochastic channel model.

Besides performance, other considerations such as algorithm complexity have to be taken into account. When the known symbols are grouped, a semi–blind criterion can be formed as the linear combination of a training sequence criterion and a blind criterion. This combination looses some information, but offers the advantage to keep the structural properties of the blind estimation problem which is the most costly part and allows to build fast semi–blind algorithms.

## 5.7 Conclusions

We have proposed a study on Cramér–Rao Bounds for blind and semi–blind FIR multichannel estimation. Semi–blind methods appear more robust and powerful than blind and training sequence methods, especially for a small number of known symbols. In the following part of the thesis, we concentrate on semi–blind methods based on DML and GML.

# A Semi–Blind Identifiability Conditions for Non–Zero Arbitrarily Dispersed Known Symbols

We examine the conditions for the semi–blind FIM to be regular in the case where the known symbols are all non–zero. The channel is assumed reducible $\mathbf{H}(z) = \mathbf{H}_I(z)\mathrm{H}_c(z)$.

The semi–blind global FIM is $[\mathcal{T}_U(h)\ \mathcal{A}]^H [\mathcal{T}_U(h)\ \mathcal{A}]$. $[\mathcal{T}_U(h)\ \mathcal{A}]$ is tall under condition $(i)$ of theorem 19, so the FIM is singular if and only if one can find $h'$ and $A'_U$ verifying:

$$\left[\ \mathcal{T}_U(h) \quad \mathcal{A}\ \right] \left[{A'_U}^T \quad -{h'}^T\right]^T = 0 \qquad \text{or} \qquad \mathcal{T}_U(h)A'_U = \mathcal{A}h' = \mathcal{T}(h')A\ . \tag{5.29}$$

This is also equivalent to:

$$\mathcal{T}(h)A' = \mathcal{T}(h')A, \text{ with } \mathcal{S}_K A' = A'_K = 0 \tag{5.30}$$

which is the blind problem except that the constraint $A'_K = 0$ is imposed:

$$(5.30) \Leftrightarrow \mathcal{T}(h_I)\mathcal{T}(h_c)A' = \mathcal{T}(h')A, \text{ with } A'_K = 0 \tag{5.31}$$

### Equivalent Monochannel Problem

**Result 7** *(5.31) is equivalent to finding $A'$ and $h'_c$ such that:*

$$\mathcal{T}(h_c)A' = \mathcal{T}(h'_c)A, \text{ with } A'_K = 0 \tag{5.32}$$

*which corresponds to a blind monochannel problem.*

Assume that $\mathbf{H}'(z) = \mathbf{H}'_I(z)\mathrm{H}'_c(z)$, $\mathbf{H}'_I(z)$ is of length $N'_I$ and $\mathrm{H}'_c(z)$ of length $N'_c$, $N'_I + N'_c - 1 = N$. $\underline{M}_I$ and $\underline{M}'_I$ are defined as:

$$\begin{cases} \underline{M}_I = \min\{M : \mathcal{T}_M(h_I) \text{ has full column rank}\} \\ \underline{M}'_I = \min\{M : \mathcal{T}_M(h'_I) \text{ has full column rank}\} \end{cases} \tag{5.33}$$

Let $\mathcal{X}$ be a matrix of length $(\underline{M}''_I+1) \times (\underline{M}''_I+N''_I)$, with $\underline{M}''_I = \max(\underline{M}_I, \underline{M}'_I)$ and $N''_I = \max(N_I, N'_I)$: $\mathcal{X}$ is filled with the element of $\boldsymbol{X} = \mathcal{T}(h)A' = \mathcal{T}(h')A$. $\mathcal{A}$ is a $(\underline{M}''_I + N''_I) \times (M - \underline{M}''_I)$ matrix.

$$\mathcal{X} = \begin{bmatrix} \boldsymbol{x}(M-1) & \cdots & \boldsymbol{x}(\underline{M}''_I) \\ \vdots & \ddots & \vdots \\ \boldsymbol{x}(M+\underline{M}''_I-1) & \cdots & \boldsymbol{x}(0) \end{bmatrix}, \mathcal{A} = \begin{bmatrix} a(M-1) & \cdots & a(\underline{M}''_I) \\ \vdots & \ddots & \vdots \\ a(M-\underline{M}''_I-N'_I) & \cdots & a(-N'_I+1) \end{bmatrix}$$

$$\mathcal{A}' = \begin{bmatrix} a'(M-1) & \cdots & a'(\underline{M}''_I) \\ \vdots & \ddots & \vdots \\ a'(M-\underline{M}''_I-N_I) & \cdots & a'(-N_I+1) \end{bmatrix}. \tag{5.34}$$

$$\mathcal{X} = \mathcal{T}_{\underline{M}_I''+1}(h_I)\mathcal{T}_{\underline{M}_I''+N_I}(h_c)\mathcal{A}' = \mathcal{T}_{\underline{M}_I''+1}(h_I')\mathcal{T}_{\underline{M}_I''+N_I'}(h_c')\mathcal{A} \tag{5.35}$$

If $M \geq 2\underline{M}_I'' + N_I'$ and $\mathcal{A}$ has at least $L_I'' + N_I'$ modes that are not zeros of $\mathrm{H}_c'(z)$, then $\mathcal{T}_{\underline{M}_I''+N_I'-1}(h_c')\mathcal{A}$ has full row rank and:

$$(5.35) \Leftrightarrow \text{range}\left\{\mathcal{T}_{\underline{M}_I''+1}(h_I')\right\} \subset \text{range}\left\{\mathcal{T}_{\underline{M}_I''+1}(h_I)\right\} \tag{5.36}$$

Using theorem 1, we have then $\boldsymbol{H}_I'(z) = \alpha \boldsymbol{H}_I(z)$. We conclude that $\mathbf{H}'(z) = \mathbf{H}_I(z)\mathrm{H}_c'(z)$ ($N_c' = N_c$), $M_I'' = M_I' = M_I$, so the previous condition $M \geq 2\underline{M}_I'' + N_I'$ is equivalent to condition $(i)$. The condition that $\mathcal{A}$ should have at least $L_I'' + N_I' = L_I + N_I$ modes that are not zeros of $\mathrm{H}_c'(z)$ is verified if there are at least $L_I + N_I + N_c - 1 = L_I + N - 1$ which is condition $(ii)$.

$$(5.31) \Leftrightarrow \mathcal{T}(h_c)A' = \mathcal{T}(h_c')A, \text{ with } A_K' = 0 \tag{5.37}$$

as $\mathcal{T}(h_I)$ is full column rank.

Next, we study the singularities of the matrix: $[\mathcal{T}(h_c) \;\; \mathcal{A}_c]$, with $\mathcal{T}(h_c)A = \mathcal{A}_c h_c$.

### FIM regularity and necessary number of known symbols

**Result 8** *The FIM has exactly* $2N_c - 1$ *singularities*

**Result 9** *For the FIM to be regular, it is necessary to have* $2N_c - 1$ *known symbols. This result is valid whatever the distribution of the known symbols over the burst.*

$[\mathcal{T}(h_c) \;\; \mathcal{A}]$ is a $M \times (M + 2N_c - 1)$ matrix, which has at least $2N_c - 1$ singularities: the $M \times M$ submatrix $\mathcal{T}_{2N-1}(h_c)$, where $\mathcal{T}_{2N-1}(h_c)$ is the version of $\mathcal{T}(h_c)$ with the $2N_c - 1$ first columns removed, is a triangular matrix with non-zero elements on the diagonal and hence is invertible. The matrix has then exactly $2N_c - 1$ singularities, and $2N_c - 1$ known symbols are necessary for the matrix to be regular and also for identifiability. The condition $(i)$ states that $M \geq M_K - N + 1 \geq 2N_c - 1N + 1$.

**Result 10** *Let* $\left[A_s^T \;\; -h_s^T\right]^T$ *be a singular vector of the blind FIM. The semi–blind FIM is regular if and only if* $A_s$ *is not equal to 0 at the position of the known symbols:*

$$\mathcal{S}_K A_s = A_{s_K} \neq 0 \tag{5.38}$$

*This happens with probability one.*

The explicit description of these singular vectors has been omitted due to lack of space. The vector $\left[A^T \;\; -h_c^T\right]^T$ is a singular vector of the blind FIM. If $\mathcal{T}_K(h_c)A_K = 0$, the FIM is singular. This happens if $A_K = 0$, which case is not considered here, or if $h_c$ is a singular vector of $\mathcal{A}_K$, with $\mathcal{T}_K(h_c)A_K = \mathcal{A}_K h_c$. This latter case will occur with probability 0 in general. However, if one assumes the $A_K$ to have at least $N_c$ modes then $\mathcal{A}_K$ is of full rank and $\mathcal{A}_K h_c$ cannot be equal to 0.

**Grouped Known Symbols** Assume that the known symbols are grouped and, for simplicity reasons, are situated at the beginning of the burst. Considering the $N_c$ first equations of (5.31), with $A'_K = 0$ we get:

$$0 = \mathcal{T}_{TS}(h)A_K = \mathcal{A}_{TS}h \tag{5.39}$$

which is impossible if the known symbols have at least $2N_c - 1$ modes In this case, the FIM is always regular.

# B Semi–Blind Identifiability Conditions for All–Zero Arbitrarily Dispersed Known Symbols

We examine the conditions for the semi–blind FIM to be 1-singular in the case where the known symbols are all equal to 0. We examine only the monochannel case which is sufficient to solve the general multichannel case. The conditions on burst length and number of modes are then the same as for the case treated in Appendix A.

We assume that we have $2N_c - 2$ zero known symbols in the burst. If among these known symbols, $N_c$ are grouped, then the matrix $[\mathcal{T}_U(h_c)\ \mathcal{A}]$ will have one row equal to zero. Its rank is at most $M - 1$: the FIM is at least 2-singular and there is no identifiability.

If there are no $N$ grouped known symbols, there will be no row equal to zero and there is at least 1 singularity. Now, we eliminate one of the columns of $\mathcal{T}_U(h_c)$ to get $\mathcal{T}_{U'}(h_c)$: this columns is chosen in order not to have $N$ consecutive columns removed from $\mathcal{T}(h_c)$. Then, with probability 1, the column space of $\mathcal{A}$ does not belong to $\mathcal{T}_{U'}(h_c)$, according to Appendix A and the FIM is exactly 1-singular.

# C Asymptotical Equivalence of DML and GML

Up to $1^{st}$ order in $\dfrac{\sigma_v^2}{\sigma_a^2}$:

$$\sigma_v^2 C_{YY}^{-1} = P^{\perp}_{\mathcal{T}_U(h)} + \frac{\sigma_v^2}{\sigma_a^2}\mathcal{T}_U(h)\left[\mathcal{T}_U^H(h)\mathcal{T}_U(h)\right]^{-2}\mathcal{T}_U^H(h) \tag{5.40}$$

Then for the Gaussian FIM, we have:

$$J_{hh}^{Gaus}(\theta)[i,j] = \frac{1}{\sigma_v^2}\left(\mathcal{A}_K^H P^{\perp}_{\mathcal{T}_U(h)}\mathcal{A}_K\right)[i,j] + \frac{\sigma_a^2}{\sigma_v^2}\mathrm{tr}\left\{\mathcal{T}_U^H\left(\frac{\partial h}{\partial h_i}\right)P^{\perp}_{\mathcal{T}_U(h)}\mathcal{T}_U\left(\frac{\partial h}{\partial h_j}\right)\right\}. \tag{5.41}$$

For $M_U \to \infty$, by the law of large numbers, the deterministic semi–blind FIM is equivalent to its expected value w.r.t. $A_U$:

$$J_{hh}^{Det}(\theta)[i,j] = \frac{1}{\sigma_v^2}\mathrm{E}_{A_U}\left(\mathcal{A}^H P^{\perp}_{\mathcal{T}_U(h)}\mathcal{A}\right)[i,j] = \frac{1}{\sigma_v^2}\mathrm{tr}\left\{\mathcal{T}^H\left(\frac{\partial h}{\partial h_i}\right)P^{\perp}_{\mathcal{T}_U(h)}\mathcal{T}\left(\frac{\partial h}{\partial h_j}\right)\mathrm{E}_{A_U}(AA^H)\right\}. \tag{5.42}$$

Now since $\mathrm{E}_{A_U}(AA^H) = \begin{bmatrix} A_K A_K^H & 0 \\ 0 & \sigma_a^2 I \end{bmatrix}$, we get:

$$J_{hh}^{Det}(\theta)[i,j] = \frac{1}{\sigma_v^2} \underbrace{\mathrm{tr}\left\{ A_K^H \mathcal{T}_K^H \left(\frac{\partial h}{\partial h_i}\right) P^{\perp}_{\mathcal{T}_U(h)} \mathcal{T}_K \left(\frac{\partial h}{\partial h_j}\right) A_K \right\}}_{\left(\mathcal{A}_K^H P^{\perp}_{\mathcal{T}_U(h)} \mathcal{A}_K\right)[i,j]} + \frac{\sigma_a^2}{\sigma_v^2} \mathrm{tr}\left\{ \mathcal{T}_U^H \left(\frac{\partial h}{\partial h_i}\right) P^{\perp}_{\mathcal{T}_U(h)} \mathcal{T}_U \left(\frac{\partial h}{\partial h_j}\right) \right\} . \tag{5.43}$$

# D Channels used in the Simulations

- Irreducible Channel:

$$\boldsymbol{H}_{well} = \begin{bmatrix} -0.4326-0.0280j & 0.1253-0.1584j & -1.1465+0.3366j & 1.1892-1.1715j & 0.3273+2.0161j \\ -1.6656-1.5420j & 0.2877+0.0911j & 1.1909+0.9190j & -0.0376-1.2130j & 0.1746+2.7042j \end{bmatrix}. \tag{5.44}$$

- Monochannel:

$$H_{mono} = \begin{bmatrix} 0.3899-0.9499j & 0.0880+0.7812j & -0.6355+0.5690j \\ & & -0.5596+0.8217j \quad 0.4437-0.2656j \end{bmatrix} \tag{5.45}$$

- Reducible Channel:

$$\boldsymbol{H}_I = \begin{bmatrix} -0.8051+0.5913j & 0.2193+0.3803j \\ 0.5287-0.6436j & -0.9219-1.0091j \\ -0.1461-0.3745j & -0.0766+1.7513j \\ 0.2481-0.4709j & 1.7382+0.7532j \end{bmatrix}, \quad \mathrm{H}_c = \begin{bmatrix} 1 & -0.1567+1.0565j \end{bmatrix}. \tag{5.46}$$

- Nearly reducible channel:

$$\mathbf{H}_{ill}(z) = \begin{bmatrix} 1.1909 z^{-3}(0.3-z)(-0.8-z)(0.21-z) \\ 1.1892 z^{-3}(0.32-z)(-0.53-z)(1.02-z) \end{bmatrix} \tag{5.47}$$

# E Numerical Evaluations of the CRBs

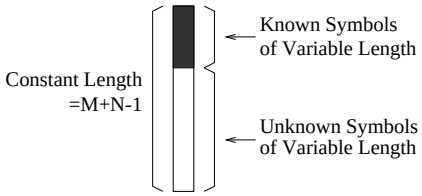



Figure 5.1: Input burst for the semi–blind mode.

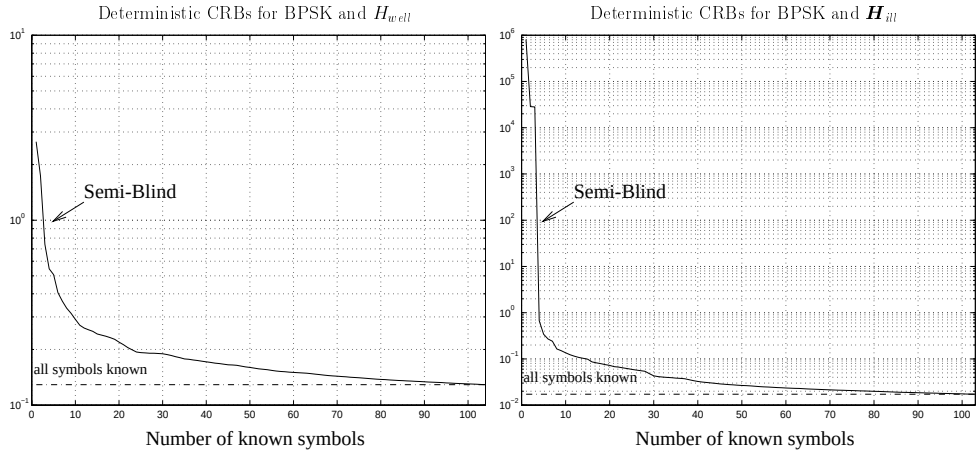



Figure 5.2: CRBs for deterministic semi–blind channel estimation w.r.t. the number of known symbols for $\boldsymbol{H}_{well}$ and $\boldsymbol{H}_{ill}$.

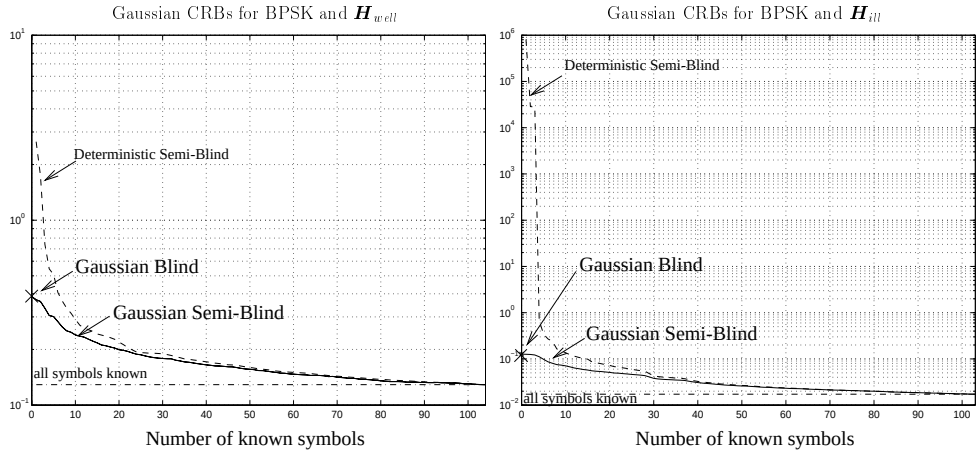



Figure 5.3: CRBs for Gaussian semi–blind channel estimation w.r.t. the number of known symbols for $\boldsymbol{H}_{well}$ and $\boldsymbol{H}_{ill}$.

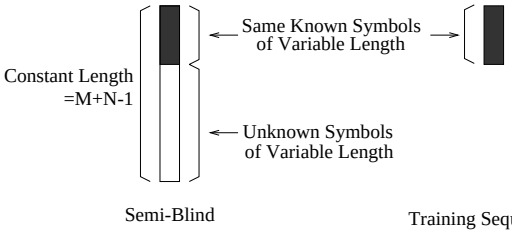



Figure 5.4: Input burst for the comparison between semi–blind and TS mode.

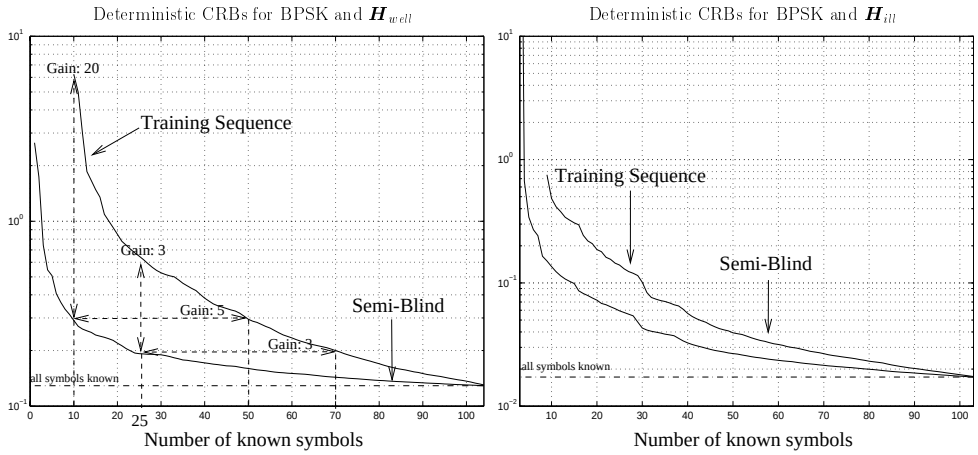



Figure 5.5: Comparison between deterministic semi–blind and TS channel estimation for $\boldsymbol{H}_{well}$ and $\boldsymbol{H}_{ill}$.

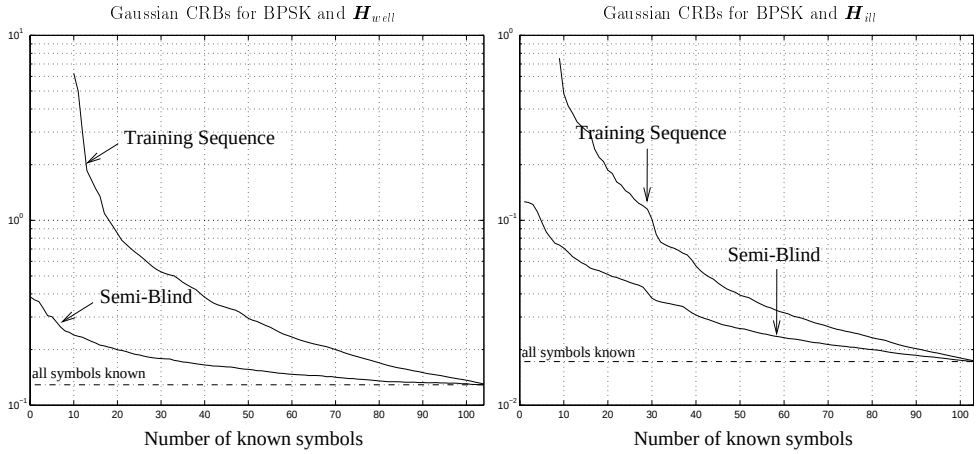



Figure 5.6: Comparison between Gaussian semi–blind and TS channel estimation for $\boldsymbol{H}_{well}$ and $\boldsymbol{H}_{ill}$.

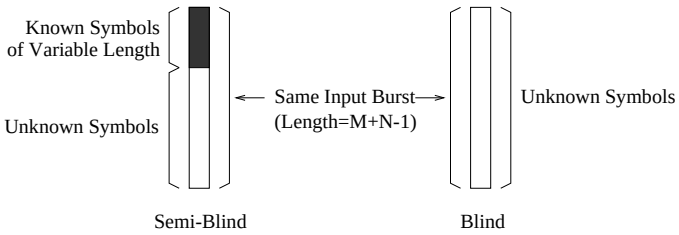



Figure 5.7: Input burst for the comparison between semi–blind and blind model.

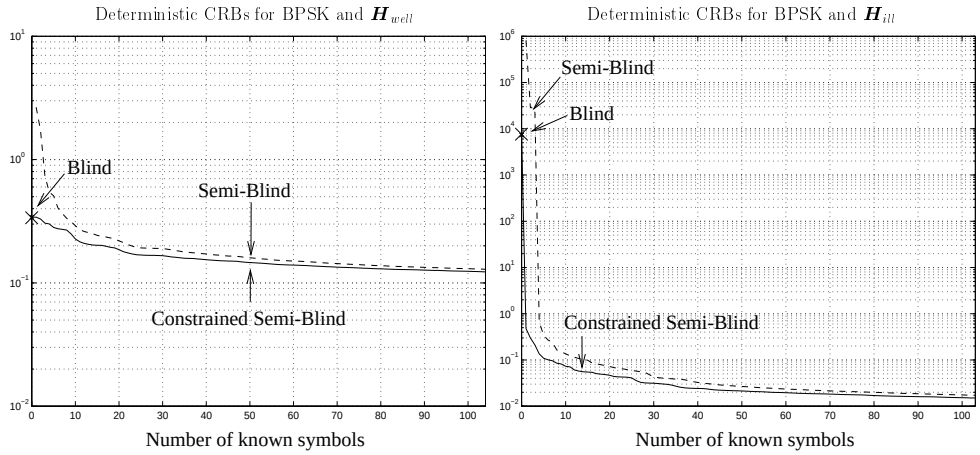



Figure 5.8: Comparison between deterministic blind and semi–blind channel estimation for $\boldsymbol{H}_{well}$ and $\boldsymbol{H}_{ill}$.

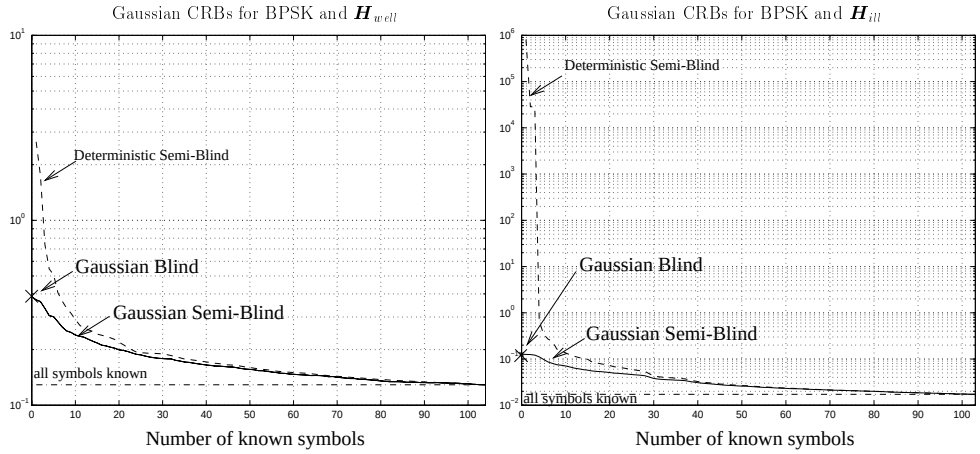



Figure 5.9: Comparison between Gaussian blind and semi–blind channel estimation for $\boldsymbol{H}_{well}$ and $\boldsymbol{H}_{ill}$.

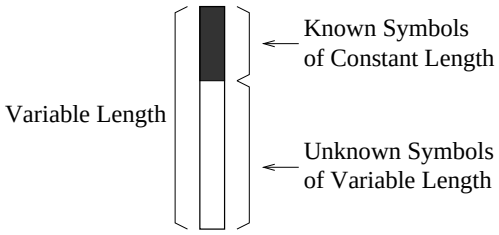



Figure 5.10: Input burst for the semi–blind mode.

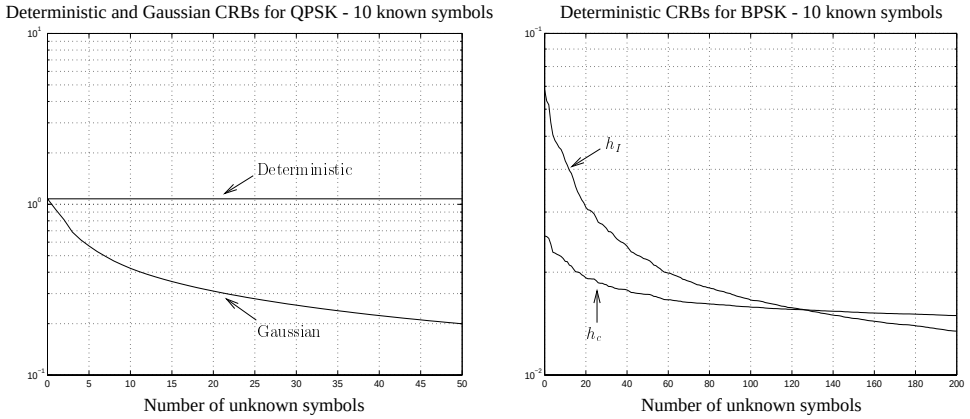



Figure 5.11: CRBs for deterministic and Gaussian semi–blind monochannel estimation (left) and deterministic CRBs for a reducible channel w.r.t. the number of unknown symbols. 10 symbols are known.

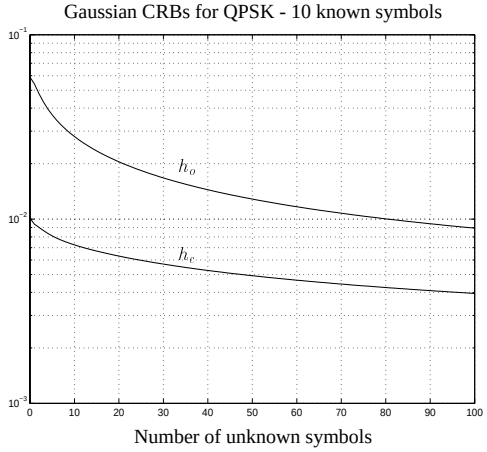



Figure 5.12: CRBs for Gaussian semi–blind channel estimation for a reducible channel w.r.t. the number of unknown symbols. 10 symbols are known.

# Part II

# Blind and Semi–Blind Channel Identification Methods

Chapter 6

# ASYMPTOTIC PERFORMANCE OF DETERMINISTIC ML AND GAUSSIAN ML

*Blind and semi–blind Deterministic ML (DML) and Gaussian ML (GML) are formulated. Their performance are derived for an asymptotic number of unknown symbols for blind estimation as well as an asymptotic number of known symbols for semi–blind estimation. The case of high SNR is also mentioned. The known symbols are considered as grouped. We express the performance w.r.t. the CRB: we prove that DML performance is above the deterministic CRB whereas GML performance is below the Gaussian CRB (computed based on the Gaussian distribution for the input symbols); both DML and GML attain the CRB at high SNR. We analyze how the information brought by the training sequence combines with the blind information and how both information are partioned between the different parameters to estimate: in particular, we study the role of blind and training sequence information in the estimation of the zeros of a reducible channel. Furthermore, the superiority of the Gaussian model is demonstrated.*

## 6.1 ML Methods

### 6.1.1 Deterministic ML (DML)

In the deterministic model, $\boldsymbol{Y} \sim \mathcal{N}(\mathcal{T}(h)A, \sigma_v^2 I)$, then the DML criterion for $\theta = [A_U^T \;\; h^T]^T$ is:

$$\max_{A_U,h} f(\boldsymbol{Y}|h) \quad \Leftrightarrow \quad \min_{A_U,h} \|\boldsymbol{Y} - \mathcal{T}(h)A\|^2 . \tag{6.1}$$

$f(\boldsymbol{Y}|h)$ is the complex probability density function when $A$ is complex and the real one when $A$ is real. $\mathcal{T}(h)A = \mathcal{T}_K(h)A_K + \mathcal{T}_U(h)A_U$, and optimizing w.r.t. the unknown symbols, we get:

$$A_U = \left(\mathcal{T}_U^H(h)\mathcal{T}_U(h)\right)^{-1} \mathcal{T}_U^H(h)\left(\boldsymbol{Y} - \mathcal{T}_K(h)A_K\right) \tag{6.2}$$

which is the output of the non-causal MMSE zero-forcing decision feedback equalizer with feedback of the known symbols. Substituting (6.2) in (6.1) we get the following minimization criterion for $h$:

$$\min_h \left(\boldsymbol{Y} - \mathcal{T}_K(h)A_K\right)^H P^{\perp}_{\mathcal{T}_U(h)} \left(\boldsymbol{Y} - \mathcal{T}_K(h)A_K\right) \tag{6.3}$$

where $P^{\perp}_{\mathcal{T}_U(h)} = I - \mathcal{T}_U(h)\left(\mathcal{T}_U^H(h)\mathcal{T}_U(h)\right)^{-1}\mathcal{T}_U^H(h)$. We denote $\mathcal{F}(\theta)$ the cost function, with $\theta = h$. For commodity reasons, when $A$ is complex, it is taken equal to $\dfrac{1}{\sigma_v^2}$ times the expression in (6.1), when $A$ is real it is $\dfrac{2}{\sigma_v^2}$ times this expression.

### 6.1.2 Gaussian ML (GML)

In the Gaussian model, $\boldsymbol{Y} \sim \mathcal{N}(\mathcal{T}_K(h)A_K, C_{YY})$, $C_{YY} = \sigma_a^2\mathcal{T}_U(h)\mathcal{T}_U^H(h) + \sigma_v^2 I$ and the GML criterion is $\max\limits_{h,\sigma_v^2} f(Y|h)$, or:

$$\min_{h,\sigma_v^2} \left\{ \ln\det C_{YY} + \left(\boldsymbol{Y} - \mathcal{T}_K(h)A_K\right)^H C_{YY}^{-1}\left(\boldsymbol{Y} - \mathcal{T}_K(h)A_K\right) \right\} . \tag{6.4}$$

We denote $\mathcal{F}(\theta)$ the cost function, with $\theta = \left[h^T \;\; \sigma_v^2\right]^T$. When $A$ is complex, it is taken equal to the expression in (6.4), when $A$ is real, it is 2 times this expression.

## 6.2 Asymptotic Performance

The asymptotic semi–blind conditions will be:

(i) $M_K \to \infty$ and $M_U \to \infty$

(ii) $\dfrac{\sqrt{M_U}}{M_K} \to 0$.

In the blind case, the condition is only $M = M_U \to \infty$. The second condition indicates that the training sequence part of the semi–blind criterion is not negligible w.r.t. the blind part as will be seen later.

The fact of considering a large number of known symbols may appear artificial as this number will be in general small. In chapters 8 and 9, we compare the simulated performance of ML semi–blind criteria to the theoretical performance we obtain by considering a large number of known symbols: the theoretical performance are found to make sense.

We will assume that the known symbols are located at the beginning of the burst.

### 6.2.1 Regular Estimation Case

We assume that the parameters are identifiable. Let $\theta$ be the complex parameter vector, $\theta_R = [\mathrm{Re}(\theta^T)\ \mathrm{Im}(\theta^T)]^T$, the real associated parameter vector ($\theta_R = \theta$ for real parameters), $\theta^o$ and $\theta_R^o$ the true values, $\hat{\theta}$ and $\hat{\theta}_R$ the ML estimates and $\Delta\theta = \hat{\theta} - \theta^o$, $\Delta\theta_R = \hat{\theta}_R - \theta_R^o$, the errors. We denote:

$$\mathcal{F}'(\theta) = \frac{\partial \mathcal{F}(\theta)}{\partial \theta_R} \qquad \text{and} \qquad \mathcal{F}'(\theta) = \frac{\partial}{\partial \theta_R}\left(\frac{\partial \mathcal{F}(\theta)}{\partial \theta_R}\right)^T . \tag{6.5}$$

We assume that the parameter estimation is consistent and that the first and second derivative of $\mathcal{F}(\theta)$ exist and are continuous. We can then proceed to the following Taylor development of $\mathcal{F}'(\theta)$ around $\theta^o$.

$$\mathcal{F}'(\hat{\theta}) = 0 = \mathcal{F}'(\theta^o) + \mathcal{F}''(\theta^o)\Delta\theta_R + o(\Delta_R) . \tag{6.6}$$

We call $\mathcal{F}''_\infty(\theta^o)$ the limit value of $\mathcal{F}''(\theta^o)$:

$$\mathcal{F}''_\infty(\theta^o) = \lim_{\substack{M_U \to \infty \\ M_K \to \infty}} \mathcal{F}''(\theta^o) . \tag{6.7}$$

Asymptotically $\mathcal{F}''(\theta^o) = \mathcal{F}''_\infty(\theta^o) + o(\mathcal{F}''(\theta^o))$, then, neglecting the first order terms, equation (6.6) becomes:

$$0 = \mathcal{F}'(\theta^o) + \mathcal{F}''_\infty(\theta^o)\Delta\theta_R \ \Rightarrow \ \left\{ \begin{array}{lcl} \Delta\theta_R & = & - \ [\mathcal{F}''_\infty(\theta^o)]^{-1}\, \mathcal{F}'(\theta^o) \\ C_{\Delta\theta_R \Delta\theta_R} & = & [\mathcal{F}''_\infty(\theta^o)]^{-1}\, \mathrm{E}\left[\mathcal{F}'(\theta^o)\mathcal{F}'(\theta^o)\right]\, [\mathcal{F}''_\infty(\theta^o)]^{-1} \end{array} \right. . \tag{6.8}$$

Let us define the matrices:

$$\mathcal{J}^{(1)}_{\theta_R\theta_R} = E\left(\frac{\partial \mathcal{F}(\theta)}{\partial \theta_R}\right)\left(\frac{\partial \mathcal{F}(\theta)}{\partial \theta_R}\right)^H \quad \text{and} \quad \mathcal{J}^{(2)}_{\theta_R\theta_R} = -E\frac{\partial}{\partial \theta_R}\left(\frac{\partial \mathcal{F}(\theta)}{\partial \theta_R}\right)^H . \tag{6.9}$$

These two matrices are Fisher-like information matrices. Indeed, if $\mathcal{F}(\theta) = \ln f(\boldsymbol{Y}, \theta)$, the distribution of the observations, $\mathcal{J}^{(1)}_{\theta_R\theta_R} = \mathcal{J}^{(2)}_{\theta_R\theta_R}$ and is equal to the Fisher Information Matrix.

In our specific cases, asymptotically, by the law of large numbers, the different quantities involved will be asymptotically equivalent to their expected value, and particularly:

$$\mathcal{F}''_{\infty}(\theta^o) = \mathcal{J}^{(2)}_{\theta_R\theta_R}\Big|_{\theta=\theta^o} . \tag{6.10}$$

(To simplify, we give up the subscript $\theta = \theta^o$ in the following). Furthermore, $\mathcal{F}'(\theta^o)$ will be a centered Gaussian random variable, so:

$$\left\{ \begin{array}{lcl} \Delta\theta_R & \sim & \mathcal{N}\left(0, C_{\Delta\theta_R\Delta\theta_R}\right) \\ C_{\Delta\theta_R\Delta\theta_R} & = & \left(\mathcal{J}^{(2)}_{\theta_R\theta_R}\right)^{-1} \left(\mathcal{J}^{(1)}_{\theta_R\theta_R}\right) \left(\mathcal{J}^{(2)}_{\theta_R\theta_R}\right)^{-1} . \end{array} \right. \tag{6.11}$$

When $\theta$ is complex, $\mathcal{J}_{\theta_R\theta_R}$ can be expressed w.r.t. the complex IMs:

$$J^{(1)}_{\varphi\psi} = E\left(\frac{\partial\mathcal{F}(\theta)}{\partial\varphi^*}\right)\left(\frac{\partial\mathcal{F}(\theta)}{\partial\psi^*}\right)^H \quad \text{and} \quad J^{(2)}_{\varphi\psi} = -E\frac{\partial}{\partial\varphi^*}\left(\frac{\partial\mathcal{F}(\theta)}{\partial\psi^*}\right)^H . \tag{6.12}$$

$\mathcal{J}^{(1)}_{\theta_R\theta_R}$ and $\mathcal{J}^{(2)}_{\theta_R\theta_R}$ can be expressed in terms of $J_{\theta\theta}$ and $J_{\theta\theta^*}$:

$$\mathcal{J}^{(1)}_{\theta_R\theta_R} = 2\begin{bmatrix} \mathrm{Re}(J^{(1)}_{\theta\theta}) & -\mathrm{Im}(J^{(1)}_{\theta\theta}) \\ \mathrm{Im}(J^{(1)}_{\theta\theta}) & \mathrm{Re}(J^{(1)}_{\theta\theta}) \end{bmatrix} + 2\begin{bmatrix} \mathrm{Re}(J^{(1)}_{\theta\theta^*}) & \mathrm{Im}(J^{(1)}_{\theta\theta^*}) \\ \mathrm{Im}(J^{(1)}_{\theta\theta^*}) & -\mathrm{Re}(J^{(1)}_{\theta\theta^*}) \end{bmatrix} \tag{6.13}$$

and the same for $\mathcal{J}^{(2)}_{\theta_R\theta_R}$.

In the DML case, we will have $J^{(1)}_{\theta\theta^*} = J^{(2)}_{\theta\theta^*} = 0$, so when the input symbols are complex, we can work directly with complex quantities, and (6.13) can be compactly written as:

$$C_{\Delta\theta\Delta\theta} = \left(J^{(2)}_{\theta\theta}\right)^{-1} J^{(1)}_{\theta\theta} \left(J^{(2)}_{\theta\theta}\right)^{-1} . \tag{6.14}$$

It is also possible to compute the performance at high SNR conditions. In this case, equation (6.14) is still valid, but the random part of the received signal coming from the noise only, no expectation is necessary in the expression of the IMs.

### 6.2.2 ML Performance under Constraints

The expression of the performance is well defined if the IM $J^{(2)}_{\theta\theta}$ is regular; we will see that $J^{(2)}_{\theta\theta}$ is equal to the FIM so in the blind DML or GML cases, equation (6.14) cannot be applied because the IM is singular.

We determine here performance under constraints, with results very similar to those seen for CRBs under constraints. Again, we assume that $\theta$ is real and consider the general constraint:

$$\mathcal{K}_\theta = 0 \tag{6.15}$$

where $\mathcal{K}_\theta : \mathbb{R}^n \to \mathbb{R}^k$ is continuously differentiable. As for the constrained CRBs, the constrained performance only depends on the local properties of the constraint set. Let us recall the definition of $\mathcal{M}_\theta$, the tangent to the constraint set at point $\theta$:

$$\mathcal{M}_\theta = \left\{ Z \in \mathbb{R}^n \, ; \;\; \left( \frac{\partial \mathcal{K}_\theta^T}{\partial \theta} \right)^T \; Z = 0 \right\} . \tag{6.16}$$

**Theorem 23** *The performance of ML methods under constraint (6.15) is:*

$$C_{\Delta\theta\Delta\theta} = \mathcal{V}_\theta \left( \mathcal{V}_\theta^T \mathcal{J}_{\theta\theta}^{(2)} \mathcal{V}_\theta \right)^{-1} \mathcal{V}_\theta^T \mathcal{J}_{\theta\theta}^{(1)} \mathcal{V}_\theta \;\; \left( \mathcal{V}_\theta^T \mathcal{J}_{\theta\theta}^{(2)} \mathcal{V}_\theta \right)^{-1} \mathcal{V}_\theta^T \tag{6.17}$$

*where $\mathcal{V}_\theta$ is an $n \times r$ matrix (where $r = rank(\mathcal{M}_\theta)$) whose columns form a basis of $\mathcal{M}_\theta$. A necessary and sufficient condition for the constrained performance to be defined is that $\mathcal{V}_\theta^T \mathcal{J}_{\theta\theta}^{(2)} \mathcal{V}_\theta$ be regular. Furthermore, $\Delta\theta \sim \mathcal{N}(0, C_{\Delta\theta\Delta\theta})$.*

*Proof:* The main idea, as explained in Chapter 3, is that locally a point $\theta$ verifying the constraint (6.15) in the neighborhood of $\theta^o$ can be approximated as belonging to $\mathcal{M}_{\theta^o}$.

$$\theta = \theta^o + \mathcal{V}_\theta^o \xi \, . \tag{6.18}$$

By the chain rule, keeping only the first order terms in $\Delta h$:

$$\mathcal{J}_{\xi\xi}^{(1)/(2)} = \mathcal{V}_{\theta^o}^T \mathcal{J}_{\theta\theta}^{(1)/(2)} \mathcal{V}_{\theta^o} \, . \tag{6.19}$$

The estimation of $\xi$ is regular and consistent and by (6.14):

$$C_{\Delta\xi} = \left( \mathcal{V}_{\theta^o}^T \mathcal{J}_{\theta\theta}^{(2)} \mathcal{V}_{\theta^o} \right)^{-1} \mathcal{V}_{\theta^o}^T \mathcal{J}_{\theta\theta}^{(1)} \mathcal{V}_{\theta^o} \left( \mathcal{V}_{\theta^o}^T \mathcal{J}_{\theta\theta}^{(2)} \mathcal{V}_{\theta^o} \right)^{-1} \, . \tag{6.20}$$

Now $\Delta\theta = \mathcal{V}_{\theta^o} \Delta\xi \Rightarrow C_{\Delta\theta\Delta\theta} = \mathcal{V}_{\theta^o} C_{\Delta\xi\Delta\xi} \mathcal{V}_{\theta^o}^T$ and we get (6.25).

□

As for the constrained CRB, the general constrained problem (6.15) is equivalent to the linearly constrained problem:

$$\left( \frac{\partial \mathcal{K}_\theta^T}{\partial \theta} \right)^T \Bigg|_{\theta=\theta^o} \theta = \left( \frac{\partial \mathcal{K}_\theta^T}{\partial \theta} \right)^T \Bigg|_{\theta=\theta^o} \theta^o \, . \tag{6.21}$$

If $\mathcal{J}_{\theta\theta}^{(1)} = \mathcal{J}_{\theta\theta}^{(2)} = FIM$, $C_{\Delta\theta\Delta\theta}$ is the constrained CRB (3.25):

$$C_{\Delta\theta\Delta\theta} = CRB_C = \mathcal{V}_\theta \left( \mathcal{V}_\theta^T FIM \mathcal{V}_\theta \right)^{-1} \mathcal{V}_\theta^T \, . \tag{6.22}$$

**Theorem 24** *Assume that* $\mathcal{J}^{(1)}_{\theta\theta}$ *and* $\mathcal{J}^{(2)}_{\theta\theta}$ *have the same null space (which will be the case for blind DML and GML) and assume that the constraint is such that* $\mathcal{V}_\theta = \left(\frac{\partial \mathcal{K}^T_\theta}{\partial \theta}\right)^\perp$ *spans this null space, then the corresponding performance are:*

$$C_{\Delta\theta\Delta\theta} = \left(\mathcal{J}^{(2)}_{\theta\theta}\right)^+ \mathcal{J}^{(1)}_{\theta\theta} \left(\mathcal{J}^{(2)}_{\theta\theta}\right)^+ . \tag{6.23}$$

This gives, for a minimal number of independent constraints, the minimal value for $\mathrm{tr}\{C_{\Delta\theta\Delta\theta}\}$: a proof is given in appendix A.

The rest of this study consists mainly of applying the results of this section to DML and GML.

## 6.3 Deterministic ML (DML)

### 6.3.1 Blind DML

We assume that the channel is irreducible: $\theta = h$. The two information matrices are:

$$\left\{ \begin{array}{l} J^{(1)}_{hh} = J^{(2)}_{hh} + J'_{hh} \\ J^{(2)}_{hh} = \frac{1}{\sigma^2_v} \mathcal{A}^H P^\perp_{\mathcal{T}(h)} \mathcal{A} = FIM \\ J'_{hh}(i,j) = \mathrm{tr}\left\{ \mathcal{T}^H \left(\frac{\partial h}{\partial h_i}\right) P^\perp_{\mathcal{T}(h)} \mathcal{T} \left(\frac{\partial h}{\partial h_j}\right) \left(\mathcal{T}^H(h)\mathcal{T}(h)\right)^{-1} \right\} . \end{array} \right. \tag{6.24}$$

$J^{(1)}_{hh}$ and $J^{(2)}_{hh}$ have the same singularity spanned by $h$ corresponding to the deterministic scale factor ambiguity. The associated real IMs $\mathcal{J}^{(1)}_{h_R h_R}$ and $\mathcal{J}^{(2)}_{h_R h_R}$ have the two same singularities as the FIM:

$$h_{S_1} = \left[ \begin{array}{c} \mathrm{Re}(h) \\ \mathrm{Im}(h) \end{array} \right] \quad \text{and} \quad h_{S_2} = \left[ \begin{array}{c} -\mathrm{Im}(h) \\ \mathrm{Re}(h) \end{array} \right] . \tag{6.25}$$

Blind performance are computed under the following constraints already used for the regularized blind CRBs that we recall here:

(1) The quadratic norm constraint:

$$h^H h = h^{o\,H} h^o \tag{6.26}$$

(2) In the complex case, the linear phase constraint:

$$h^{o\,T}_{S_2} h = h^{o\,T}_{S_2} h^o = 0 . \tag{6.27}$$

As for the FIM, the ambiguous sign factor left by the constraints, is adjusted using the constraint ${h^o}^H h > 0$ (this sign factor does not influence performance computation).

We prove consistency of the channel estimate under these constraints in appendix B. Following the notations of section 3.5, the constraint is:

$$\mathcal{K}_{\theta_R} = \begin{bmatrix} h_R^T h_R - {h^o}_R^T h_R^o \\ h_{S_2}^{oT} h_R \end{bmatrix} = 0\,. \tag{6.28}$$

At the true parameter value:

$$\left.\frac{\partial \mathcal{K}_{\theta_R}^T}{\partial \theta_R}\right|_{\theta=\theta^o} = \left[2h_R^o \;\; h_{S_2}^o\right], \tag{6.29}$$

which spans the noise subspace of $J_{h_R h_R}^{(1)/(2)}$. So, according to theorem 24:

$$C_{\Delta h_R \Delta h_R} = \mathcal{J}_{h_R h_R}^{(2)+} \mathcal{J}_{h_R h_R}^{(1)} \mathcal{J}_{h_R h_R}^{(2)+}\,. \tag{6.30}$$

The compact complex error correlation matrix is then:

$$C_{\Delta h \Delta h} = J_{hh}^{(2)+} J_{hh}^{(1)} J_{hh}^{(2)+}\,. \tag{6.31}$$

As seen in Chapter 4, the asymptotic CRB associated to the estimation of $h$ with $A$ as nuisance parameter with constraints (6.26), (6.27) is $J_{hh}^{(2)^+}$:

$$CRB_{hh} = J_{hh}^{(2)^+}\,. \tag{6.32}$$

Asymptotically, $\Delta h \sim \mathcal{N}(0, C_{\Delta h \Delta h})$ with:

$$C_{\Delta h \Delta h} = CRB_{hh} + J_{hh}^{(2)^+} J'_{hh} J_{hh}^{(2)^+}\,. \tag{6.33}$$

The second term in (6.32) is positive: DML for $h$ does not reach the CRB. The estimation of the channel is indeed coupled with the estimation of the unknown symbols which cannot be estimated consistently at low SNR. The coupling prevents the channel estimate from being efficient. At high SNR however, the CRB is attained (as mentioned in [19] also).

### 6.3.2 Semi-Blind DML

Under condition (i), the $N-1$ observations containing both known and unknown symbols can be neglected and the training sequence and blind contributions can be separated in the criterion (6.3) as:

$$\|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_K\|^2 + \boldsymbol{Y}_B^H P_{\mathcal{T}_B(h)}^{\perp} \boldsymbol{Y}_B \tag{6.34}$$

with $\mathcal{T}_{TS}(h) = \mathcal{T}_{M_K - N+1}(h)$ and $\mathcal{T}_B(h) = \mathcal{T}_{M-M_K}(h)$, $\boldsymbol{Y}_{TS}$ and $\boldsymbol{Y}_B$ designate resp. the observations with known and unknown symbols only (see figure 6.1). Condition (ii) allows

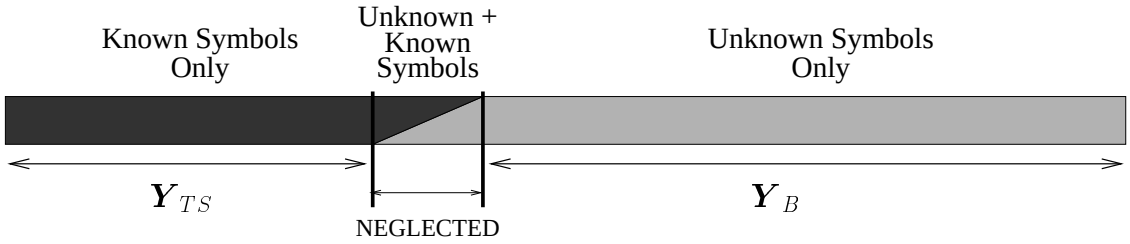



Figure 6.1: Output burst: overlap zone containing known and unknown symbols is neglected

the training sequence part to be non negligible w.r.t. the blind part. Indeed by the law of large number (that we can apply here because the known symbols are assumed grouped), the DML cost function (6.3) is equivalent to:

$$M_K \left( \mathrm{E} \left\| \boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_K \right\|^2 + O\left( \frac{1}{\sqrt{M_K}} \right) \right) + M_U \left( \mathrm{E} \left( \boldsymbol{Y}_B^H P^{\perp}_{\mathcal{T}_B(h_I)} \boldsymbol{Y}_B \right) + O\left( \frac{1}{\sqrt{M_U}} \right) \right) . \tag{6.35}$$

The training sequence term is not negligible w.r.t. the blind part if $\frac{\sqrt{M_U}}{M_K} \to 0$. Consistency of the estimation of $h$ is proved in appendix B, and result (6.11) can be applied to DML. The information matrices decompose as a training sequence and a blind part:

$$\begin{cases} J^{(1)}_{hh} = J^{(1)}_{hh,TS} + J^{(1)}_{hh,B} \\ J^{(2)}_{hh} = J^{(2)}_{hh,TS} + J^{(2)}_{hh,B} \end{cases} \tag{6.36}$$

$$\begin{cases} J^{(1)}_{hh,TS} = J^{(2)}_{hh,TS} = \dfrac{1}{\sigma_v^2} \mathcal{A}_{TS}^H \mathcal{A}_{TS} \\ J^{(1)}_{hh,B} = J^{(2)}_{hh,B} + J'_{hh,B} \\ J^{(2)}_{hh,B} = \dfrac{1}{\sigma_v^2} \mathcal{A}_B^H P^{\perp}_{\mathcal{T}_B(h)} \mathcal{A}_B \\ J'_{hh,B}(i,j) = \mathrm{tr} \left\{ \mathcal{T}_B^H \left( \dfrac{\partial h}{\partial h_i} \right) P^{\perp}_{\mathcal{T}_B(h)} \mathcal{T}_B \left( \dfrac{\partial h}{\partial h_j} \right) \left( \mathcal{T}_B^H(h) \mathcal{T}_B(h) \right)^{-1} \right\} \end{cases} \tag{6.37}$$

The CRB for $h$ is:

$$CRB_{hh} = \left( J^{(2)}_{hh} \right)^{-1} . \tag{6.38}$$

Using equation (6.14), $\Delta h \sim \mathcal{N}(0, C_{\Delta h})$ with:

$$C_{\Delta h \Delta h} = CRB_{hh} + \left( J^{(2)}_{hh} \right)^{-1} J'_{hh} \left( J^{(2)}_{hh} \right)^{-1} . \tag{6.39}$$

The second term in (6.39) is positive: semi–blind DML for $h$ does not reach the CRB.

At high SNR however, we can prove that the CRB is attained and:

$$C_{\Delta h \Delta h} = \sigma_v^2 \left[ \mathcal{A}_{TS}^H \mathcal{A}_{TS} + \mathcal{A}_B^H P^{\perp}_{\mathcal{T}_B(h)} \mathcal{A}_B \right]^{-1} . \tag{6.40}$$

### Monochannel

Assume that $\mathbf{H}(z)$ is a monochannel, $P^{\perp}_{\mathcal{T}_U(h)} = \begin{bmatrix} 0 & 0 \\ 0 & I_{M_K-N+1} \end{bmatrix}$. The DML criterion is reduced to the training sequence criterion:

$$\min_h \|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h)A_K\|^2 . \tag{6.41}$$

The monochannel gets then estimated by training and the blind part of the criterion brings no information.

### Reducible Channel

We study the general case of a reducible channel. We assume that we have detected the structure of the channel (the number of zeros): $\mathbf{H}(z) = \mathbf{H}_I(z)\mathrm{H}_c(z)$ and study the role of the blind and training sequence parts in the estimation of $\mathbf{H}_I(z)$ and of the zeros. The estimation parameter is here: $\theta = [h_I^T\ \bar{h}_c^T]^T$, where $\bar{h}_c$ is deduced from $h_c$ by eliminating its first element equal to 1.

It can be verified that $P^{\perp}_{\mathcal{T}_U(h)} = P^{\perp}_{\mathcal{T}_{U'}(h_I)}$ where $\mathcal{T}_{U'}(h_I)$ is $\mathcal{T}(h_I)$ with the last $M_K - N_c + 1$ columns removed. The semi–blind cost function becomes then:

$$\|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h)A_K\|^2 + \boldsymbol{Y}_B^H P^{\perp}_{\mathcal{T}_B(h_I)}\boldsymbol{Y}_B . \tag{6.42}$$

The information matrices are:

$$\begin{cases} J^{(1)}_{h_Ih_I} = J^{(1)}_{h_Ih_I,TS} + J^{(1)}_{h_Ih_I,B} \\ J^{(2)}_{h_Ih_I} = J^{(2)}_{h_Ih_I,TS} + J^{(2)}_{h_Ih_I,B} \\ J^{(1)}_{h_ch_c} = J^{(2)}_{h_ch_c} = J_{h_ch_c,TS} \end{cases} \tag{6.43}$$

$$\begin{cases} J^{(1)}_{h_Ih_I,TS} = J^{(2)}_{h_Ih_I,TS} = \dfrac{1}{\sigma_v^2}\mathcal{A}^H_{I_{TS}}P^{\perp}_{\bar{\mathcal{A}}_{TS}}\mathcal{A}_{I_{TS}} \\ J^{(1)}_{h_Ih_I,B} = J^{(2)}_{h_Ih_I,B} + J'_{h_Ih_I,B} \\ J^{(2)}_{h_Ih_I,B} = \dfrac{1}{\sigma_v^2}\mathcal{A}^H_{I_B}P^{\perp}_{\mathcal{T}_B(h_I)}\mathcal{A}_{I_B} \\ J'_{h_Ih_I,B}(i,j) = \mathrm{tr}\left\{\mathcal{T}_B^H\left(\dfrac{\partial h_I}{\partial h_{I_i}}\right)P^{\perp}_{\mathcal{T}_B(h_I)}\mathcal{T}_B\left(\dfrac{\partial h_I}{\partial h_{I_j}}\right)(\mathcal{T}_B^H(h_I)\mathcal{T}_B(h_I))^{-1}\right\} \\ J_{h_ch_c,TS} = \dfrac{1}{\sigma_v^2}\mathcal{A}^H_{c_{TS}}\mathcal{T}^H_{TS}(h_I)\left[I - \mathcal{A}^H_{I_{TS}}\left(\mathcal{A}^H_{I_B}P^{\perp}_{\mathcal{T}_B(h_I)}\mathcal{A}_{I_B}\right)^{-1}\mathcal{A}_{I_{TS}}\right]\mathcal{T}_{TS}(h_I)\mathcal{A}_{c_{TS}} \end{cases} \tag{6.44}$$

with notations defined by:

$$\begin{cases} \bar{\mathcal{A}}_{TS} = \mathcal{T}_{TS}(h_I)\bar{\mathcal{A}}_{c_{TS}} \quad \text{and} \quad \mathcal{T}_{TS}(\bar{h}_c)A_K = \bar{\mathcal{A}}_{c_{TS}}\bar{h}_c \\ \mathcal{T}_{TS}(h_I)\mathcal{T}_{TS}(h_c)A_K = \mathcal{A}_{I_{TS}}h_I \quad \text{and} \quad \mathcal{T}_B(h_I)\mathcal{T}_B(h_c)A_U = \mathcal{A}_{I_B}h_I . \end{cases} \tag{6.45}$$

The CRB for $h_I$ and $\bar{h}_c$ are asymptotically:

$$CRB_{h_I h_I} = \left(J^{(2)}_{h_I h_I}\right)^{-1} \text{ and } CRB_{\bar{h}_c \bar{h}_c} = \left(J^{(1)}_{\bar{h}_c \bar{h}_c}\right)^{-1} . \tag{6.46}$$

Using equation (6.14), $\Delta h_I \sim \mathcal{N}(0, C_{\Delta h_I \Delta h_I})$ and:

$$C_{\Delta h_I \Delta h_I} = CRB_{h_I h_I} + \left(J^{(2)}_{h_I h_I}\right)^{-1} J'_{h_I h_I} \left(J^{(2)}_{h_I h_I}\right)^{-1} . \tag{6.47}$$

$\Delta \bar{h}_c \sim \mathcal{N}(0, C_{\Delta \bar{h}_c \Delta \bar{h}_c})$, hence $\Delta \bar{h}_c = O_p\left(\frac{1}{\sqrt{M_K}}\right)$ and

$$C_{\Delta \bar{h}_c \Delta \bar{h}_c} = CRB_{\bar{h}_c \bar{h}_c} + \left(\bar{\mathcal{A}}^H_{TS} \bar{\mathcal{A}}_{TS}\right)^{-1} \bar{\mathcal{A}}^H_{TS} \mathcal{A}_{I_{TS}} \left(J^{(2)}_{h_I h_I}\right)^{-1} J'_{h_I h_I} \left(J^{(2)}_{h_I h_I}\right)^{-1} \mathcal{A}^H_{I_{TS}} \bar{\mathcal{A}}_{TS} \left(\bar{\mathcal{A}}^H_{TS} \bar{\mathcal{A}}_{TS}\right)^{-1} . \tag{6.48}$$

Again, the CRB for both $h_I$ and $\bar{h}_c$ is not attained. The training part and the blind part of semi–blind DML brings information to the estimation $h_I$, and $\Delta h_I$ evolves asymptotically in $\frac{1}{\sqrt{M_K}}$ as $M_K$ grows to infinity, and in $\frac{1}{\sqrt{M_U}}$ as $M_U$ grows to infinity.

Blind information asymptotically plays no role in the estimation of the zeros of the channel. When $M_U \to \infty$, $C_{\Delta \bar{h}_c \Delta \bar{h}_c}$ tends to a constant: asymptotically the zeros of the channel are only estimated by training. At high SNR again, the different CRBs are attained.

A thiner analysis could have been done in considering separately the estimation of the scale factor that cannot be blindly identified and a "normalized" version of the irreducible part of the channel: $h_I = \alpha_I \bar{h}_I$: The general results described below for the estimation of the coefficients of $\bar{h}_c$ are also valid for the estimation of the scale factor.

## 6.4 Gaussian ML (GML)

We will treat directly the case of a reducible channel. $\theta = \left[h_I^T \;\; \bar{h}_c^T \;\; \sigma_v^2\right]^T$. The GML cost function is the probability density function of $\boldsymbol{Y}$ with the input symbols considered as Gaussian random variables, but not with the right distribution, so $\mathcal{J}^{(1)}_{\theta_R \theta_R} \neq \mathcal{J}^{(2)}_{\theta_R \theta_R} \neq FIM$.

In appendix B, we prove the consistency of the parameter estimate by blind or semi–blind GML. As $J^{(1)/(2)}_{\theta \theta^*} \neq 0$, the complex CRB cannot be used here and we will distinguish between complex and real parameters, *i.e.* complex and real input constellations.

### 6.4.1 Blind GML

When the input symbols are complex, let $h_R = [\mathrm{Re}(h_I^T) \;\; \mathrm{Im}(h_I^T) \;\; \mathrm{Re}(\bar{h}_c^T) \;\; \mathrm{Im}(\bar{h}_c^T)]^T$ and $\overline{\theta}_R = [h_R^T \;\; \sigma_v^2]^T$, the estimation parameter. $\mathcal{J}^{(1)}_{\overline{\theta}_R \overline{\theta}_R}$ and $\mathcal{J}^{(2)}_{\overline{\theta}_R \overline{\theta}_R}$ can be computed thanks to the quantities:

$$\begin{cases} J^{(2)}_{\theta\theta} = J_{\theta\theta} \\ J^{(1)}_{\theta\theta} = J^{(2)}_{\theta\theta} - J'_{\theta\theta} = J_{\theta\theta} - J'_{\theta\theta} . \end{cases} \tag{6.49}$$

$$\left\{\begin{array}{l} J^{(2)}_{\theta\theta}(i,j) = \mathrm{tr}\left\{ C_{YY}^{-1}\left(\dfrac{\partial C_{YY}}{\partial \theta_i^*}\right) C_{YY}^{-1}\left(\dfrac{\partial C_{YY}}{\partial \theta_j^*}\right)^H \right\} \\ J'_{\theta\theta}(i,j) = \left[(\mathrm{E}a_k^2)^2 - \mathrm{E}a_k^4\right] \mathrm{tr}\left\{ \mathcal{T}^H(h) C_{YY}^{-1}\left(\dfrac{\partial C_{YY}}{\partial \theta_i^*}\right) C_{YY}^{-1}\mathcal{T}(h)\mathcal{T}^H(h) C_{YY}^{-1}\left(\dfrac{\partial C_{YY}}{\partial \theta_j^*}\right)^H \mathcal{T}(h) \right\} \\ +(\mathrm{E}a_k^2)^2 \mathrm{diag}^T\left\{ \mathcal{T}^H(h) C_{YY}^{-1}\left(\dfrac{\partial C_{YY}}{\partial \theta_i^*}\right) C_{YY}^{-1}\mathcal{T}(h)\right\} \mathrm{diag}\left\{ \mathcal{T}^H(h) C_{YY}^{-1}\left(\dfrac{\partial C_{YY}}{\partial \theta_j^*}\right)^H C_{YY}^{-1}\mathcal{T}(h)\right\} \end{array}\right. \tag{6.50}$$

$$\text{where:} \quad \left\{\begin{array}{l} \dfrac{\partial C_{YY}}{\partial h^*_{I_i}} = \sigma_a^2 \mathcal{T}(h_I)\mathcal{T}^H(h_c)\mathcal{T}^H\left(\dfrac{\partial h_I^*}{\partial h^*_{I_i}}\right) \\ \dfrac{\partial C_{YY}}{\partial \bar{h}^*_{c_i}} = \sigma_a^2 \mathcal{T}(h_I)\mathcal{T}^H\left(\dfrac{\partial h_c^*}{\partial \bar{h}^*_{c_i}}\right)\mathcal{T}^H(h_I) \\ \dfrac{\partial C_{YY}}{\partial \sigma_v^2} = \dfrac{1}{2} I\,. \end{array}\right. \tag{6.51}$$

diag(.) designate here the vector of the diagonal elements of its argument. The same kind of relations as (6.49) and (6.50) hold also for $J^{(1)/(2)}_{\theta\theta^*}$:

Due to the continuous phase factor ambiguity, $\mathcal{J}^{(2)}_{\overline{\theta}_R\overline{\theta}_R}$ is singular with one singularity spanned by $\theta_S = [h_{S_2}^T \;\; 0^T_{(N_c-1)\times 1} \;\; 0]^T$. Blind estimation performance are computed under the linear constraint:

$$\theta_S^{oH}\theta_R = 0 \quad \Leftrightarrow \quad h_{S_2}^{oH} h_R = 0 \tag{6.52}$$

which allows to determine the phase factor up to a sign factor.

This constraint gives the minimal value for the performance of the estimation on $\overline{\theta}_R$. $C_{\Delta h_{I_R}\Delta h_{I_R}}$ and $C_{\Delta \bar{h}_{c_R}\Delta \bar{h}_{c_R}}$ are the appropriate submatrix of:

$$\begin{array}{rcl} C_{\overline{\theta}_R\overline{\theta}_R} & = & \left(\mathcal{J}^{(2)}_{\overline{\theta}_R\overline{\theta}_R}\right)^+ \mathcal{J}^{(1)}_{\overline{\theta}_R\overline{\theta}_R}\left(\mathcal{J}^{(2)}_{\overline{\theta}_R\overline{\theta}_R}\right)^+ \\ & = & CRB_{hh} + \left(\mathcal{J}^{(2)}_{\overline{\theta}_R\overline{\theta}_R}\right)^+ \mathcal{J}'_{\theta_R\theta_R}\left(\mathcal{J}^{(2)}_{\overline{\theta}_R\overline{\theta}_R}\right)^+ . \end{array} \tag{6.53}$$

$\Delta h_{I_R} \sim \mathcal{N}(0, C_{\Delta h_I \Delta h_I})$ and evolves in $\dfrac{1}{\sqrt{M_U}}$ as $M_U \to \infty$; $\Delta h_{c_R} \sim \mathcal{N}(0, C_{\Delta \bar{h}_c \Delta \bar{h}_c})$ and evolves in $\dfrac{1}{\sqrt{M_U}}$ as $M_U \to \infty$ which is due to the fact that blind GML can estimate the zeros of the channel.

At high SNR, the influence of the estimation of $\sigma_v^2$ on the estimation of the channel becomes negligible, and performance for the estimation of $h_I$ is the same as in the deterministic case.

We will not detail the computations for the case of real input symbols. The same kind of results holds also. As we have local identifiability, $\mathcal{J}^{(2)}_{\theta_R\theta_R}$ is regular and the error correlation matrices of $h_I$ and $\bar{h}_c$ are the appropriate submatrices of $\mathcal{J}_{\theta_R\theta_R} = \left(\mathcal{J}^{(2)}_{\theta_R\theta_R}\right)^{-1} \mathcal{J}^{(1)}_{\theta_R\theta_R}\left(\mathcal{J}^{(2)}_{\theta_R\theta_R}\right)^{-1}$.

### 6.4.2 Semi-Blind GML

As DML, under conditions (i) and (ii) and neglecting the observations containing known and unknown symbols at the same time, the GML criterion can be decomposed into a training sequence and blind contribution as:

$$\frac{1}{\sigma_v^2}\|\boldsymbol{Y}_{TS}-\mathcal{T}_{TS}(h)A_K\|^2+M_K\ln\sigma_v^2+\ln\det C_{Y_BY_B}+\boldsymbol{Y}_B^H C_{Y_BY_B}^{-1}\boldsymbol{Y}_B\,. \tag{6.54}$$

The IMs are:

$$\begin{cases} J_{\theta\theta}^{(2)}=J_{\theta\theta,TS}+J_{\theta\theta,B}^{(2)}=J_{\theta\theta,TS}+J_{\theta\theta,B} \\ J_{\theta\theta}^{(1)}=J_{\theta\theta}^{(2)}-J'_{\theta\theta,B}\,. \end{cases} \tag{6.55}$$

with same kind of relations for $J_{\theta\theta^*}^{(2)}$ and $J_{\theta\theta^*}^{(1)}$; the blind IMs can be written as as in (6.50) and the TS based IMs are:

$$J_{\theta\theta,TS}=\begin{bmatrix} \frac{1}{\sigma_v^2}[\mathcal{A}_{I_{TS}}\,\bar{\mathcal{A}}_{c_{TS}}]^H[\mathcal{A}_{I_{TS}}\,\bar{\mathcal{A}}_{c_{TS}}] & 0 \\ 0 & \frac{M_K}{4\sigma_v^4} \end{bmatrix} \quad \text{and} \quad J_{\theta\theta^*,TS}=\begin{bmatrix} 0 & 0 \\ 0 & \frac{M_K}{4\sigma_v^4}\,. \end{bmatrix} \tag{6.56}$$

We have:

$$C_{\bar{\theta}_R\bar{\theta}_R}=CRB+\left(J_{\bar{\theta}_R\bar{\theta}_R}^{(2)}\right)^{-1}J_{\bar{\theta}_R\bar{\theta}_R}^{(1)}\left(J_{\bar{\theta}_R\bar{\theta}_R}^{(2)}\right)^{-1}\,. \tag{6.57}$$

Both $\Delta h_I$ and $\Delta\bar{h}_c$ evolves in $\frac{1}{M_U}$ and $\frac{1}{M_K}$ as $M_U\to\infty$ and $M_K\to\infty$ respectively.

## 6.5 Numerical Evaluations

We tested three different kinds of channels (the notations are independent of the ones in Chapter 5):

- an irreducible channel $\boldsymbol{H}_{well}$,
- a channel with a nearly common zero $\boldsymbol{H}_{ill}$
- a reducible channel $\mathbf{H}(z)=\mathbf{H}_I(z)\mathrm{H}_c(z)$.

The channel coefficients were chosen randomly and can be found in appendix D. The SNR is at 10dB, the input symbols belong to a QPSK constellation and are i.i.d.. We plot the quantity: $\sqrt{\mathrm{tr}(C_{\Delta h_R\Delta h_R})}/\|h\|$.

In figure 6.2, the blind performance are plotted for $\boldsymbol{H}_{well}$ and $\boldsymbol{H}_{ill}$ for a burst length of 150 (blind performance corresponds to 0 known symbols in the curves). Performance of blind GML are computed under constraint (6.52) but also under the same constraints as blind DML to allow significant comparison between DML and GML. We see that blind GML

outperforms blind DML and especially for $\boldsymbol{H}_{ill}$. We can also note that DML does not reach the CRB.

On the same plot, the semi-blind GML and DML performance computed under constraint (6.26) and (6.27) are also plotted w.r.t. the number of known symbols, starting from 30 known symbols in the burst: we can see the gain brought by the known symbols compared to the blind mode.

The semi–blind DML and GML curves are shown in figure 6.3 and compared to the training sequence based estimation performance. The performance when all the input symbols are known is also shown as a reference. We see a certain gain of semi-blind techniques w.r.t. the training sequence technique, which cannot be significant as the number of known symbols is large. Here again GML appears better than DML. The same comments as those already done in the CRB study are basically also valid here.

In figure 6.4, the reducible case is shown. For a fixed number of 30 known symbols, we plot the performance w.r.t. the number of unknown symbols in the burst. The performance for the estimation of $H_c$ by DML will tend to be constant as the number of unknown symbols grows. GML profits from the blind information, and the slope of the curve will eventually evolve in $\frac{1}{M_U}$.

As already stated, our performance expressions, valid for an asymptotic number of known symbols, make also sense for this number is small. It is also possible to give expressions taking into account dispersed known symbols or the overlap zone (figure 6.1), as CRBs do (using $\mathcal{T}_K(h)$ and $\mathcal{T}_U(h)$): these expressions were also found to have sense.

## 6.6 Conclusions

We have derived the asymptotic performance of the blind and semi–blind deterministic ML and Gaussian ML for an asymptotic amount of known and unknown symbols but also for a high SNR. The performance have been compared to the CRBs: both DML and GML attain the CRB at high SNR only. This performance study is of particular importance as DML and GML offer lower performance bound on resp. all the deterministic and Gaussian methods.

# A Minimal Performance

We prove that $\left(\mathcal{J}_{\theta\theta}^{(2)}\right)^{+}\mathcal{J}_{\theta\theta}^{(1)}\left(\mathcal{J}_{\theta\theta}^{(2)}\right)^{+}$ corresponds to the constrained performance which gives, for a minimal number of independent constraints, the lowest value for $\mathrm{tr}\left\{C_{\Delta\theta\Delta\theta}\right\}$.

Let $\mathcal{K}_\theta = 0$ be a set of constraints on $\theta$; $\mathrm{range}\{\mathcal{V}_\theta\} = \left(\mathrm{range}\left\{\dfrac{\partial \mathcal{K}_\theta^T}{\partial\theta}\right\}\right)^{\perp}$. We assume that $\mathcal{V}_\theta^T\mathcal{J}_{\theta\theta}\mathcal{V}_\theta$ is invertible. The corresponding constrained performance is:

$$C_{\Delta\theta\Delta\theta} = \mathcal{V}_\theta\left(\mathcal{V}_\theta^T\mathcal{J}_{\theta\theta}\mathcal{V}_\theta\right)^{-1}\mathcal{V}_\theta^T\mathcal{J}_{\theta\theta}^{(1)}\mathcal{V}_\theta\left(\mathcal{V}_\theta^T\mathcal{J}_{\theta\theta}\mathcal{V}_\theta\right)^{-1}\mathcal{V}_\theta^T \tag{6.58}$$

Let the eigendecomposition of $\mathcal{J}_{\theta\theta}^{(2)} = S_1^{(2)}\Lambda_1^{(2)}S_1^{(2)^T} = S_1\Lambda_1{S_1}^T$ and $\mathcal{J}_{\theta\theta}^{(1)} = S_1^{(1)}\Lambda_1^{(1)}S_1^{(1)^T}$. $\mathcal{V}_\theta$ has components on $S_1$ and $S_2$ (the columns of $S_2$ form a basis of the null space of $\mathcal{J}_{\theta\theta}^{(1)}$ and $\mathcal{J}_{\theta\theta}^{(2)}$): $\mathcal{V}_\theta = S_1Q_1 + S_2Q_2$.

$$\mathcal{V}_\theta\left(\mathcal{V}_\theta^TS_1\Lambda_1{S_1}^T\mathcal{V}_\theta\right)^{-1}\mathcal{V}_\theta^T = \mathcal{V}_\theta\left({S_1}^T\mathcal{V}_\theta\right)^{-1}\Lambda_1^{-1}\left(\mathcal{V}_\theta^TS_1\right)^{-1}\mathcal{V}_\theta^T = \mathcal{V}_\theta Q_1^{-1}\Lambda_1^{-1}Q_1^{-H}\mathcal{V}_\theta^T \tag{6.59}$$

$$\begin{aligned} C_{\Delta\theta\Delta\theta} = {} & \left(S_1 + S_2Q_2Q_1^{-1}\right)\Lambda_1^{-1}\left(S_1 + S_2Q_2Q_1^{-1}\right)^T \\ & S_1^{(1)}\Lambda_1^{(1)}S_1^{(1)^T}\left(S_1 + S_2Q_2Q_1^{-1}\right)\Lambda_1^{-1}\left(S_1 + S_2Q_2Q_1^{-1}\right)^T \end{aligned} \tag{6.60}$$

$$C_{\Delta\theta\Delta\theta} = \left(S_1 + S_2Q_2Q_1^{-1}\right)\Lambda_1^{-1}S_1^TS_1^{(1)}\Lambda_1^{(1)}S_1^{(1)^T}S_1\Lambda_1^{-1}\left(S_1 + S_2Q_2Q_1^{-1}\right)^T \tag{6.61}$$

$$\mathrm{tr}\left\{C_{\Delta\theta\Delta\theta}\right\} = \mathrm{tr}\left\{\left(\mathcal{J}_{\theta\theta}^{(2)}\right)^{+}\mathcal{J}_{\theta\theta}^{(1)}\left(\mathcal{J}_{\theta\theta}^{(2)}\right)^{+}\right\} + \mathrm{tr}\left\{Q_2Q_1^{-1}\Lambda_1^{-1}S_1^T\mathcal{J}_{\theta\theta}^{(1)}S_1\Lambda_1^{-1}Q_1^{-T}Q_2^T\right\} \tag{6.62}$$

The second term is positive, so $\mathrm{tr}\left\{C_{\Delta\theta\Delta\theta}\right\} \geq \mathrm{tr}\left\{\left(\mathcal{J}_{\theta\theta}^{(2)}\right)^{+}\mathcal{J}_{\theta\theta}^{(1)}\left(\mathcal{J}_{\theta\theta}^{(2)}\right)^{+}\right\}$, with equality if $Q_2 = 0$, *i.e.* $\mathrm{range}\left\{\dfrac{\partial\mathcal{K}_\theta}{\partial\theta}\right\} = \mathrm{null}\left(\mathcal{J}_{\theta\theta}^{(1)}\right) = \mathrm{null}\left(\mathcal{J}_{\theta\theta}^{(2)}\right)$.

# B Consistency of Blind and Semi–Blind DML and GML

## B.1 Blind DML

As $M \to \infty$, the DML cost function tends uniformly to its expected value:

$$\boldsymbol{Y}^HP_{\mathcal{T}(h)}^{\perp}\boldsymbol{Y} \to \mathrm{tr}\left\{P_{\mathcal{T}(h)}^{\perp}\mathrm{E}\left(\boldsymbol{Y}\boldsymbol{Y}^H\right)\right\} \tag{6.63}$$

with $\mathrm{E}\left(\boldsymbol{Y}\boldsymbol{Y}^H\right) = \mathcal{T}(h^o)A^oA^{oH}\mathcal{T}^H(h^o) + \sigma_v^2I$.

$$\boldsymbol{Y}^HP_{\mathcal{T}(h)}^{\perp}\boldsymbol{Y} \to \mathrm{tr}\left\{A^{oH}\mathcal{T}^H(h^o)P_{\mathcal{T}(h)}^{\perp}\mathcal{T}(h^o)A^o\right\} + \sigma_v^2\mathrm{tr}\left\{P_{\mathcal{T}(h)}^{\perp}\right\}I \tag{6.64}$$

The term $\sigma_v^2\mathrm{tr}\left\{P_{\mathcal{T}(h)}^{\perp}\right\}I$ is constant, so the DML criterion is equivalent to:

$$\min_{\text{constrained } h}\left\{A^{oH}\mathcal{T}^H(h^o)P_{\mathcal{T}(h)}^{\perp}\mathcal{T}(h^o)A^o\right\} \tag{6.65}$$

which is the noiseless DML criterion. $h^o$ nulls the criterion and is the unique solution of the minimization problem under the blind identifiability conditions. $h^o$ verifies also both constraints (6.26) and (6.27) which give then consistent estimate.

### B.2 Semi–blind DML

Asymptotically, the semi-blind DML criterion (6.34) is equivalent to:

$$\min_h \left\{ A^{oH} \left[\mathcal{T}^H(h^o) - \mathcal{T}(h)\right]^H \left[\mathcal{T}^H(h^o) - \mathcal{T}(h)\right] A^o \quad + \quad A^{oH}\mathcal{T}^H(h^o) P^{\perp}_{\mathcal{T}(h)} \mathcal{T}(h^o) A^o \right\} \tag{6.66}$$

which is nulled by $h^o$.

### B.3 GML

As $M \to \infty$, the GML cost function tends to its expected value:

$$\mathcal{F}_{GML}(\theta) \to \ln\det\{C_{YY}\} + \mathrm{tr}\left\{C_{YY}^{-1}\mathrm{E}\left(\boldsymbol{Y} - \mathcal{A}_K h\right)\left(\boldsymbol{Y} - \mathcal{A}_K h\right)^H\right\} \tag{6.67}$$

The gradient of $\mathrm{E}\left(\mathcal{F}(\theta)\right)$ w.r.t. $\theta_i^*$ is:

$$\begin{aligned}\mathrm{tr}\left\{C_{YY}^{-1}\frac{\partial C_{YY}}{\partial \theta_i^*}\right\} - \mathrm{tr}\left\{C_{YY}^{-1}\frac{\partial C_{YY}}{\partial \theta_i^*} C_{YY}^{-1}\mathrm{E}\left(\boldsymbol{Y}-\mathcal{A}_K h\right)\left(\boldsymbol{Y}-\mathcal{A}_K h\right)^H\right\} - \frac{\partial h^H}{\partial \theta_i^*}\mathcal{A}_K^H C_{YY}^{-1} E\left(\boldsymbol{Y}-\mathcal{A}_K h\right) \\ = \mathrm{tr}\left\{C_{YY}^{-1}\frac{\partial C_{YY}}{\partial \theta_i^*}\left(I - C_{YY}^{-1}C_{YY}(h,\sigma_v^2)\right)\right\}\end{aligned} \tag{6.68}$$

with $C_{YY}(h,\sigma_v^2) = \mathrm{E}\left(\boldsymbol{Y} - \mathcal{A}_K h\right)\left(\boldsymbol{Y} - \mathcal{A}_K h\right)^H$. The gradient is nulled by $(h^o, \sigma_v^{o\,2})$, and as the Hessian of $\mathrm{E}\mathcal{F}(\theta)$, *i.e.* the FIM is positive semi–definite, this is the unique minimum of the cost function.

## C Equivalence between DML and GML

The semi–blind GML criterion is:

$$\min_{h,\sigma_v^2}\left\{\ln\det C_{YY} + \boldsymbol{Y}^H C_{YY}^{-1}\boldsymbol{Y}\right\} \tag{6.69}$$

$$C_{YY}^{-1} = \left(\sigma_a^2\mathcal{T}_U(h)\mathcal{T}_U^H(h) + \sigma_v^2 I\right)^{-1} = \frac{1}{\sigma_v^2}\left[I - \mathcal{T}_U(h)\left(\mathcal{T}_U^H(h)\mathcal{T}_U(h) + \frac{\sigma_v^2}{\sigma_a^2}I\right)^{-1}\mathcal{T}_U^H(h)\right] \tag{6.70}$$

by the matrix inversion lemma and as $\dfrac{\sigma_v^2}{\sigma_a^2} \to 0$:

$$C_{YY}^{-1} = \frac{1}{\sigma_v^2}P^{\perp}_{\mathcal{T}_U(h)} \tag{6.71}$$

So the second term in (6.69) is equivalent to $\frac{1}{\sigma_v^2}\boldsymbol{Y}^H P^{\perp}_{\mathcal{T}_U(h)}\boldsymbol{Y}$, which is the DML cost function.

$C_{YY}$ has $Mm - M - N + 1$ eigenvalues equal to $\sigma_v^2$ and $M + N - 1$ eigenvalues equal to $\sigma_a^2\lambda_i(h)$ where $\lambda_i(h)$ is an eigenvalue of $\mathcal{T}(h)\mathcal{T}^H(h)$ depending an $h$ only. The only contribution of $h$ in $\ln\det C_{YY}$ comes from the $\lambda_i(h)$'s and can be proved to be negligible w.r.t. $\boldsymbol{Y}^H C_{YY}^{-1}\boldsymbol{Y}$. The estimation of $h$ is decoupled from that of $\sigma_v^2$ and the GML cost function is equivalent to the DML one.

# D Channels used in the Simulations

- Irreducible Channel $\boldsymbol{H}_{well}$:

  The complex channel is randomly chosen and is irreducible:

$$\boldsymbol{H}_{well} = \begin{bmatrix} -0.4326-0.0280j & 0.1253-0.1584j & -1.1465+0.3366j \\ -1.6656-1.5420j & 0.2877+0.0911j & 1.1909+0.9190j \\ & & 1.1892-1.1715j \quad 0.3273+2.0161j \\ & & -0.0376-1.2130j \quad 0.1746+2.7042j \end{bmatrix}. \tag{6.72}$$

- Nearly reducible channel $\boldsymbol{H}_{ill}$:

$$\boldsymbol{H}_{ill} = \begin{bmatrix} z^{-3}\left(0.5711 - 0.3999j\right)\left(0.5 + 0.3j - z\right)\left(0.6 - 1.2j - z\right)\left(-0.2 - z\right) \\ z^{-3}\left(0.6900 + 0.8156j\right)\left(0.45 + 0.32j - z\right)\left(-0.8 - 0.2j - z\right)\left(-1.2 - z\right) \end{bmatrix} \tag{6.73}$$

- Reducible Channel:

$$\boldsymbol{H}_I = \begin{bmatrix} -0.8051+0.5913j & 0.2193+0.3803j \\ 0.5287-0.6436j & -0.9219-1.0091j \\ -0.1461-0.3745j & -0.0766+1.7513j \\ 0.2481-0.4709j & 1.7382+0.7532j \end{bmatrix}, \quad H_c = \begin{bmatrix} 1 & -0.1567+1.0565j \end{bmatrix}. \tag{6.74}$$

# E Numerical Evaluations of the Performance

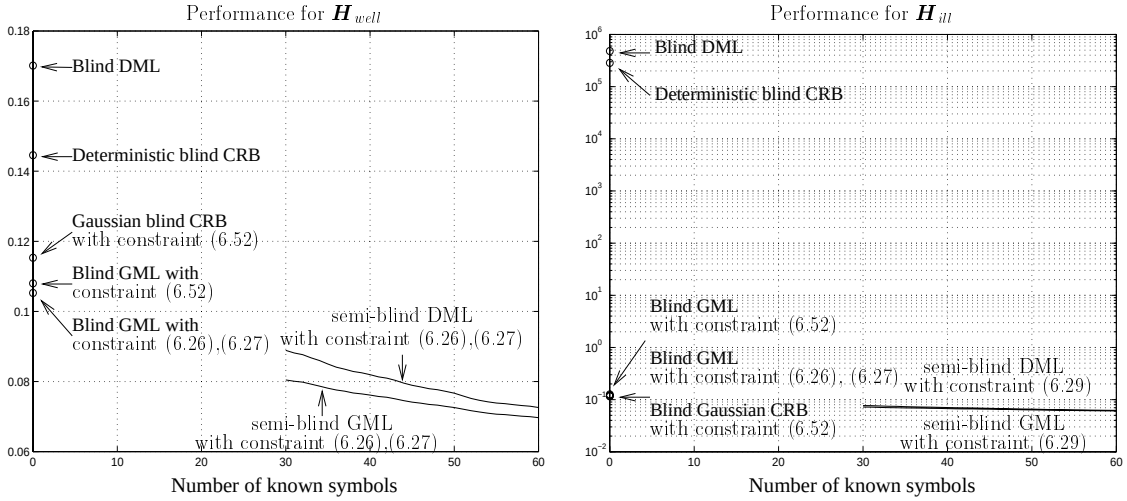



Figure 6.2: Blind (and Semi-blind) DML and GML: irreducible and nearly reducible case.

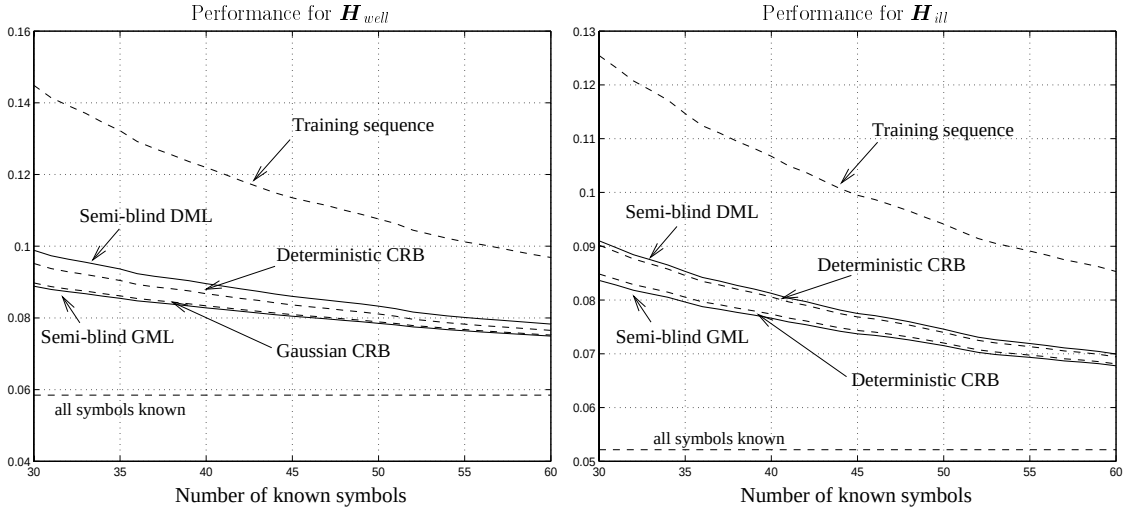



Figure 6.3: Semi-blind DML and GML: irreducible and nearly reducible case.

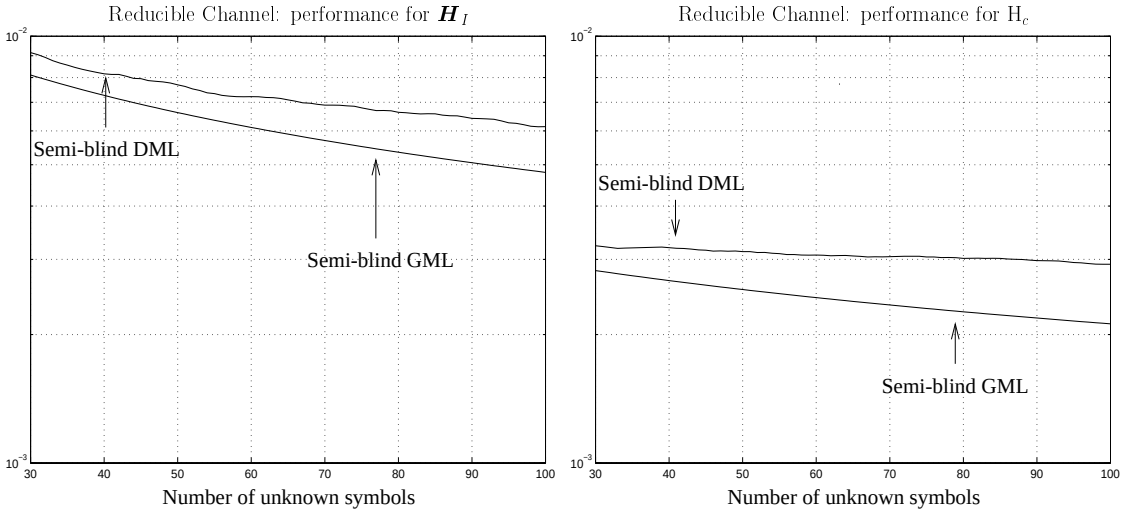



Figure 6.4: Semi-blind DML and GML: reducible case.

Chapter 7

# BLIND DETERMINISTIC MAXIMUM–LIKELIHOOD METHODS

*As we will develop semi–blind DML criterion that contain a blind part and a training sequence part, it is important that the blind part be solved by a powerful method, this is why we concentrate on finding low–computational, quasi–optimal solutions to solve DML. As stated in Chapter 1, deterministic methods have the property of giving the exact channel for a finite amount of data in the absence of noise. This is the property that is exploited in this chapter. Two DML–based algorithms are presented. The first one is a modification of the Iterative Quadratic ML (IQML) algorithm which gives biased estimates of the channel and performs poorly at low SNR due to the presence of noise. We "denoise" the IQML cost function by eliminating the noise contribution: the resulting algorithm Denoised IQML (DIQML) gives consistent estimates and outperforms IQML. Furthermore DIQML is asymptotically globally convergent and insensitive to the initialization. Its asymptotic performance does not match the ML performance, though. The second strategy, called Pseudo-Quadratic ML (PQML), is naturally denoised. The denoising is however more efficient than in DIQML: PQML gives the same asymptotic performance as DML, not DIQML though, but requires a consistent initialization which can be given by SRM or DIQML. We will furthermore compare DIQML and PQML to the alternating minimization technique w.r.t. the symbols and the channel used to solve DML. A performance study and simulations are presented.*

## 7.1 Blind Deterministic ML

Let us recall the blind DML criterion for joint estimation of $A$ and $h$:

$$\max_{A,h} f(\boldsymbol{Y}|h) \quad \Leftrightarrow \quad \min_{A,h} \|\boldsymbol{Y} - \mathcal{T}(h)A\|^2 . \tag{7.1}$$

We assume here that the blind deterministic conditions [DetB] are verified. The channel is then identifiable up to a scale factor. We impose the non-triviality constraint $\|h\| = 1$. This constraint is not sufficient to completely identify the channel as it leaves a phase ambiguity: the phase constraint (6.27) will be imposed later, in the performance study of the proposed algorithms.

We solve here the DML criterion in $h$ (with $A$ as nuisance parameter):

$$\min_{\|h\|=1} \boldsymbol{Y}^H P^{\perp}_{\mathcal{T}(h)} \boldsymbol{Y} . \tag{7.2}$$

$P^{\perp}_{\mathcal{T}(h)} = I - \mathcal{T}(h) \left(\mathcal{T}^H(h)\mathcal{T}(h)\right)^{-1} \mathcal{T}^H(h)$.

## 7.2 Linear Parameterization of the Noise Subspace

The DML criterion is highly non linear and its optimization would require costly solutions in the form (7.2). The key to a computationally attractive solution of the DML problem is a linear parameterization of the noise subspace. We consider here a linear parameterization in terms of channel coefficients (a parameterization in terms of prediction quantities was also presented in [13]). Let $\mathbf{H}^{\perp}(z)$ be such a parameterization: it verifies $\mathbf{H}^{\perp}(z)\mathbf{H}(z) = 0$ and $\mathcal{T}(h^{\perp})\mathcal{T}(h) = 0$; $\mathcal{T}(h^{\perp})$ is the convolution matrix of $\mathbf{H}^{\perp}(z)$ and the columns of $\mathcal{T}^H(h^{\perp})$ span the entire noise subspace. In the case $m = 2$ in which the multichannel has 2 subchannels, the obvious choice for $\mathbf{H}^{\perp}(z)$ is:

$$\mathbf{H}^{\perp}(z) = [-\mathrm{H}_2(z) \;\; \mathrm{H}_1(z)] . \tag{7.3}$$

For a larger number of subchannels, different choices are available [19], [20]. An example is [20]:

$$\mathbf{H}^{\perp}(z) = \begin{bmatrix} -\mathrm{H}_2(z) & \mathrm{H}_1(z) & 0 & \cdots & 0 \\ 0 & -\mathrm{H}_3(z) & \mathrm{H}_2(z) & \cdots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \mathrm{H}_m(z) & 0 & \cdots & 0 & -\mathrm{H}_1(z) \end{bmatrix} . \tag{7.4}$$

## 7.3 Subchannel Response Matching (SRM)

The Subchannel Response Matching (SRM) algorithm is based on this linear parameterization of the noise subspace. Using the commutativity of convolution and the linearity of $\mathcal{T}(h^\perp)$ in $h$, we can write $\mathcal{T}(h^\perp)\boldsymbol{Y}$ as:

$$\mathcal{T}(h^\perp)\boldsymbol{Y} = \mathcal{Y}h \tag{7.5}$$

where $\mathcal{Y}$ is a matrix filled with the elements of the observation vector $\boldsymbol{Y}$. In the noiseless case, $\boldsymbol{Y} = \mathcal{T}(h)A$ and we have $\mathcal{T}(h^\perp)\boldsymbol{Y} = \mathcal{Y}h = 0$: from this relation, the channel can be uniquely determined up to a scale factor. SRM requires the channel to be irreducible; the burst length requirements are higher than in [DetB] [19].

When there is noise, $\mathcal{Y}h \neq 0$ and the SRM criterion is solved in the least–squares sense under the constraint $\|h\| = 1$:

$$\min_{\|h\|=1} h^H \, \mathcal{Y}^H \mathcal{Y} \, h \, . \tag{7.6}$$

The resulting $h$ is the minimal eigenvector of $\mathcal{Y}^H\mathcal{Y}$. Different choices for the linear parameterization of the noise subspace give different channel estimates: certain parameterizations give indeed biased estimates [20]. The parameterization described in (7.4) gives consistent estimates.

SRM appears as a non-weighted version of the Iterative Quadratic ML (IQML) algorithm, and is used in [19] to initialize IQML. We will use it to initialize our algorithms also.

## 7.4 Iterative Quadratic ML (IQML)

Since $P^\perp_{\mathcal{T}(h)} = P_{\mathcal{T}^H(h^\perp)}$, (7.2) can be written as:

$$\min_{\|h\|=1} \boldsymbol{Y}^H \mathcal{T}^H(h^\perp) \left(\mathcal{T}(h^\perp)\mathcal{T}^H(h^\perp)\right)^+ \mathcal{T}(h^\perp)\boldsymbol{Y} \, . \tag{7.7}$$

$\mathcal{T}(h^\perp)\mathcal{T}^H(h^\perp)$ is singular for $m > 2$, which is why the Moore-Penrose pseudo–inverse needs to be introduced.

The Iterative Quadratic ML (IQML) algorithm solves (7.7) iteratively in such a way that at each step a quadratic problem appears. Let $\mathcal{R}(h) \stackrel{\triangle}{=} \mathcal{T}(h^\perp)\mathcal{T}^H(h^\perp)$, then (7.7) becomes:

$$\min_{\|h\|=1} \boldsymbol{Y}^H \mathcal{T}^H(h^\perp)\mathcal{R}^+(h)\mathcal{T}(h^\perp)\boldsymbol{Y} \, . \tag{7.8}$$

At iteration (i) of IQML, the denominator $\mathcal{R}(h)$ is computed based on the estimate from the previous iteration $h^{(i-1)}$ and is considered as constant for the current iteration. Hence, as $\mathcal{T}(h^\perp)$ is linear in $h$, the criterion (7.7) becomes quadratic. Denoting the constant denominator $\mathcal{R}(h) = \mathcal{R}$, the IQML criterion can be rewritten as:

$$\min_{\|h\|=1} h^H \mathcal{Y}^H \mathcal{R}^+ \mathcal{Y} \, h \, . \tag{7.9}$$

Under the constraint $\|h\| = 1$, $h$ is estimated as the minimal eigenvector of the matrix $\mathcal{Y}^H \mathcal{R}^+ \mathcal{Y}$.

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

$$\min_{\|h\|=1} \boldsymbol{X}^H \mathcal{T}^H(h^\perp) \mathcal{R}^+ \mathcal{T}(h^\perp) \boldsymbol{X} \tag{7.10}$$

where $\boldsymbol{X} = \mathcal{T}(h^o)A$ is the noise-free received signal. As $\mathcal{T}(h^{o\perp})\boldsymbol{X} = \mathcal{X}h^o = 0$, $h^o$ nulls exactly the criterion regardless of the initialization. At high SNR, a first iteration of IQML gives a consistent estimate of the channel whatever the initialization of $\mathcal{R}(h)$ (provided that $\text{Null}(\mathcal{R}^+) \cap \text{Range}(\mathcal{X}) = 0$, which is guaranteed in general). And it can be proven [19] that a second iteration gives the exact ML estimate.

At low SNR however, this method is biased. Indeed, consider the asymptotic situation in which the number of data $M$ grows to infinity. By the law of large numbers, the IQML criterion becomes equivalent to its expected value which is:

$$\begin{aligned} &\min_{\|h\|=1} \text{tr}\{\mathcal{T}^H(h^\perp) \mathcal{R}^+ \mathcal{T}(h^\perp) E(\boldsymbol{Y}\boldsymbol{Y}^H)\} = \\ &\min_{\|h\|=1} \left\{ \text{tr}\{\mathcal{T}^H(h^\perp) \mathcal{R}^+ \mathcal{T}(h^\perp) \boldsymbol{X}\boldsymbol{X}^H\} + \sigma_v^2 \text{tr}\{\mathcal{T}^H(h^\perp) \mathcal{R}^+ \mathcal{T}(h^\perp)\} \right\} \end{aligned} \tag{7.11}$$

since $E(\boldsymbol{Y}\boldsymbol{Y}^H) = \boldsymbol{X}\boldsymbol{X}^H + \sigma_v^2 I$. $h^o$ nulls exactly the first term, but is not in general a minimal eigenvector of the second term and hence of the sum. Then, due to the presence of noise, $h^o$ is not asymptotically a stationary point of the algorithm and IQML performs poorly even if initialized by a consistent channel estimate.

We propose here a method to "denoise" the IQML criterion: this denoised criterion minimized in the IQML way will compensate the IQML bias and gives a consistent channel estimate.

## 7.5 Denoised Iterative Quadratic ML (DIQML)

### 7.5.1 Asymptotic Amount of Data

The asymptotic noise contribution to the DML criterion is $\sigma_v^2 \,\text{tr}\left\{P_{\mathcal{T}^H(h^\perp)}\right\}$. The denoising strategy consists simply in removing this asymptotic noise term, or more exactly an estimate of it $\widehat{\sigma_v^2} \text{tr} P_{\mathcal{T}^H(h^\perp)}$ (where $\widehat{\sigma_v^2}$ will be a consistent estimate of the noise variance), from the DML criterion which becomes:

$$\begin{aligned} &\min_{\|h\|=1} \text{tr}\left\{P_{\mathcal{T}^H(h^\perp)} \left(YY^H - \widehat{\sigma_v^2} I\right)\right\} \quad \Leftrightarrow \\ &\min_{\|h\|=1} \left\{ h^H \mathcal{Y}^H \mathcal{R}^+(h) \mathcal{Y} h - \widehat{\sigma_v^2} \text{tr}\{\mathcal{T}(h^\perp) \mathcal{R}^+(h) \mathcal{T}^H(h^\perp)\} \right\} . \end{aligned} \tag{7.12}$$

Note that this operation does not change the solution of the DML criterion as $\widehat{\sigma_v^2}\,\mathrm{tr}\{P_{\mathcal{T}^H(h^\perp)}\} = \widehat{\sigma_v^2}(M(m-1)-N+1)I$ is constant w.r.t. $h$.

(7.12) is solved in the IQML way: considering $\mathcal{R}(h) = \mathcal{R}$ as constant, the optimization problem is quadratic:

$$\min_{\|h\|=1} h^H \left\{ \mathcal{Y}^H \mathcal{R}^+ \mathcal{Y} - \widehat{\sigma_v^2}\, \mathcal{D} \right\} h \tag{7.13}$$

where $h^H \mathcal{D} h = \mathrm{tr}\{\mathcal{T}^H(h^\perp)\mathcal{R}^+\mathcal{T}(h^\perp)\}$.

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

$$\min_{\|h\|=1} \boldsymbol{X}^H \mathcal{T}^H(h^\perp)\mathcal{R}^+\mathcal{T}(h^\perp)\boldsymbol{X}\,. \tag{7.14}$$

We find again the main features of the high SNR IQML algorithm:

- The first iteration gives a consistent estimate of the channel.
- The second iteration gives asymptotically the global minimizer of DIQML. Unlike in the high SNR case, this global minimizer is not the ML minimizer though, as will be seen in appendix A.
- This behavior holds whatever the initialization

At high SNR global convergence is also guaranteed as it is for the original IQML algorithm but this time DIQML gives the ML solution.

### 7.5.2 Finite Amount of Data

The choice of $\widehat{\sigma_v^2}$ turns out to be crucial. In practice, with large but finite amount of data $M$, and the true noise variance value, the central matrix in (7.13) is indefinite, and the minimization problem is no longer well posed. The solution in this case would be the eigenvector corresponding to the smallest eigenvalue, which is negative. Simulations have shown that performance does not improve upon IQML in this case. The central matrix $\mathcal{Q} = \mathcal{Y}^H\mathcal{R}^+\mathcal{Y} - \lambda\mathcal{D}$ should be constrained to be positive semi-definite.

As a consistent estimate of $\sigma_v^2$, we choose here a certain $\lambda$ that renders $\mathcal{Q} = \mathcal{Y}^H\mathcal{R}^+\mathcal{Y} - \lambda\mathcal{D}$ exactly positive semi-definite with one singularity. The DIQML criterion becomes:

$$\min_{\|h\|=1,\lambda} h^H \left\{ \mathcal{Y}^H \mathcal{R}^+ \mathcal{Y} - \lambda \mathcal{D} \right\} h \tag{7.15}$$

with constraint that $\mathcal{Q}$ be positive semi-definite. $\lambda$ the generalized minimal eigenvalue of $\mathcal{Y}^H\mathcal{R}^+\mathcal{Y}$ and $\mathcal{D}$ and $h$ is the associated generalized eigenvector. Asymptotically, $\lambda \to \sigma_v^2$, and the criterion becomes equivalent to (7.14), and asymptotic global convergence applies for $h$ and for $\sigma_v^2$, with the same properties as mentioned earlier.

Other attempts have been undertaken to denoise the IQML strategy. Kristensson [68] applies the same strategy in the DOA (Direction Of Arrival) context: as estimate of the noise variance, he chooses the one which in the context of blind channel estimation would correspond to the minimum value of the SRM criterion. It can indeed be verified that asymptotically:

$$\begin{array}{rcll} \hat{h}_{SRM}^H \left(\mathcal{Y}^H \mathcal{Y}\right) \hat{h}_{SRM} & = & \sigma_v^2 (M-N+1) & \text{for } m=2 \\ \hat{h}_{SRM}^H \left(\mathcal{Y}^H \mathcal{Y}\right) \hat{h}_{SRM} & = & 2\sigma_v^2 (M-N+1) & \text{for } m>2 \end{array} \tag{7.16}$$

with $\|\hat{h}_{SRM}\|^2 = 1$. For a finite amount of data, the noise variance estimate given by SRM underestimates the true $\sigma_v^2$ on the average: indeed, as $\hat{h}_{SRM}$ minimizes the SRM criterion, $\hat{h}_{SRM}^H \left(\mathcal{Y}^H \mathcal{Y}\right) \hat{h}_{SRM} \leq h^{o\,H} \left(\mathcal{Y}^H \mathcal{Y}\right) h^o / \|h^o\|^2$, taking the expected value on both sides, we get $\mathrm{E}\widehat{\sigma_{v\,SRM}^2} \leq \sigma_v^2$.

So for the realizations in which $\widehat{\sigma_v^2}$ is smaller than $\sigma_v^2$, $\mathcal{Q}$ can be positive (but not always) and in that case, the DIQML principle works. However, there will be realizations in which $\widehat{\sigma_{v\,SRM}^2}$ overestimates $\sigma_v^2$: in that case $\mathcal{Q}$ is not positive and DIQML does not improve upon IQML. After the submission of our generalized eigenvalue approach in [69], we came accross [70] in which IQML is presented with different constraints, one of which corresponds to our generalized eigenvalue strategy.

## 7.6 Pseudo-Quadratic ML (PQML)

The principle of PQML introduced in the context of sinusoids in noise estimation [71] and then applied to DML in [72]. The gradient of the DML cost function may be arranged as $\mathcal{P}(h)h$, where $\mathcal{P}(h)$ is ideally a positive semi-definite matrix. The ML solution satisfies

$$\mathcal{P}(h)h = 0, \tag{7.17}$$

which is solved under the constraint $\|h\| = 1$ by the PQML strategy as follows. At iteration (i), $\mathcal{P}(h)$ is considered constant, computed thanks to the result of the previous iteration/initialization $h^{(i-1)}$; as $\mathcal{P}(h)$ is positive semi-definite, the problem becomes quadratic and $h$ is the minimal eigenvector of $\mathcal{P}(h)$. This solution is used to reevaluate $\mathcal{P}(h)$ and further iterations may be performed.

The difficulty consists in defining the right $\mathcal{P}(h)$, and especially with the positive semi-definiteness constraint. Denoting $\mathcal{T}\left(\dfrac{\partial h^\perp}{\partial h_i}\right) = \Delta\mathcal{T}_i^\perp$, the gradient of the DML cost function consists of two terms:

$$\begin{array}{c} \left(\mathcal{P}(h)\underline{\boldsymbol{h}}\right)(i) = \\ \boldsymbol{Y}^H \Delta\mathcal{T}_i^{\perp\,H} \mathcal{R}^+(h) \mathcal{T}(\underline{\boldsymbol{h}}^\perp) \boldsymbol{Y} \;\; - \;\; \boldsymbol{Y}^H \mathcal{T}^H(h^\perp) \mathcal{R}^+(h) \left[\mathcal{T}(\underline{\boldsymbol{h}}^\perp) \Delta\mathcal{T}_i^{\perp\,H}\right] \mathcal{R}^+(h) \mathcal{T}(h^\perp) \boldsymbol{Y}\,. \end{array} \tag{7.18}$$

Here, we consider that $h$ is complex and complex derivation w.r.t. $h^*$ is applied; for a real $h$, the results are similar. We assume that the pseudo–inverse is computed by regularization, as recommended in [73]. In that case, for the purpose of taking derivations, we just need to consider the derivative of a regular inverse.

In each iteration, $\mathcal{P}(h)$ will be considered as constant: $\underline{\boldsymbol{h}}$ designates the instances of $h$ that we consider as variable (on which minimization will be done) and $h$ designates the instances of $h$ that are considered as part of the constant $\mathcal{P}(h)$. The first term of $\mathcal{P}(h)\underline{\boldsymbol{h}}$ is $\mathcal{Y}^H\mathcal{R}^+(h)\mathcal{Y}\underline{\boldsymbol{h}}$, and the second term is $\mathcal{B}^H(h)\mathcal{B}(h)\underline{\boldsymbol{h}}$, with $\boldsymbol{Y}^H\mathcal{T}^H(h^\perp)\mathcal{R}^+(h)\mathcal{T}(\underline{\boldsymbol{h}}^\perp) = \underline{\boldsymbol{h}}^T\mathcal{B}^T(h)$. Then, $\mathcal{P}(h)$ has the following form:

$$\mathcal{P}(h) = \mathcal{Y}^H\mathcal{R}^+(h)\mathcal{Y} - \mathcal{B}^H(h)\mathcal{B}(h)\,. \tag{7.19}$$

As $M \to \infty$, the second term of $\mathcal{P}(h)$ tends to its expected value by the law of large numbers. In appendix A, we prove that $\mathrm{E}\left(\mathcal{B}^H(h)\mathcal{B}(h)\right)$ has a noise component equal to $\sigma_v^2 D$, the asymptotic noise component of the IQML Hessian, but it also has a non–zero signal component for $h \neq h^o$. This prevents PQML from being asymptotically insensitive to the initialization, unlike DIQML. However, when $\mathcal{P}(h)$ is evaluated at a consistent estimate of $h$, the previously mentioned signal component becomes negligible and the global convergence applies here also. PQML gives furthermore better performance than DIQML, and in fact, offers asymptotically the same performance as ML.

The matrix $\mathcal{P}(h)$ is indefinite for a finite data length $M$, and applying the PQML strategy directly will not work. In [72], $h$ is chosen as the eigenvector corresponding to the smallest absolute eigenvalue of $\mathcal{P}(h)$; it gives poor performance except at very high SNR.

PQML is closely related to DIQML as the first term of (7.13) and (7.19) are the same and $\mathrm{E}(\mathcal{B}^H(h^o)\mathcal{B}(h^o)) = \sigma_v^2\mathcal{D}(h^o)$. By analogy with DIQML for which $\mathcal{Q}(h)$ was also indefinite for finite $M$, we introduce an arbitrary $\lambda$ such that $\mathcal{Y}^H\mathcal{R}^+\mathcal{Y} - \lambda\mathcal{B}^H\mathcal{B}$ is exactly positive semi-definite. PQML then becomes the following minimization problem:

$$\min_{\|h\|=1,\lambda} h^H \left\{\mathcal{Y}^H\mathcal{R}^+\mathcal{Y} - \lambda\mathcal{B}^H\mathcal{B}\right\} h \tag{7.20}$$

with a semi-definite positivity constraint on the central matrix. The minimizing $h$ is the minimal generalized eigenvector of $\mathcal{Y}^H\mathcal{R}^+\mathcal{Y}$ and $\mathcal{B}^H\mathcal{B}$, and $\lambda$ the minimal generalized eigenvalue. Asymptotically for a consistent initialization, there is global convergence for $h$, as described previously, and as well as $\lambda$ ($\to 1$).

The identifiability conditions for both DIQML and PQML are the same as for SRM.

## 7.7 Asymptotic Performance

In appendix A, we compute the performance of DIQML and PQML under our usual deterministic constraints for $h$: $h^H h = h^{o\,H} h^o$ and $h_{S_2}^H h = 0$. We prove the following results:

- PQML has better performance than DIQML.

- PQML has the same asymptotic performance as ML. The PQML global minimizer is different however from the ML global minimizer.

At high SNR, DIQML, PQML and DIQML exhibit the same performance.

## 7.8 Alternating Quadratic ML (AQML)

In addition to comparing the performance of DIQML and PQML to the optimal ML performance, we will compare them to an algorithm we call Alternating Quadratic ML (AQML).

### 7.8.1 Alternating Minimization

AQML proceeds by alternating minimizations w.r.t. $A$ and w.r.t $h$ of the DML criterion:

$$\min_{h,A} \|\boldsymbol{Y} - \mathcal{T}(h)A\|^2 \tag{7.21}$$

(1) Initialization: $\hat{h}^{(0)}$.

(2) Iteration $(i+1)$:

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

$$\widehat{A}^{(i+1)} = \left(\mathcal{T}^H(\hat{h}^{(i)})\mathcal{T}(\hat{h}^{(i)})\right)^{-1}\mathcal{T}^H(\hat{h}^{(i)})\boldsymbol{Y} \tag{7.22}$$

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

$$\hat{h}^{(i+1)} = \left(\hat{\mathcal{A}}^{(i+1)\,H}\hat{\mathcal{A}}^{(i+1)}\right)^{-1}\hat{\mathcal{A}}^{(i+1)\,H}\boldsymbol{Y} \tag{7.23}$$

(3) Repeat (2) until $(\widehat{A}^{(i+1)}, \hat{h}^{(i+1)}) \sim (\widehat{A}^{(i)}, \hat{h}^{(i)})$.

At any iteration $(i+1)$, we assume that the algorithm gives a unique solution: $\mathcal{T}(\hat{h}^{(i)})$ has full-column rank (*i.e.* $\hat{\boldsymbol{H}}(z)$ is irreducible), as well as $\hat{\mathcal{A}}^{(i+1)}$, otherwise as suggested in [74], we take the minimum-norm solution (*i.e.* the regular inverse is replaced by the pseudo-inverse). That case is unlikely though if $h^o$ and $\mathcal{A}^o$ are well-conditioned.

### 7.8.2 Convergence Study

In appendix B, a convergence study of AQML is proposed. Any AQML iteration is shown to strictly decrease the cost function until a fixed point is attained. And the global minimizer of DML (not necessarily its local minimizers though) is a fixed point of AQML. Convergence is reached after an infinite number of iterations.

An interesting consequence of this algorithm is that even with a short data burst and appropriate initialization, the algorithm will converge to the global ML minimizer, which is not the case for PQML. However the great disadvantage of AQML is its slow convergence which prohibits its use.

### 7.8.3 Asymptotic Behavior of AQML

The ML criterion for $(A, h)$ is asymptotically (in the number of data) equivalent to :

$$\min_{h,A} \left\{ \|\boldsymbol{X} - \mathcal{T}(h)A\|^2 + \|\boldsymbol{V}\|^2 \right\} \Leftrightarrow \min_{h,A} \|\boldsymbol{X} - \mathcal{T}(h)A\|^2 . \tag{7.24}$$

So the asymptotic behavior of AQML is equivalent to its behavior in the noise–free case. In [29], it has been shown that in the noise–free case, the true quantities are the only fixed point of AQML for which $h$ is irreducible and $A$ is sufficiently exciting. Hence asymptotically, AQML is essentially globally convergent.

## 7.9 Simulation Results

We consider an irreducible channel $\boldsymbol{H}$ of length $N = 4$ with $m = 2$ subchannels, complex and randomly generated. The input symbols are drawn from an i.i.d. QPSK constellation. The initialization of the DIQML/PQML algorithms is done by SRM.

In figure 7.1, we plot the Normalized MSE (NMSE): NMSE= $\|h - \hat{h}\|^2/\|h\|^2$, the DML cost function, the generalized eigenvalue for PQML and the ratio between the generalized eigenvalue and $\sigma_v^2$ for DIQML, averaged over 500 Monte-Carlo runs of the noise. The burst length is $M = 100$ and SNR=10dB. We notice that the averaged minimal generalized eigenvalue of DIQML tends to the noise variance $\sigma_v^2$ and that of PQML to 1, while remaining smaller than these values in both cases. After 1 or 2 iterations, DIQML and PQML reach their steady state.

In Figure 7.1 and 7.2, the NMSEs are shown for burst length 100 and 200 and SNR values of 10dB and 20 dB. They are compared to the theoretical performance of DIQML and PQML, the last one being also the DML performance. The CRB is also shown. An improvement w.r.t. the SRM initialization can be observed for both algorithms, especially for PQML which outperforms DIQML. Performance can be seen to be closed to the theoretical performance. We do not plot IQML results which are much worse than SRM, DIQML and PQML.

At last, in figure 7.3, we compare PQML and DIQML to AQML, where the slow convergence of AQML can be noticed.

## 7.10 Conclusion

We have presented two methods, DIQML and PQML, to solve the DML problem. DIQML is asymptotically globally convergent but does not reach the ML performance. PQML reaches asymptotically the ML performance with a consistent initialization, which can be given by SRM or DIQML. Semi–blind extensions of PQML is presented in the next chapter and will be shown to give better performance than their blind counterparts. Furthermore, a (blind and semi–blind) extension of PQML has been proposed in a multiuser context (Spatial Division Multiple Access (SDMA)) [23].

# A Performance Study of DIQML and PQML

## A.1 Asymptotic behavior of PQML

### For an Inconsistent Initialization

The element $(i,j)$ of the Hessian $\mathcal{P}(h)$ of the PQML cost function can be written as:

$$\mathcal{P}(h)(i,j) = \underbrace{\boldsymbol{Y}^H \Delta\mathcal{T}_i^{\perp\,H} \mathcal{R}^+(h) \Delta\mathcal{T}_j^{\perp} \boldsymbol{Y}}_{\mathcal{P}_1(h)(i,j)} - \underbrace{\boldsymbol{Y}^H \mathcal{T}^H(h^\perp) \mathcal{R}^+(h) [\Delta\mathcal{T}_j^{\perp} \Delta\mathcal{T}_i^{\perp\,H}] \mathcal{R}^+(h) \mathcal{T}(h^\perp) \boldsymbol{Y}}_{\mathcal{P}_2(h)(i,j)}. \tag{7.25}$$

$\mathrm{E}\,\boldsymbol{Y}\boldsymbol{Y}^H = \sigma_a^2 \mathcal{T}(h^o)\mathcal{T}^H(h^o) + \sigma_v^2 I$ implies:

$$\mathrm{E}\,\mathcal{P}_1(h)(i,j) = \sigma_a^2 \mathrm{tr}\left\{\Delta\mathcal{T}_i^{\perp\,H} \mathcal{R}^+(h) \Delta\mathcal{T}_j^{\perp} \mathcal{T}(h^o)\mathcal{T}^H(h^o)\right\} + \sigma_v^2 \mathrm{tr}\left\{\Delta\mathcal{T}_i^{\perp\,H} \mathcal{R}^+(h) \Delta\mathcal{T}_j^{\perp}\right\} \tag{7.26}$$

$$\begin{aligned} \mathrm{E}\,\mathcal{P}_2(h)(i,j) = {} & \sigma_a^2 \mathrm{tr}\left\{\mathcal{T}^H(h^\perp)\mathcal{R}^+(h)\left[\Delta\mathcal{T}_j^{\perp}\Delta\mathcal{T}_i^{\perp\,H}\right]\mathcal{R}^+(h)\mathcal{T}(h^\perp)\mathcal{T}(h^o)\mathcal{T}^H(h^o)\right\} \\ & + \sigma_v^2 \mathrm{tr}\left\{\Delta\mathcal{T}_i^{\perp\,H}\mathcal{R}^+(h)\Delta\mathcal{T}_j^{\perp}\right\} \end{aligned} \tag{7.27}$$

Then for $h \neq h^o$, $\mathrm{E}\mathcal{P}(h)(i,j) \neq \sigma_a^2 \mathrm{tr}\left\{\Delta\mathcal{T}_i^{\perp\,H}\mathcal{R}^+(h)\Delta\mathcal{T}_j^{\perp}\mathcal{T}(h^o)\mathcal{T}^H(h^o)\right\}$ (*i.e.* the noise–free IQML central matrix) because of the signal contribution in $\mathrm{E}\mathcal{P}_2(h)(i,j)$.

### For a Consistent Initialization

Assume $h$ is a consistent estimate of $h^o$, *i.e.* $h = h^o + \Delta h$. We denote by $\mathcal{P}_{21}(h)$ the first term of $\mathcal{P}_2(h)$:

$$\begin{aligned} \mathrm{E}\mathcal{P}_{21}(h^o + \Delta h) &= \sigma_a^2 \mathrm{tr}\left\{\mathcal{T}^H(h^o)\mathcal{T}^H(\Delta h^\perp)\mathcal{R}^+(h^o)\Delta\mathcal{T}_j^{\perp}\Delta\mathcal{T}_i^{\perp\,H}\mathcal{R}^+(h^o)\mathcal{T}(\Delta h^\perp)\mathcal{T}(h^o)\right\} \\ &= O(\|\Delta h\|^2) \end{aligned} \tag{7.28}$$

which is of order 2 in $\Delta h$ and the other terms in $\mathrm{E}\mathcal{P}(h)$ can be verified to be of order 1. So $\mathcal{P}_{21}(h)$ is negligible and, asymptotically, the role of $\mathcal{P}_2$ is to remove the noise contribution in $\mathcal{P}_1$.

## A.2 Generalized Eigenvector Problem: $Av = \lambda Bv$

Let A and B be n-by-n matrices. Here are some useful properties of the generalized eigenvectors and eigenvalues of $A$ and $B$.

- There are n eigenvalues if and only if rank(B)=n.
- If A and B are symmetric:
  - The $\lambda_i$'s are real.

– Let $Av_i = \lambda_i Bv_i$ and $Av_k = \lambda_k Bv_k$, $\lambda_i \neq \lambda_k$.

$$v_i^H A v_k = v_i^H B v_k = 0 \tag{7.29}$$

– If rank(B)=n, $\{v_i\}_{i=1,\cdots,n}$ form a basis.

– If A and B are symmetric and positive semi–definite: $\lambda_i \geq 0$.

## A.3 Performance of DIQML and PQML

We consider the following general blind channel estimation problem:

$$\min_{h,\lambda} h^H \left\{ \widehat{A}(\boldsymbol{Y}, h^c) - \lambda \widehat{B}(\boldsymbol{Y}, h^c) \right\} h \tag{7.30}$$

where $h^c$ is a consistent estimate of $h$, and compute its asymptotic performance under our usual constraints: $h^H h = h^{oH} h^o$ and $h_{S_2}^{oT} h_R = h_{S_2}^{oT} h_R^o = 0$

As the data length tends to $\infty$, $\widehat{A}(\boldsymbol{Y}, h^c) \overset{M\to\infty}{\longrightarrow} A^o(h^c)$ and $\widehat{B}(\boldsymbol{Y}, h^c) \overset{M\to\infty}{\longrightarrow} B^o(h^c)$. $\widehat{A}(\boldsymbol{Y}, h^c) = \frac{1}{M} \mathcal{Y}^H \mathcal{R}^+(h^c) \mathcal{Y}$ for DIQML and PQML, and $\widehat{B}(\boldsymbol{Y}, h^c) = \frac{1}{M} \mathcal{D}(h^c)$ for DIQML and $\widehat{B}(\boldsymbol{Y}, \hat{h}) = \frac{1}{M} \mathcal{B}^H(h^c) \mathcal{B}(h^c)$ for DIQML.

It can be shown that the channel estimation performance given by (7.30) is the same when one uses $\widehat{A}(\boldsymbol{Y}, h^o) = \widehat{A}(\boldsymbol{Y})$ and $\widehat{B}(\boldsymbol{Y}, h^o) = \widehat{B}(\boldsymbol{Y})$ instead of $\widehat{A}(\boldsymbol{Y}, h^c)$ and $\widehat{B}(\boldsymbol{Y}, h^c)$. Asymptotically, we have:

$$\begin{cases} \widehat{A}(\boldsymbol{Y}) &= A^o + \widetilde{A}(\boldsymbol{Y}) \\ \widehat{B}(\boldsymbol{Y}) &= B^o + \widetilde{B}(\boldsymbol{Y}) \end{cases} \tag{7.31}$$

### Asymptotic Expressions of $\Delta\lambda$

The solution of (7.30) for $\lambda$ and $h$ is the minimal generalized eigenvalue and corresponding eigenvector of $\widehat{A}(\boldsymbol{Y})$ and $\widehat{B}(\boldsymbol{Y})$.

$$\widehat{A}(\boldsymbol{Y})\hat{h} - \hat{\lambda}\widehat{B}(\boldsymbol{Y})h = 0 \;\; \Rightarrow \;\; \hat{\lambda} = \frac{\hat{h}^H \widehat{A}(\boldsymbol{Y})\, \hat{h}}{\hat{h}^H \widehat{B}(\boldsymbol{Y})\, \hat{h}}. \tag{7.32}$$

We denote $\hat{h} = h^o + \Delta h$ and $\hat{\lambda} = \lambda^o + \Delta\lambda$. Then keeping only the first order terms, we get:

$$\Delta\lambda = \frac{h^{oH} \left[\widehat{A}(\boldsymbol{Y}) - \lambda^o \widehat{B}(\boldsymbol{Y})\right] h^o}{h^{oH} B^o h^o} \tag{7.33}$$

### Asymptotic Expressions of $\Delta h$ and $C_{\Delta h \Delta h}$

After substitution of the solution for $\lambda$, the estimation problem for $h$ becomes:

$$\min_h \;\; \left\{ h^H \left\{ \widehat{A}(\boldsymbol{Y}) - \hat{\lambda}(\boldsymbol{Y}) \widehat{B}(\boldsymbol{Y}) \right\} h = \mathcal{F}(h) \right\} \tag{7.34}$$

Let $\mathcal{V}^o$ be a matrix the columns of which form a basis of the orthogonal complement of $h^o$. Proceeding as in [75] for the asymptotic analysis of the DML performance, keeping only the first order terms in $\Delta h$, we get the following asymptotic expressions for $\Delta h$ and $C_{\Delta h \Delta h}$ with constraints:

$$\left\{ \begin{array}{rcl} \Delta h & = & \mathcal{V}^o \left( \mathcal{V}^{oH} J_{hh}^{(2)} \mathcal{V}^o \right)^{-1} \mathcal{V}^{oH} \dfrac{\partial \mathcal{F}(h)}{\partial h^*} \\ C_{\Delta h} & = & \mathcal{V}^o \left( \mathcal{V}^{oH} J_{hh}^{(2)} \mathcal{V}^o \right)^{-1} \mathcal{V}^{oH} J_{hh}^{(1)} \mathcal{V}^o \left( \mathcal{V}^{oH} J_{hh}^{(2)} \mathcal{V}^o \right)^{-1} \mathcal{V}^{oH} \end{array} \right. \tag{7.35}$$

where:

$$\left\{ \begin{array}{rcl} J_{hh}^{(1)} & = & \mathrm{E} \left( \dfrac{\partial \mathcal{F}(h)}{\partial h^*} \right) \left( \dfrac{\partial \mathcal{F}(h)}{\partial h^*} \right)^H \\ J_{hh}^{(2)} & = & \mathrm{E} \dfrac{\partial}{\partial h^*} \left( \dfrac{\partial \mathcal{F}(h)}{\partial h^*} \right)^H \end{array} \right. \tag{7.36}$$

and $\mathcal{F}(h)$ is the cost function in (7.34).

$$J_{hh}^{(2)} = \mathrm{E} \left( \widehat{A}(\boldsymbol{Y}) - \hat{\lambda}(\boldsymbol{Y}) \widehat{B}(\boldsymbol{Y}) \right) = A^o - \lambda^o B^o \tag{7.37}$$

As $A^o - \lambda^o B^o$ admits $h^o$ has (unique) singular eigenvector,

$$\mathcal{V}^o \left( \mathcal{V}^{oH} J_{hh}^{(2)} \mathcal{V}^o \right)^{-1} \mathcal{V}^{oH} = (A^o - \lambda^o B^o)^+ \ . \tag{7.38}$$

Hence:

$$\Delta h = (A^o - \lambda^o B^o)^+ \left( \widehat{A}(\boldsymbol{Y}) - \hat{\lambda} \widehat{B}(\boldsymbol{Y}) \right) h^o \ . \tag{7.39}$$

### Application to DIQML and PQML

Specializing (7.39) to DIQML and PQML, we get asymptotically:

$$\begin{array}{rcl} \Delta h & = & (\mathcal{X}^H \mathcal{R}^+ \mathcal{X})^+ \left( \widehat{A}(\boldsymbol{Y}) - \hat{\lambda} \widehat{B}(\boldsymbol{Y}) \right) h^o \\ & = & (\mathcal{X}^H \mathcal{R}^+ \mathcal{X})^+ \left( \widehat{A}(\boldsymbol{Y}) - \lambda^o \widehat{B}(\boldsymbol{Y}) - \Delta\lambda B^o(\boldsymbol{Y}) \right) h^o \end{array} \tag{7.40}$$

For ML, the same kind of analysis gives [75]:

$$\Delta h_{ML} = (\mathcal{X}^H \mathcal{R}^+ \mathcal{X})^+ \left( \widehat{A}(\boldsymbol{Y}) - \lambda^o \widehat{B}(\boldsymbol{Y}) \right) h^o \tag{7.41}$$

where $\widehat{A}(\boldsymbol{Y})$ and $\widehat{B}(\boldsymbol{Y})$ are the same as in the PQML case. And so the estimate given by DIQML and PQML is different from the ML estimate.

We introduce:

$$\left\{ \begin{array}{rcl} \mathcal{W}_{\lambda^o} & = & \mathrm{E} \left\{ \left( \hat{A}(\boldsymbol{Y}) - \lambda^o \hat{B}(\boldsymbol{Y}) \right) h^o h^{oH} \left( \hat{A}(\boldsymbol{Y}) - \lambda^o \hat{B}(\boldsymbol{Y}) \right)^H \right\} \\ \mathcal{W}_{\hat{\lambda}} & = & \mathrm{E} \left\{ \left( \hat{A}(\boldsymbol{Y}) - \hat{\lambda} \hat{B}(\boldsymbol{Y}) \right) h^o h^{oH} \left( \hat{A}(\boldsymbol{Y}) - \hat{\lambda} \hat{B}(\boldsymbol{Y}) \right)^H \right\} \end{array} \right. \tag{7.42}$$

We get:

$$\mathcal{W}_{\hat{\lambda}} = \mathcal{W}_{\lambda^o} + \frac{h^{o\,H}\mathcal{W}_{\lambda^o}h^o}{(h^{o\,H}B^o h^o)^2}B^o h^o h^{o\,H}B^{o\,H} - \frac{B^o h^o h^{o\,H}\mathcal{W}_{\lambda^o}}{h^{o\,H}B^o h^o} - \frac{\mathcal{W}_{\lambda^o}h^o h^{o\,H}B^{o\,H}}{h^{o\,H}B^o h^o} \tag{7.43}$$

- **Performance of DIQML**
  For DIQML, we have:

$$\mathcal{W}_{\lambda^o}(i,j) = \sigma_v^2\left(\mathcal{X}^H\mathcal{R}^+\mathcal{X}\right)(i,j) + \sigma_v^4 \mathrm{tr}\left\{\Delta\mathcal{T}_i^{\perp\,H}\mathcal{R}^+\Delta\mathcal{T}_j^{\perp}\right\} \tag{7.44}$$

  ($B^o = \mathcal{D}$ in equation (7.13)), and:

$$C_{\Delta h}^{\mathrm{DIQML}} = \sigma_v^2\left(\mathcal{X}^H\mathcal{R}^+\mathcal{X}\right)^+ + \sigma_v^4\left(\mathcal{X}^H\mathcal{R}^+\mathcal{X}\right)^+\left(B^o - \frac{B^o h^o h^{o\,H}B^o}{h^{o\,H}B^o h^o}\right)\left(\mathcal{X}^H\mathcal{R}^+\mathcal{X}\right)^+ \tag{7.45}$$

- **Performance of PQML**

  For PQML, we get:

$$\mathcal{W}_{\lambda^o}(i,j) = \sigma_v^2\left(\mathcal{X}^H\mathcal{R}^+\mathcal{X}\right)(i,j) + \sigma_v^4 \mathrm{tr}\left\{\Delta\mathcal{T}_i^{\perp\,H}\mathcal{R}^+\Delta\mathcal{T}_j^{\perp}P_{\mathcal{T}(h)}\right\} \tag{7.46}$$

  $\mathcal{W}_{\lambda^o}$ has one singular vector: $h^o$. Hence, $\mathcal{W}_{\lambda^o}h^o = 0$, so the last three terms in (7.43), due to $\Delta\lambda$, disappear and $\mathcal{W}_{\hat{\lambda}} = \mathcal{W}_{\lambda^o}$. $\mathcal{W}_{\lambda^o}$ is the same as the one for DML [75]. In fact, for PQML, $\Delta\lambda$ is of order $\frac{1}{M}$, whereas $\widetilde{A}(\boldsymbol{Y})h^o$ and $\widetilde{B}(\boldsymbol{Y})h^o$ are of order $\frac{1}{\sqrt{M}}$, so that in (7.30) $\hat{\lambda}(\boldsymbol{Y})$ can be replaced by $\lambda^o$.

$$C_{\Delta h}^{\mathrm{DIQML}} = \sigma_v^2\left(\mathcal{X}^H\mathcal{R}^+\mathcal{X}\right)^+ + \sigma_v^4\left(\mathcal{X}^H\mathcal{R}^+\mathcal{X}\right)^+\left(\left(\mathrm{tr}\left\{\Delta\mathcal{T}_i^{\perp\,H}\mathcal{R}^+\Delta\mathcal{T}_j^{\perp}P_{\mathcal{T}(h)}\right\}\right)\right)\left(\mathcal{X}^H\mathcal{R}^+\mathcal{X}\right)^+ \tag{7.47}$$

$$C_{\Delta h}^{\mathrm{PQML}} = C_{\Delta h}^{\mathrm{DIQML}} - \sigma_v^4\left(\mathcal{X}^H\mathcal{R}^+\mathcal{X}\right)^+\left(\left(\mathrm{tr}\left\{\Delta\mathcal{T}_i^{\perp\,H}\mathcal{R}^+\Delta\mathcal{T}_j^{\perp}P_{\mathcal{T}^H(h^\perp)}\right\}\right)\right)\left(\mathcal{X}^H\mathcal{R}^+\mathcal{X}\right)^+ \tag{7.48}$$

where $((A_{i,j}))$ is another notation for matrix $A$. PQML has better performance than DIQML and $C_{\Delta h}^{PQML} = C_{\Delta h}^{DML}$.

# B Convergence Study of AQML

The study proposed here uses results pointed out in [74] for the convergence study of ILSE (Iterative Least-Squares with Enumeration).

**Definition 1** *Let $\mathcal{F} : \mathcal{U} \to \mathcal{U}$ a mapping from a point of $\mathcal{U}$ to a point in $\mathcal{U}$. A fixed point $u^f \in \mathcal{V}$ of $\mathcal{F}$ verifies $\mathcal{F}(u^{(f)}) = u^{(f)}$.*

**Definition 2** *A function $f$ is a descent function for the mapping $\mathcal{F}$ if:*

1. *$f : \mathcal{V} \to \mathcal{R}$ is non-negative and continuous.*

2. *$f(\bar{u}) < f(u)$ for $\bar{u} = \mathcal{F}(u)$ and $u$ is not a fixed point of $\mathcal{F}$.*

3. *$f(\bar{u}) \leq f(u)$ for $\bar{u} = \mathcal{F}(u)$ and $u$ is a fixed point of $\mathcal{F}$.*

Let $u_n = (A^{(n)}, h^{(n)})$. AQML can be seen as an iterative process that generates the sequence $\{u_n\}$, defined as $u_{n+1} = \mathcal{F}(u_n)$. Let $f$ be the DML cost function; we prove that $f$ is a descent function: AQML decreases the DML cost function at each iteration.

$$\|\boldsymbol{Y} - \mathcal{T}(h^{(i+1)})A^{(i+1)}\| = \min_h \|\boldsymbol{Y} - \mathcal{T}(h)A^{(i+1)}\| \leq \|\boldsymbol{Y} - \mathcal{T}(h^{(i)})A^{(i+1)}\| \tag{7.49}$$

There is strict inequality if $h^{(i+1)} \neq h^{(i)}$, and equality if $h^{(i+1)} = h^{(i)}$.

$$\|\boldsymbol{Y} - \mathcal{T}(h^{(i)})A^{(i+1)}\| = \min_A \|\boldsymbol{Y} - \mathcal{T}(h^{(i)})A\| \leq \|\boldsymbol{Y} - \mathcal{T}(h^{(i)})A^{(i)}\| \tag{7.50}$$

There is strict inequality if $A^{(i+1)} \neq A^{(i)}$, and equality if $A^{(i+1)} = A^{(i)}$. AQML decreases strictly the cost function until a fixed point is attained: the points of convergence of AQML are fixed points. Convergence is reached after an infinite number of steps, unlike ILSE which needs only a finite number of steps.

We prove that the global minimizer of DML is a fixed point of AQML. The DML global minimizer $\hat{h}_{ML}$ verifies:

$$\hat{h}_{ML} = arg \min_h \|\boldsymbol{Y}^H - \mathcal{T}(h) \left(\mathcal{T}^H(h)\mathcal{T}(h)\right)^{-1} \mathcal{T}^H(h)\boldsymbol{Y}\|^2 \tag{7.51}$$

The minimal value of the ML cost function is:

$$\|\boldsymbol{Y} - \mathcal{T}(\hat{h}_{ML}) \underbrace{\left(\mathcal{T}^H(\hat{h}_{ML})\mathcal{T}(\hat{h}_{ML})\right)^{-1} \mathcal{T}^H(\hat{h}_{ML})\boldsymbol{Y}}_{=\widehat{A}} \|^2 . \tag{7.52}$$

If we now initialize AQML by $\widehat{A}$ to compute an estimate $\hat{h}$ of the channel, then:

$$\|\boldsymbol{Y} - \mathcal{T}(\hat{h})\widehat{A}\|^2 < \|\boldsymbol{Y} - \mathcal{T}(h) \left(\mathcal{T}^H(h)\mathcal{T}(h)\right)^{-1} \mathcal{T}^H(h)\boldsymbol{Y}\|^2 \tag{7.53}$$

which contradicts the fact that $\hat{h}_{ML}$ is the global minimizer.

# C Simulations

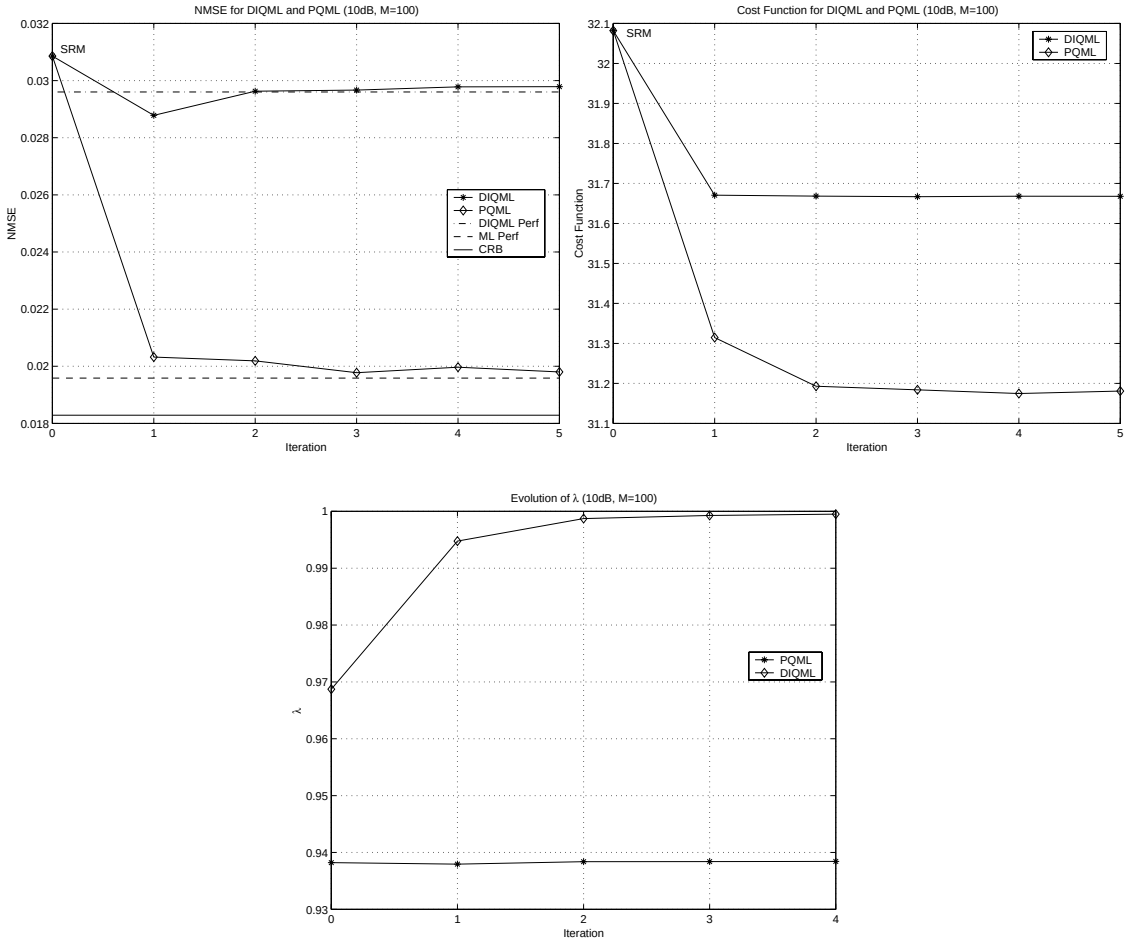



Figure 7.1: NMSE, cost function, generalized eigenvalue for DIQML and PQML at 10dB, for a burst length of 100.

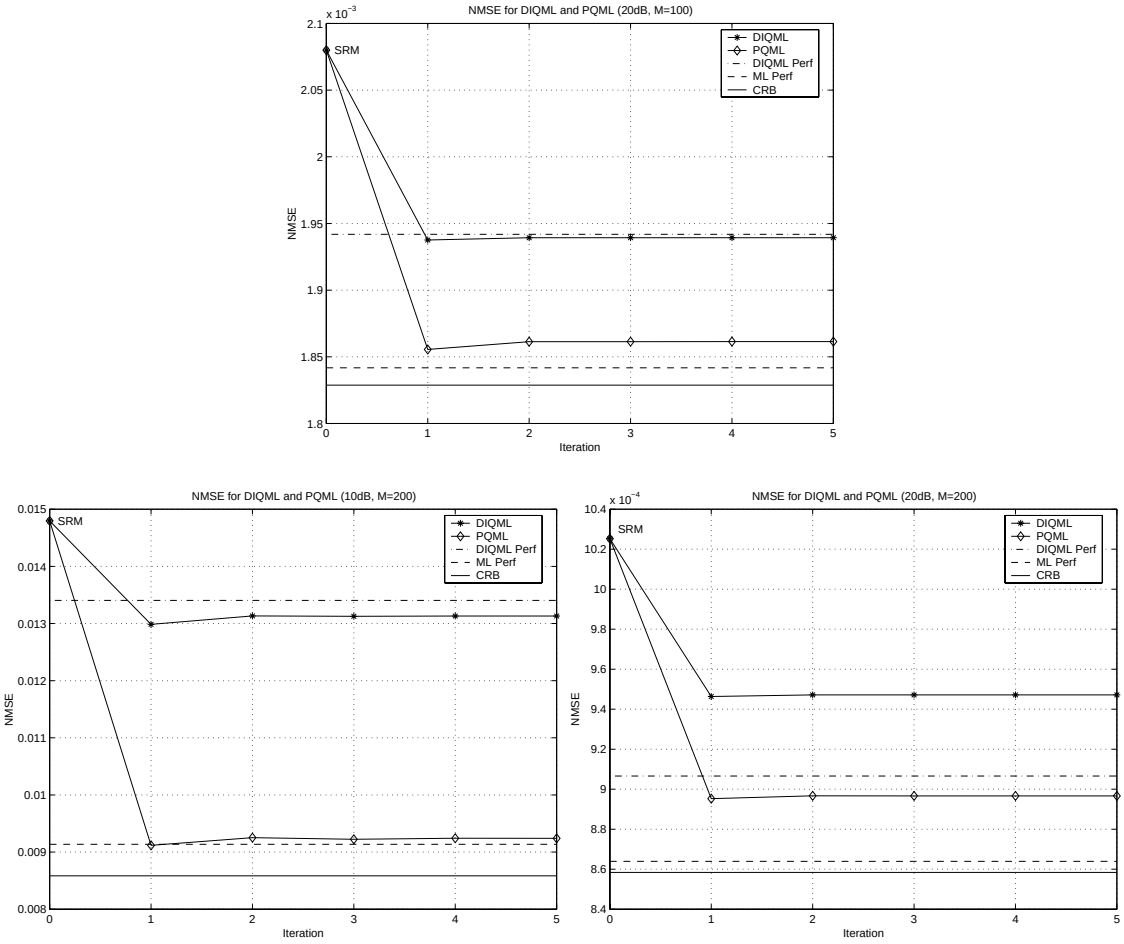



Figure 7.2: NMSE for DIQML and PQML at 10dB and 20dB for a burst length of 100 and 200.

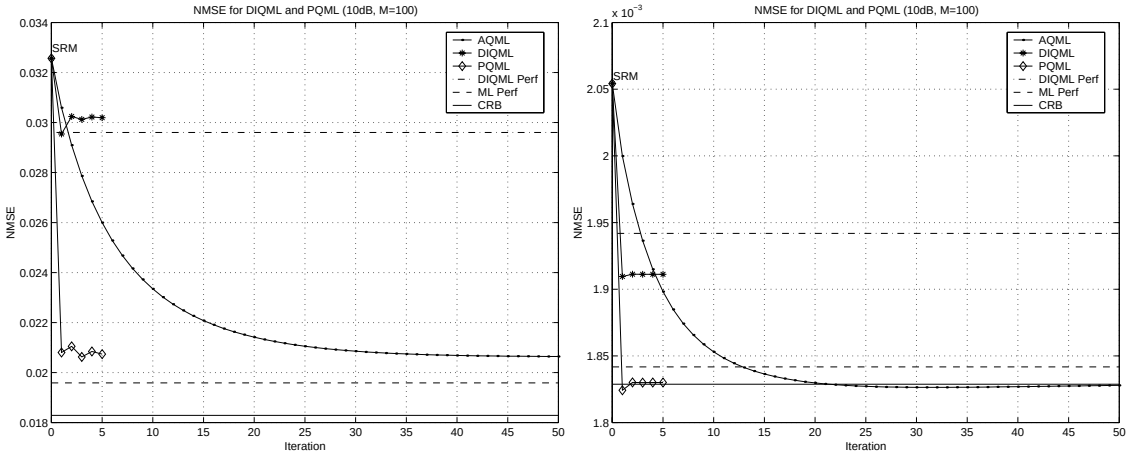



Figure 7.3: Comparison between DIQML, PQML and AQML at 10dB and 20 dB.

Chapter 8

# SEMI-BLIND METHODS BASED ON DETERMINISTIC MAXIMUM–LIKELIHOOD

*In this chapter, we propose semi–blind methods based on DML. Optimal semi–blind methods are first presented that can take into account the information coming from known symbols arbitrarily dispersed over the burst: these methods use the AQML principle. In a second step, the known symbols are assumed grouped: this allows to build suboptimal semi–blind criteria that are formed of a linear combination of a training sequence based criterion and a blind criterion: the weights of this combination are optimal in the ML sense. Three criteria including three different training sequence based estimation methods are optimized using the semi–blind PQML as well as the AQML strategy. We also construct a semi–blind criterion based on SRM which is used as initialization of the DML based semi–blind algorithms. Simulations are presented to illustrate the performance of these algorithms, and comparisons between them are discussed. At last, we suggest a way to build semi–blind criteria combining some blind criterion and a training sequence based criterion. The subspace fitting based semi–blind criterion is given as an example.*

## 8.1 Semi–Blind Methods

### 8.1.1 Optimal Semi–Blind Approaches

In a first step, we consider optimal solutions to the semi–blind estimation problem. Optimal semi–blind algorithms should fulfill a certain number of conditions:

- They should exploit all the information coming from the known and the unknown symbols in the burst, and especially the observations containing known and unknown symbols at the same time. This could be a difficult task, as the classical training sequence based estimation cannot do it and blind estimation does not do it (and considers the known symbols as unknown).
- They should work when the known symbols are arbitrarily dispersed.
- Semi–blind identifiability conditions should also be respected: for example, the methods should work for only one known symbol for irreducible channels, which is not systematically verified by the suboptimal methods.
- With a sufficient number of known symbols, the optimal semi–blind methods should be able to identify any channel, and especially monochannels.

Optimal semi–blind methods are methods that naturally incorporate the knowledge of symbols. Maximum–Likelihood methods fulfill this condition: DML, GML, FA-ML, SML criteria incorporate the known symbols (and for some of them also the unknown symbols). Methods estimating directly the input symbols like [76] are also good semi–blind candidates.

### 8.1.2 Suboptimal Semi–Blind Approaches

When the choice is possible, it is better to have grouped known symbols in general. Indeed, in the contrary case, one looses the blind problem structure on which low computational algorithms can be built. Also, as already seen in chapter 4, it is preferable for performance reasons.

So, in a second step, we considered suboptimal semi–blind approaches which exist when the known symbols are grouped. The proposed semi–blind methods are again based on ML. In that case, a ML based semi–blind criterion can be written as:

$$\text{Semi–blind criterion} = \alpha_1 \text{ Training sequence criterion} + \alpha_2 \text{ Blind criterion}.$$

The weights $\alpha_1$ and $\alpha_2$ are the optimal weights in the ML sense: they are not arbitrary and are deduced from the semi–blind ML problem. Such methods were initiated in [77] and [69].

### 8.1.3 Linear Combination of a Blind and a TS criterion

We will see that it is possible to build semi–blind algorithms by simply linearly combining a training sequence based criterion and a blind criterion (which does not allow to incorporate the knowledge of symbols as ML criteria do). All the difficulty consists then in finding the right weights $\alpha_1$ and $\alpha_2$. One solution would be to determine the weights that optimize the semi–blind performance: an analytical solution is difficult to find and could necessitate a computationally demanding search method. In some of the proposed semi–blind algorithms, the weights take into account the number of data on which the training sequence based criterion and the blind criterion are based. Other possibilities are user-defined weights based on SNR conditions for example. These kinds of choices may appear however arbitrary and not always appropriate. And, in fact, we may even wonder why the linear combination semi–blind solution would be legitimate.

We propose to combine a blind and a training sequence based criterion by optimally weighting the 2 criteria (when the blind criterion is based on least–squares). We treat examples where the optimal weighting matrix is approximated by a diagonal matrix. The resulting criterion becomes then indeed a linear combination of the two criteria. Semi–blind SRM and subspace fitting based criteria built that way will be proposed.

## 8.2 Semi-Blind AQML: a Semi–Blind Optimal Algorithm

Our purpose is to solve the semi–blind DML criterion:

$$\min_{h,A_U} \|\boldsymbol{Y} - \mathcal{T}(h)A\|^2 = \min_{h,A_U} \|\boldsymbol{Y} - \mathcal{T}_K(h)A_K - \mathcal{T}_U(h)A_U\|^2 . \tag{8.1}$$

The alternating minimization strategy already used to solve blind DML can be adapted to semi–blind estimation:

(1) Initialization $\hat{h}^{(0)}$:

(2) Iteration $(i+1)$:

- Minimization w.r.t. $A$, $h = \hat{h}^{(i)}$:

$$\min_{A_U} \|\boldsymbol{Y} - \mathcal{T}(\hat{h}^{(i)})A\|^2 \Leftrightarrow \min_{A_U} \|\boldsymbol{Y} - \mathcal{T}_K(\hat{h}^{(i)})A_K - \mathcal{T}_U(\hat{h}^{(i)})A_U\|^2 \tag{8.2}$$

$$\Rightarrow \widehat{A}_U^{(i+1)} = \left(\mathcal{T}_U^H(\hat{h}^{(i)})\mathcal{T}_U(\hat{h}^{(i)})\right)^{-1} \mathcal{T}_U^H(\hat{h}^{(i)}) \left(\boldsymbol{Y} - \mathcal{T}_K(\hat{h}^{(i)})A_K\right) \tag{8.3}$$

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

$$\Rightarrow \hat{h}^{(i+1)} = \left(\hat{\mathcal{A}}^{(i+1)\,H}\hat{\mathcal{A}}^{(i+1)}\right)^{-1} \hat{\mathcal{A}}^{(i+1)\,H}\boldsymbol{Y} \tag{8.4}$$

$\hat{\mathcal{A}}^{(i+1)}$ is constructed from $A_U^{(i+1)}$ and from $A_K$.

(3) Repeat 1 until $(\widehat{A}^{(i+1)}, \hat{h}^{(i+1)}) \sim (\widehat{A}^{(i)}, \hat{h}^{(i)})$.

Semi–blind AQML inherits the advantages and disadvantages of its blind counterpart. We can prove as in the blind case that at each iteration of semi–blind AQML, the cost function decreases until a fixed point is reached; and again with good initialization, AQML converges to the global minimum. In the case where the known symbols are dispersed, we can see how the blind problem part looses its structure as $\mathcal{T}_U(h)$ has no particular structure properties.

The AQML strategy can also be used to solve semi–blind DML with the Finite Alphabet (FA) constraint in the input symbols: (8.3) is just followed by a decision step.

Semi–blind AQML and the FA-AQML are semi–blind optimal as described previously.

## 8.3 Three Suboptimal Algorithms based on PQML

In this section, we assume that the known symbols are grouped and for simplicity reasons that they are located at the beginning of the burst. We present here three semi–blind algorithms based on the PQML: similar algorithms could have been constructed based on DIQML also, which we will not consider here as it does not perform as well as PQML.

### 8.3.1 PQML Principle for Semi–Blind Estimation

The general semi-blind PQML strategy applies as follows: the gradient of the cost function may be written as $\mathcal{P}(h)h + \mathcal{S}(h)$ where $\mathcal{P}(h)$ is ideally positive definite. At each iteration, we consider $\mathcal{P}(h)$ and $\mathcal{S}(h)$ as constant, and $h$ is the solution of a linear system, which is used to reevaluate $\mathcal{P}(h)$ and $\mathcal{S}(h)$ to perform other iterations.

### 8.3.2 Splitting the Data

The output burst can be decomposed into 3 parts, see figure 8.1 (top):

1. The observations containing only known symbols.

2. The $N-1$ overlap observations containing known and unknown symbols.

3. The observations containing only unknown symbols.

The proposed semi–blind criteria will consider a decomposition of the data into 2 parts, with the overlap zone assimilated to the training part or the to the blind part of the semi–blind criteria.

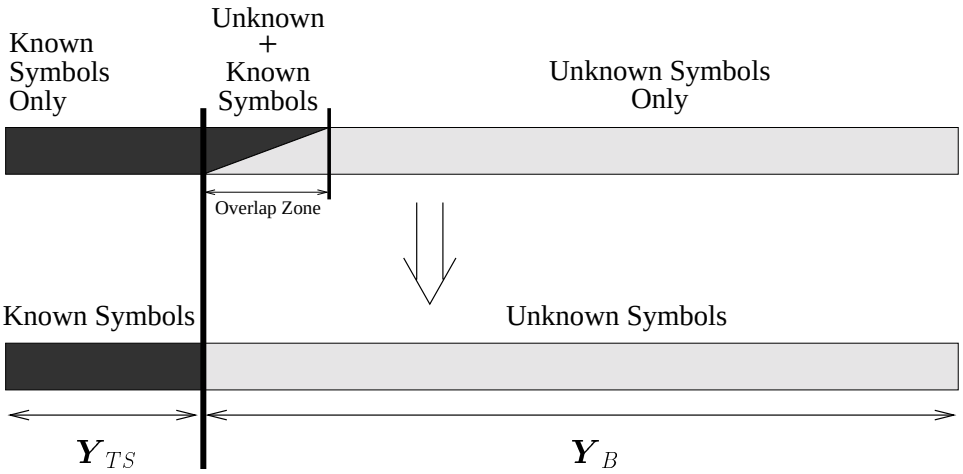



Figure 8.1: Output Burst: split of the data for LS–PQML.

### 8.3.3 Least Squares–PQML (LS–PQML)

#### Split of the Data

In the first semi–blind approach, the overlap zone is assimilated to the blind part of the semi–blind criterion. The data is split as $\boldsymbol{Y} = \left[\boldsymbol{Y}_{TS}^T \;\; \boldsymbol{Y}_B^T\right]^T$, see figure 8.1:

- $\boldsymbol{Y}_{TS} = \mathcal{T}_{TS}(h)A_{TS} + \boldsymbol{V}_{TS}$ groups all the observations containing only known symbols: the input symbols $A_{TS}$ are naturally modeled as known deterministic quantities.
- $\boldsymbol{Y}_B = \mathcal{T}_B(h)A_B + \boldsymbol{V}_B$ groups all the observations containing unknown symbols and especially the overlap observations, where we do not exploit the knowledge of the symbols, which will be treated as unknown. Some information is then lost. This loss of information can be critical especially when the training sequence is very short, of less than $N$ symbols (see the identifiability section).

We apply the DML principle to:

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_{TS} \\ \boldsymbol{Y}_B \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} \mathcal{T}_{TS}(h)A_{TS} \\ \mathcal{T}_B(h)A_B \end{bmatrix}, \sigma_v^2 I \right) \tag{8.5}$$

As $\boldsymbol{Y}_{TS}$ and $\boldsymbol{Y}_B$ are decoupled in term of noise, the DML criterion for $\boldsymbol{Y}$ is the sum of the DML criterion for $\boldsymbol{Y}_{TS}$ and $\boldsymbol{Y}_B$:

$$\min_{h,A_B} \left\{ \|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h)A_{TS}\|^2 + \|\boldsymbol{Y}_B - \mathcal{T}_B(h)A_B\|^2 \right\} . \tag{8.6}$$

This criterion can be optimized by alternating minimizations between $h$ or $A_B$. We can also solve w.r.t. $A_B$ and substitute the solution to get the semi–blind DML criterion for $h$:

$$\min_{h} \left\{ \|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h)A_{TS}\|^2 + \boldsymbol{Y}_B^H P_{\mathcal{T}_B^H(h^\perp)} \boldsymbol{Y}_B \right\} . \tag{8.7}$$

In the following section, we minimize this criterion by semi–blind PQML and prove that it is simply equivalent to optimizing the blind part of the criterion by PQML and the training sequence part by Least-Squares.

### Semi–blind PQML

The gradient of the cost function is:

$$\begin{aligned} & \mathcal{P}_B(h)h - \mathcal{A}_{TS}^H \left(\boldsymbol{Y}_{TS} - \mathcal{A}_{TS}h\right) = 0 \\ \Leftrightarrow \quad & \left[\mathcal{P}_B(h) + \mathcal{A}_{TS}^H\mathcal{A}_{TS}\right] h = \mathcal{A}_{TS}^H \boldsymbol{Y}_{TS} \end{aligned} \tag{8.8}$$

where: $\mathcal{T}_{TS}(h)A_{TS} = \mathcal{A}_{TS}h$. $\mathcal{P}_B(h)h$ corresponds to the blind decomposition of equation (7.19). Our quantities of interest are:

$$\begin{cases} \mathcal{P}(h) = \mathcal{P}_B(h) + \mathcal{A}_{TS}^H\mathcal{A}_{TS} = \mathcal{Y}_B^H\mathcal{R}_B^+\mathcal{Y}_B - \mathcal{B}_B^H\mathcal{B}_B + \mathcal{A}_{TS}^H\mathcal{A}_{TS} \\ \mathcal{S}(h) = -\mathcal{A}_{TS}^H Y_{TS}\,. \end{cases} \tag{8.9}$$

For finite $M$, $\mathcal{Y}_B^H\mathcal{R}_B^+\mathcal{Y}_B - \mathcal{B}_B^H\mathcal{B}_B$ is indefinite: in general, the presence of the training sequence term $\mathcal{A}_{TS}^H\mathcal{A}_{TS}$ allows $\mathcal{P}(h)$ to be positive definite. The generalized eigenvalue strategy could then be avoided; we observed however a better behavior of the algorithm when using it, the convergence speed particularly is higher. Our semi–blind criterion then becomes:

$$\min_{h,\lambda} \left\{ \|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h)A_{TS}\|^2 + h^H \left\{ \mathcal{Y}_B^H\mathcal{R}_B^+\mathcal{Y}_B - \lambda\, \mathcal{B}_B^H\mathcal{B}_B \right\} h \right\} \tag{8.10}$$

with the semi–definite positivity constraint of $\mathcal{P}_B(h)$. When introducing this generalized eigenvalue, the first blind term of $\mathcal{P}_B(h)$ becomes positive semi–definite and the presence of the second training sequence term allows $\mathcal{P}(h)$ to be positive. At a given iteration, $h$ has for expression:

$$h = \left(\mathcal{A}_{TS}^H\mathcal{A}_{TS} + \mathcal{Y}_B^H\mathcal{R}_B^+\mathcal{Y}_B - \lambda\mathcal{B}_B^H\mathcal{B}_B\right)^{-1} \mathcal{A}_{TS}^H Y_{TS} \tag{8.11}$$

where the different quantities are computed thanks to the previous iteration.

### Identifiability

LS–PQML is a suboptimal way of solving the semi–blind problem and the semi–blind identifiability conditions about the number of known symbols necessary no longer hold exactly. For irreducible channels, the previous criterion (8.10) needs at least $N$ known symbols to be well–defined: $\mathcal{P}_B(h)$ is indeed positive semi–definite with 1 singularity, with $N$ known symbols $\mathcal{A}_{TS}^H\mathcal{A}_{TS}$ has rank 1, which is sufficient to allow $\mathcal{P}(h)$ to be positive definite. For a reducible channel with $N_c - 1$ zeros, asymptotically $\mathcal{P}_B(h) \to \mathcal{X}_B^H\mathcal{R}_B^+\mathcal{X}_B$ has $N_c$ singularities, and $N + N_c$ known symbols are necessary to have a well–conditioned problem. Furthermore, monochannels cannot be identified, because the blind criterion is not defined for monochannels.

### 8.3.4 Alternating Quadratic PQML (AQ–PQML)

Here, the overlap zone will be assimilated to the training part of the criterion. The data is split as $\boldsymbol{Y} = [\boldsymbol{Y}_{AQ}^T \ \boldsymbol{Y}_B^T]^T$, see figure 8.2:

- $\boldsymbol{Y}_{AQ} = \mathcal{T}_{AQ}(h)A_{AQ} + \boldsymbol{V}_{AQ} = \mathcal{T}'_K(h)A_{TS} + \mathcal{T}'_U(h)A'_U + \boldsymbol{V}_{AQ}$ groups all the observations containing the known symbols $A_{TS}$ and especially the overlap observations. The unknown symbols in $\boldsymbol{Y}_{AQ}$, $A'_U$, are considered as deterministic.

- $\boldsymbol{Y}_B = \mathcal{T}_B(h)A_U + \boldsymbol{V}_B$ groups all the observations containing only unknown symbols: all the input symbols are considered as deterministic unknown quantities

DML is applied to $[\boldsymbol{Y}_{AQ}^T \ \boldsymbol{Y}_B^T]^T$:

$$\min_{h,A_B,A'_U} \left\{ \|\boldsymbol{Y}_{TS} - \mathcal{T}_{AQ}A_{AQ}\|^2 + \|\boldsymbol{Y}_B - \mathcal{T}_B(h)A_B\|^2 \right\} . \tag{8.12}$$

Semi–blind AQML proceeds as:

1. Initialization $\hat{h}^{(0)}$

2. Iteration (i+1):

   - AQML on $\boldsymbol{Y}_{AQ}$, initialized by $\hat{h}^{(i)}$.
     Criterion $\min_{h,A'_U} \|\boldsymbol{Y}_{AQ} - \mathcal{T}_{AQ}(h)A_{AQ}\|^2 = \min_{h,A'_U} \|\boldsymbol{Y}_{AQ} - \mathcal{T}'_K(h)A_{TS} - \mathcal{T}'_U(h)A'_U\|^2$ solved by alternating minimization on $A'_U$ and $h$. We keep only the estimate of $A'_U$, to form the new estimate of $\widehat{A}_{AQ}$: $\widehat{A}_{AQ}^{(i+1)} = [{A'}_U^{(i+1)^T} \ A_K^T]^T$.

   - Solve the semi–blind criterion to get $\hat{h}^{(i+1)}$:
     We can minimize alternatively between $A_B$ and $h$ starting from $\hat{h}^{(i)}$ based on the criterion :

     $$\min_{h,A_B} \left\{ \|\boldsymbol{Y}_{AQ} - \mathcal{T}_{AQ}(h)A_{AQ}^{(i+1)}\|^2 + \|\boldsymbol{Y}_B - \mathcal{T}(h)A_B\| \right\} . \tag{8.13}$$

     We can also solve the PQML based criterion:

     $$\min_{h,\lambda} \left\{ \|\boldsymbol{Y}_{AQ} - \mathcal{T}_{AQ}(h)A_{AQ}^{(i+1)}\|^2 + h^H \left( \mathcal{Y}_B^H \mathcal{R}_B^+(\hat{h}^{(i)}) \mathcal{Y}_B - \lambda \mathcal{B}_B^H(\hat{h}^{(i)}) \mathcal{B}_B(\hat{h}^{(i)}) \right) h \right\} . \tag{8.14}$$

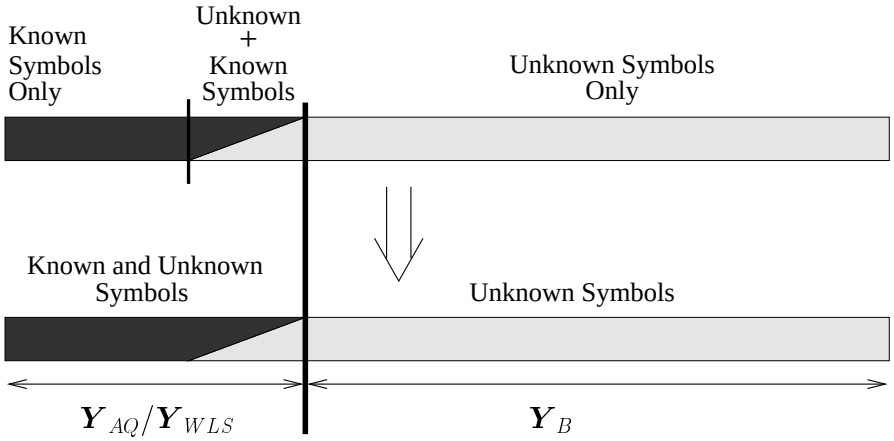



Figure 8.2: Output Burst: split of the data for AQ–PQML and PQML-LS.

### 8.3.5 Weighted–Least–Squares–PQML (WLS–PQML)

#### Split of the Data

WLS–PQML is based on the same decomposition as AQ–PQML. It mixes a deterministic and a Gaussian point of view: the unknown symbols $A'_U$ in the overlap zone are modeled as i.i.d. Gaussian random variables of mean 0 and variance $\sigma_a^2$. We denote $\boldsymbol{Y}_{WLS} = \mathcal{T}_{WLS}(h)A_{WLS} + \boldsymbol{V}_{WLS} = \mathcal{T}'_K(h)A_{TS} + \mathcal{T}'_U(h)A'_U + \boldsymbol{V}_{WLS}$. GML is applied to $\boldsymbol{Y}_{WLS}$ and DML to $\boldsymbol{Y}_B$:

$$\left\{ \begin{array}{l} \boldsymbol{Y} = \left[ \begin{array}{c} \boldsymbol{Y}_{WLS} \\ \boldsymbol{Y}_B \end{array} \right] \sim \mathcal{N}\left( \left[ \begin{array}{c} \mathcal{T}'_K(h)A_{TS} \\ \mathcal{T}_B(h)A_B \end{array} \right], \left[ \begin{array}{cc} C_{Y_{WLS}Y_{WLS}} & 0 \\ 0 & \sigma_v^2 I \end{array} \right] \right), \\ C_{Y_{WLS}Y_{WLS}} = \sigma_a^2 \mathcal{T}'_U(h)\mathcal{T}'_U(h) + \sigma_v^2 I \end{array} \right. \tag{8.15}$$

The mixed ML criterion is:

$$\min_{h,\sigma_v^2} \left\{ \ln\det C_{Y_{WLS}Y_{WLS}} + \left(\boldsymbol{Y}_{WLS} - \mathcal{T}'_K(h)A_{TS}\right)^H C^{-1}_{Y_{WLS}Y_{WLS}} \left(\boldsymbol{Y}_{WLS} - \mathcal{T}'_K(h)A_{TS}\right) + \ln\det \sigma_v^2 I + \frac{1}{\sigma_v^2} \left\| \boldsymbol{Y}_B^H - \mathcal{T}_B(h)A_B \right\|^2 \right\}. \tag{8.16}$$

Again, we optimize this criterion by semi–blind PQML, considering $\sigma_v^2$ as known: in practice, it will be estimated apart. We prove that it is simply equivalent to solving the blind part of the criterion in the PQML way and the training sequence part by weighted least-squares, *i.e.* we neglect the term in det and in the second term, consider $C_{Y_{WLS}Y_{WLS}}$ as constant (computed from the previous iteration).

### Semi–Blind PQML

If we apply the PQML strategy to the training part of the cost function:

$$\ln\det C_{Y_{WLS}Y_{WLS}} + \left(\boldsymbol{Y}_{WLS} - \mathcal{T}'_K(h)A_{TS}\right)^H C^{-1}_{Y_{WLS}Y_{WLS}} \left(\boldsymbol{Y}_{WLS} - \mathcal{T}'_K(h)A_{TS}\right) \tag{8.17}$$

we can prove that it can be approximated by the optimally WLS criterion. The mixed criterion becomes:

$$\min_h \left\{ \|\boldsymbol{Y}_{WLS} - \mathcal{T}_{WLS}(h)A_{WLS}\|^2_{C^{-1}_{Y_{WLS}Y_{WLS}}} + \frac{1}{\sigma_v^2}\boldsymbol{Y}_B^H P_{\mathcal{T}_B^H(h^\perp)}\boldsymbol{Y}_B \right\} . \tag{8.18}$$

The PQML quantities are:

$$\begin{cases} \mathcal{P}(h) = \dfrac{1}{\sigma_v^2}\mathcal{P}_B(h) + \mathcal{A}^H_{WLS}C^{-1}_{Y_{WLS}Y_{WLS}}\mathcal{A}_{WLS} \\ \mathcal{S}(h) = -\mathcal{A}^H_{WLS}C^{-1}_{Y_{WLS}Y_{WLS}}\boldsymbol{Y}_{WLS} \end{cases} \tag{8.19}$$

with $\mathcal{T}'_K(h)A_{TS} = \mathcal{A}_{WLS}h$. At each iteration, the solution for $h$ is:

$$h = \left( \frac{1}{\sigma_v^2}\left(\mathcal{Y}_B^H\mathcal{R}_B^+\mathcal{Y}_B - \lambda\mathcal{B}_B^H\mathcal{B}_B\right) + \mathcal{A}^H_{WLS}C^{-1}_{Y_{WLS}Y_{WLS}}\mathcal{A}_{WLS} \right)^{-1} \mathcal{A}^H_{WLS}C^{-1}_{Y_{WLS}Y_{WLS}}Y_{WLS} . \tag{8.20}$$

where $\lambda$ is the minimal generalized eigenvalue of $\mathcal{Y}_B^H\mathcal{R}_B^+\mathcal{Y}_B$ and $\mathcal{B}_B^H\mathcal{B}_B$.

Now, alternating minimizations between $h$ and $A_B$ can be done on the criterion:

$$\min_{h,A_B} \left\{ \|\boldsymbol{Y}_{WLS} - \mathcal{T}_{WLS}(h)A_{WLS}\|^2_{C^{-1}_{Y_{WLS}Y_{WLS}}} + \frac{1}{\sigma_v^2}\|\boldsymbol{Y}_B - \mathcal{T}_B(h)A_B\|^2 \right\} . \tag{8.21}$$

AQ–PQML and WLS–PQML outperforms LS–PQML because the information coming from the known symbols in the overlap zone is used.

### Identifiability

For an irreducible channel, AQ–PQML and WLS–PQML are defined with only 1 known symbol. For a reducible channel with $N_c - 1$ zeros, $N_c$ known symbols are sufficient to have a well–conditioned problem.

## 8.3.6 Performance

The semi–blind performance can be seen from 2 points of view.

- Asymptotic number of unknown and known symbols.

  In this case, the overlap received data containing known and unknown symbols at the same time can be neglected: LS-PQML, WLS-PQML and AQ-PQML becomes

equivalent and reach the semi–blind DML performance. Assume we have a consistent estimate of the channel, an iteration gives the global minimizer (results that can be obtained by the same asymptotic reasonings as for the blind PQML in Chapter 7).

- Asymptotic number of unknown symbols, finite number of known symbols.

  We prove in [78] that the performance expressions we have in the previous asymptotic conditions are also valid when the number of known symbols is small. This fact is verified by simulations.

## 8.4 Initialization of the Semi–Blind ML Algorithms

The PQML based semi–blind algorithms needs an initialization. A natural initialization is by a semi–blind algorithm based on SRM, which will not need any initialization (we will see that we will need in fact only an estimate of the norm of the channel).

### 8.4.1 Semi–Blind SRM as a Linear Combination of Blind SRM and TS Criterion

Semi–Blind SRM illustrates the fact that it may be difficult to build a semi–blind criterion as a linear combination of a blind and a TS criterion. Indeed, consider the following cost function:

$$\|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h)A_{TS}\|^2 + \alpha\, h^H \mathcal{Y}_B^H \mathcal{Y}_B h\,. \tag{8.22}$$

(used the decomposition of figure 8.1). The blind SRM criterion $min_h h^H \mathcal{Y}^H \mathcal{Y} h$ gives unbiased estimates only under the constant norm constraint for the channel. As the semi–blind criterion is optimized without constraints, the blind SRM part gives biased estimates which renders the performance of the semi–blind algorithm poor.

For the criterion to be unbiased, the term $\mathcal{Y}_B^H \mathcal{Y}_B$ needs to be denoised. We remove $\lambda_{min}(\mathcal{Y}_B^H \mathcal{Y}_B)$ (the minimum eigenvalue of $\mathcal{Y}_B^H \mathcal{Y}_B$), the resulting matrix $\mathcal{Y}_B^H \mathcal{Y}_B - \lambda_{min}(\mathcal{Y}_B^H \mathcal{Y}_B) I$ has exactly one singularity.

Once the criterion is denoised, the choice for the constant $\alpha$ remains unsolved. A way to determine this factor would be to minimize the asymptotic performance of the semi–blind SRM channel estimate (computed with $M_U$ and $M_K$ considered as infinite). In our case, it is impossible analytically, and search techniques would represent an increase in complexity.

In the next section, we construct semi–blind SRM as an approximation of DIQML: the blind SRM part will be automatically denoised.

### 8.4.2 Semi–Blind SRM as an Approximation of DIQML

We know that the semi–blind ML criterion gives the optimal weights between the blind and training sequence part:

$$\|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h)A_{TS}\|^2 + h^H \mathcal{Y}_B^H \mathcal{R}^+(h)\mathcal{Y}_B h\,. \tag{8.23}$$

We neglect the off-diagonal terms of $\mathcal{R}(h)$: $\mathcal{R}(h) = \mathcal{D}(h)$. For $m = 2$, the diagonal is constant with elements equal to $\|h\|^2$. For $m > 2$, the diagonal contains the squared norm of each set of 2 subchannels. For example, for $m = 3$, the first three diagonal elements are: $\|\mathrm{H}_1\|_F^2 + \|\mathrm{H}_2]\|_F^2$, $\|\mathrm{H}_2\|_F^2 + \|\mathrm{H}_3\|_F^2$, $\|\mathrm{H}_1\|_F^2 + \|\mathrm{H}_3\|_F^2$ and are repeated along the diagonal $M$ times.

With this approximation, (8.23) becomes:

$$\|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h)A_{TS}\|^2 + h^H \mathcal{Y}_B^H \mathcal{D}^+(h)\mathcal{Y}_B h \tag{8.24}$$

and we optimize it in the DIQML way in order to denoise it; $\mathcal{D}(h) = \mathcal{D}$ is considered as constant. The semi–blind criterion becomes:

$$\min_h \left\{ \|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h)A_{TS}\|^2 + h^H \left( \mathcal{Y}_B^H \mathcal{D}^{-1} \mathcal{Y}_B - \widehat{\sigma_v^2}\mathcal{D}_v \right) h \right\} \tag{8.25}$$

where $h^H \mathcal{D}_v h = \mathrm{tr}\left\{ \mathcal{T}^H(h^\perp)\mathcal{D}^{-1}\mathcal{T}(h^\perp) \right\}$, and $\widehat{\sigma_v^2}$ will be the generalized eigenvalue of $\mathcal{Y}_B^H \mathcal{D}^{-1} \mathcal{Y}_B$ and $\mathcal{D}_v$.

The norm of the different subchannels, used to compute $\mathcal{D}$, can be recovered by an estimate of the denoised second–order moment of a data sample $r_{yy}(0) = \sigma_a^2 \boldsymbol{H}\boldsymbol{H}^H$: $\hat{r}_{yy}(0) - \widehat{\sigma_v^2} I = \sum_{i=0}^{M-1} \boldsymbol{y}(k)\boldsymbol{y}^H(k) - \widehat{\sigma_v^2} I$. As will be seen in the simulations, at low SNR, the weight on the blind part should be in fact lower than the true value of $\mathcal{D}$, $\mathcal{D}^o$. So instead of estimating the energy of each channel by the denoised $r_{YY}(0)$, we determine it by the noisy one: the resulting $\mathcal{D}$ will be lower than $\mathcal{D}^o$.

In general, the different channels will have the same energy, so that $\mathcal{D}$ can be considered as a constant diagonal matrix with element $\widehat{\|h\|^2}\frac{m}{2}$. When $m = 2$, this is not an approximation. With $\mathcal{D}$ as constant diagonal, $\widehat{\sigma_v^2}$ is the minimal eigenvalue of $\mathcal{Y}_B^H \mathcal{Y}_B$, and the semi–blind criterion becomes:

$$\min_h \left\{ \|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h)A_{TS}\|^2 + \frac{2}{m}\frac{1}{\widehat{\|h\|^2}}\; h^H \left( \mathcal{Y}_B^H \mathcal{Y}_B - \lambda_{min}(\mathcal{Y}_B^H \mathcal{Y}_B) \right) h \right\}. \tag{8.26}$$

An alternative to this semi–blind SRM criterion, is to use the decomposition of figure 8.6 and use WLS or AQML to solve its training sequence part as for PQML. We will call criterion (8.26) LS-SRM and the alternatives WLS-SRM and AQ-SRM.

To make a link with section 8.6, we can note that (8.23) can be seen as an optimally weighted combination of blind SRM and TS based criterion. The semi–blind SRM is deduced by treating correctly the IQML strategy (the denoising) and approximating the weighting.

## 8.5 Simulations

We illustrate here by simulations the behavior of the different semi–blind algorithms. Most of the simulations are based on 1000 Monte–Carlo runs of noise and the input symbols, as well as of the channels. We tested channels with randomly chosen coefficients, and GSM channels according to two typical GSM propagation environments: Typical Urban (TU) and Hilly Terrain (HI) as specified in the ETSI standard [79].

### 8.5.1 Simulations for semi–blind SRM

**Non Weighted, Non Denoised Semi–Blind SRM** In figure 8.5, we see the effect of not denoising the semi–blind SRM criterion: the NMSE is plotted w.r.t. $\alpha$ in equation (8.22). For $\alpha = 1$ (which corresponds to a simple concatenation of the blind and the TS criterion equations), performance is worse than with TS channel estimation. We furthermore notice that this incorrectly built semi–blind criterion is very sensitive to the value of $\alpha$.

**Weighted and Denoised Semi–Blind SRM** In figure 8.6, we plot the NMSE w.r.t. $\alpha\|H\|^2$. At relatively high SNRs, we can see that the optimal $\alpha$ is closed to $1/\|H\|^2$. At lower SNRs, however, the optimal $\alpha$ is lower than $1/\|H\|^2$. We call $\alpha_n$ the $\alpha$ in (8.26) obtained from the noisy received signal covariance matrix and $\alpha_d$, the one obtained from the denoised covariance matrix. We show the case of 7 known symbols for which TS estimation is not defined. From this simulations, we can conclude that the approximation of the weighting matrix is valid; furthermore, we see that the criterion is relatively insensitive to the value of $\alpha$.

**Underestimation of the Channel Order** In figure 8.7, we see the effect of underestimating the channel order and how semi–blind estimation allows to overcome this problem which blind deterministic methods cannot do.

**Channel with a Common Zero** In figure 8.8, a channel with a common zero is tested.

**Channel with more than 2 Subchannels** In Figure 8.9, the case of a channel with 6 subchannels is shown. Semi–blind SRM as in (8.25) and as in (8.26) is tested. For subchannels with different energies, (8.25) gives slightly better results than (8.26), but the gain is not significant. So, approximation (8.26), for $m > 2$, is valid.

**Semi–Blind SRM vs SNR** At last, in figure 8.11 and 8.12, we plot LS-SRM, WLS-SRM and AQ-SRM and compare it to TS estimation as well as blind estimation (with the scale factor adjusted using the training sequence). Two kinds of channels are tested: random channels and GSM channels (model TU). Ten Monte–Carlo realizations of the channels are done and for each realization, 100 realizations of the noise and the input symbols. We see the gain

brought by semi–blind techniques w.r.t. to blind and TS methods. Blind SRM performs particularly poorly: this is due to the fact that among all the realizations of the channel several ones completely failed (we did not reject the worst realizations), but this did not happen for semi–blind SRM. WLS-SRM appears as the best method (especially for the GSM channels). Semi–blind performance is very closed to the optimal theoretical performance of semi–blind ML, except for difficult experimental conditions: GSM channels and 6 known symbols (for which TS estimation is not defined).

### 8.5.2 Simulations for Semi–Blind PQML

Simulations for the PQML based semi–blind algorithms are presented in figures 8.13–8.18 for severe experimental conditions. The semi–blind algorithms are compared to the semi–blind FA method (see section 8.2).

We show the NMSE given by all the semi–blind algorithms in (8.13) for a randomly chosen channel ($N = 4$, $m = 2$) and 1000 Monte-Carlo runs of the noise and input symbols; 7 symbols are known (which is the limit for TS identifiability). The PQML based semi–blind criterion and AQML are initialized either by TS or LS–SRM. The semi–blind algorithms improve dramatically TS performance and give better results that blind PQML. LS–SRM gives performance very closed to the optimal ML performance so that semi–blind PQML cannot really bring a significant improvement. The FA algorithm outperforms the semi–blind criteria, but we notice a certain sensitivity to the initialization at 5dB.

To illustrate the lack of robustness of blind methods, we show in figure 8.14 simulations with Monte–Carlo runs on random channels also: the particularly poor performance of blind estimation can be noticed.

GSM channels are also tested. For TU channels ($N = 4$, $m = 2$), 7 known symbols and $M = 100$ (figure 8.15), semi–blind PQML becomes more sensitive to bad experimental conditions, especially at 5dB. For a very short burst of $M = 50$, the semi–blind algorithms performs as well as the FA method at 5dB and 10dB. This remark is also valid for the experiments at 5 known symbols (figure 8.17) for which TS estimation does not work. HI channels ($N = 7$, $m = 2$) are at last tested (figure 8.18) for which again, except at 20dB, FA methods do not improve significantly semi–blind methods.

### 8.5.3 Conclusions Drawn from the Simulations

From the simulations, we can draw the following important conclusions:

- The proposed semi–blind methods outperforms the TS based methods but also blind methods which lack robustness to poor experimental conditions.
- Semi–blind methods are very good candidates to initialize the FA algorithm, especially when the training sequence is too short to estimate the channel. For severe experimental

conditions, it is even more prudent to use semi–blind criteria because FA methods tend to converge to local minima.

- Among all the criteria proposed, the best one appears to be WLS-PQML.
- Semi–Blind SRM, a particularly simple semi–blind algorithm, appears also as very attractive as its performance is closed to WLS–PQML.

## 8.6 Semi–Blind Criteria as a Combination of a Blind and a TS Based Criteria

### 8.6.1 Linear Combination

Some algorithms have been proposed that linearly combine a TS and a blind criteria: in [80] a semi–blind criterion is proposed based on the (non–weighted) Subspace Fitting (SF) criterion; in [81, 82], another one is based on the blind CMA criterion.

Finding the right weights is not an easy task. Take the example of SF based semi–blind cost function:

$$\alpha \, \|\widehat{P}^{\perp}_{\mathcal{T}_L^{(B)}(\hat{h})} \mathcal{T}_L^{(B)}(h)\|^2 + \|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_{TS}\|^2 \quad \Leftrightarrow \quad \alpha \, h^H \mathcal{S}^H \mathcal{S} \, h + \|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_{TS}\|^2 \, . \tag{8.27}$$

We adopt here the decomposition of figure 8.1; $L$ is the size of the convolution matrix considered. In [80], $\alpha$ was chosen equal to the number of data on which the blind criterion is based, *i.e.* $M_U$.

In figures 8.3, we plot the NMSE of channel estimation w.r.t. $\alpha$ for different size $L$, for 20dB, 20 known symbols and 10dB, 10 known symbols. For $L = N$, the semi–blind criterion is relatively insensitive to the value of $\alpha$. For $L$ larger than $N$, however, it is visibly very sensitive to its value, especially for a small SNR and small number of known symbols. The choice $\alpha = M_U$ can give performance worse than that for training sequence based estimation. These simulations suggest that the linearly combined semi–blind algorithm is sensitive to the dimension of the noise subspace which varies when $L$ varies.

As explained for the semi–blind SRM example, trying to determine the value of $\alpha$ that minimizes the (theoretical) semi–blind performance expression is not a viable solution, especially for computational reasons.

### 8.6.2 Weighted Combination

One solution to build a criteria combining a blind and a TS criteria is to optimally weight the 2 criteria. The semi–blind SRM criterion is built this way, and the proposed simplification, which appears in a linear combination form, is based on the weighting matrix of the blind criterion.

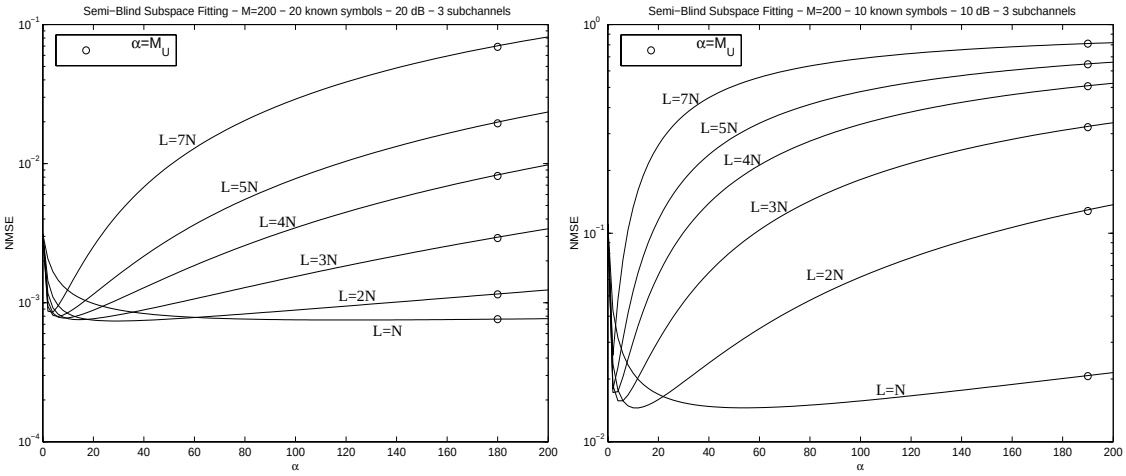



Figure 8.3: Semi–blind subspace fitting built as a linear combination of blind SF and training sequence based criteria.

Although we did not test it yet (for lack of time), the proper semi-blind SF cost function would be :

$$\min_h \left\{ h^H \mathcal{S}^H \mathcal{W}^+ \mathcal{S} h + \frac{1}{\sigma_v^2} \| \boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_{TS} \|^2 \right\} \tag{8.28}$$

where $\mathcal{W} = \mathrm{E}\left(\mathcal{S}\, h h^H \mathcal{S}^H\right)$ is the optimal weighting matrix, whose expression can be found in [83]. $\mathcal{W}$ is of the form $\mathcal{W} = \frac{1}{M_U}\left(\sigma_v^2 \mathcal{W}^{(o)} + \sigma_v^4 \mathcal{W}^{(1)}\right)$ and is computed thanks to a consistent channel estimate, as for the blind WSF criterion. $\mathcal{W}$ gives naturally the term $M_U$ as component of the weighting of blind SF part.

Considering that the weighting matrix contains the information about the dimension $\mathcal{N}$ of the noise subspace in the criterion, we propose to replace $\mathcal{W}$ by $\frac{1}{M_U}\sigma_v^2 \mathcal{N} I$, the semi–blind SF criterion is then

$$\min_h \left\{ \frac{M_U}{\mathcal{N}}\, h^H \mathcal{S}^H \mathcal{S} h + \| \boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_{TS} \|^2 \right\} . \tag{8.29}$$

Let us specify that this is just a first try, and that we will be looking for a better justification (and maybe improvement) for this semi–blind criterion. In fact, this weighting turns out to give satisfactory results, as illustrated in figure 8.4 (Semi–blind SF (1) is the criterion (8.27) and Semi–blind SF (2) is the criterion (8.29)), with $\alpha$ as coefficient of the blind part.

## 8.7 Conclusion

In this chapter we have derived semi–blind algorithms all based on DML. We have seen through simulations that semi–blind algorithms perform better than their blind counterpart,

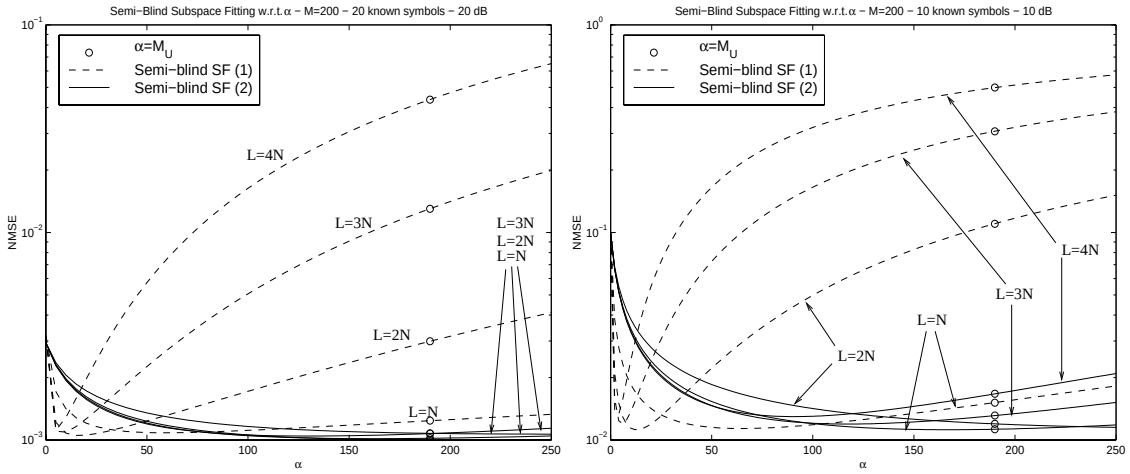



Figure 8.4: Semi–blind subspace fitting built as a linear combination of blind SF and training sequence based criteria.

especially for ill–conditioned channels like GSM channels. We have seen that semi–blind algorithms perform well when TS methods cannot work because the training sequence is too short. In this case they appear as particularly appropriate to initialize methods that exploit the finite alphabet of the input symbols. The best method appears to be WLS–PQML, based on optimally weighted least squares for the TS part of the criterion. Being able to take this overlap observations was one the challenges of semi–blind for it to work very well with few known symbols, *i.e.* fewer symbols that what is required to estimate the channel by training sequence, and also bad experimental conditions. The example of semi–blind SRM and SF have shown us that semi–blind criteria combining a certain blind criterion and a TS criterion are not trivial to construct and we have proposed a way to manage to proceed to a successful combination.

# A Simulations for the Semi–Blind Algorithms

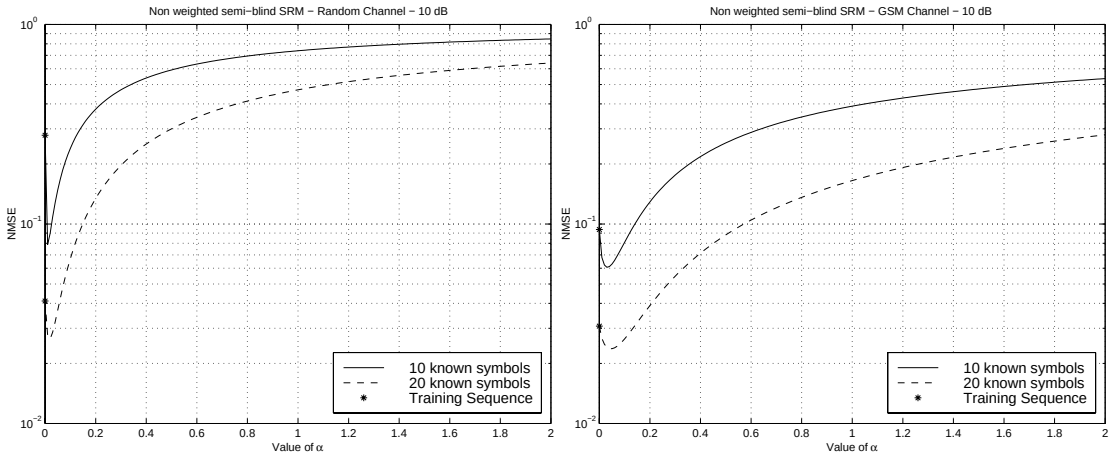



Figure 8.5: Performance of the **non–weighted**, **non–denoised** Semi–Blind SRM (SB–SRM) w.r.t the scalar $\alpha$ for a random channel (left) and a GSM channel (right). For $\alpha = 1$, semi–blind SRM performs worse than TS estimation ($\alpha = 0$).

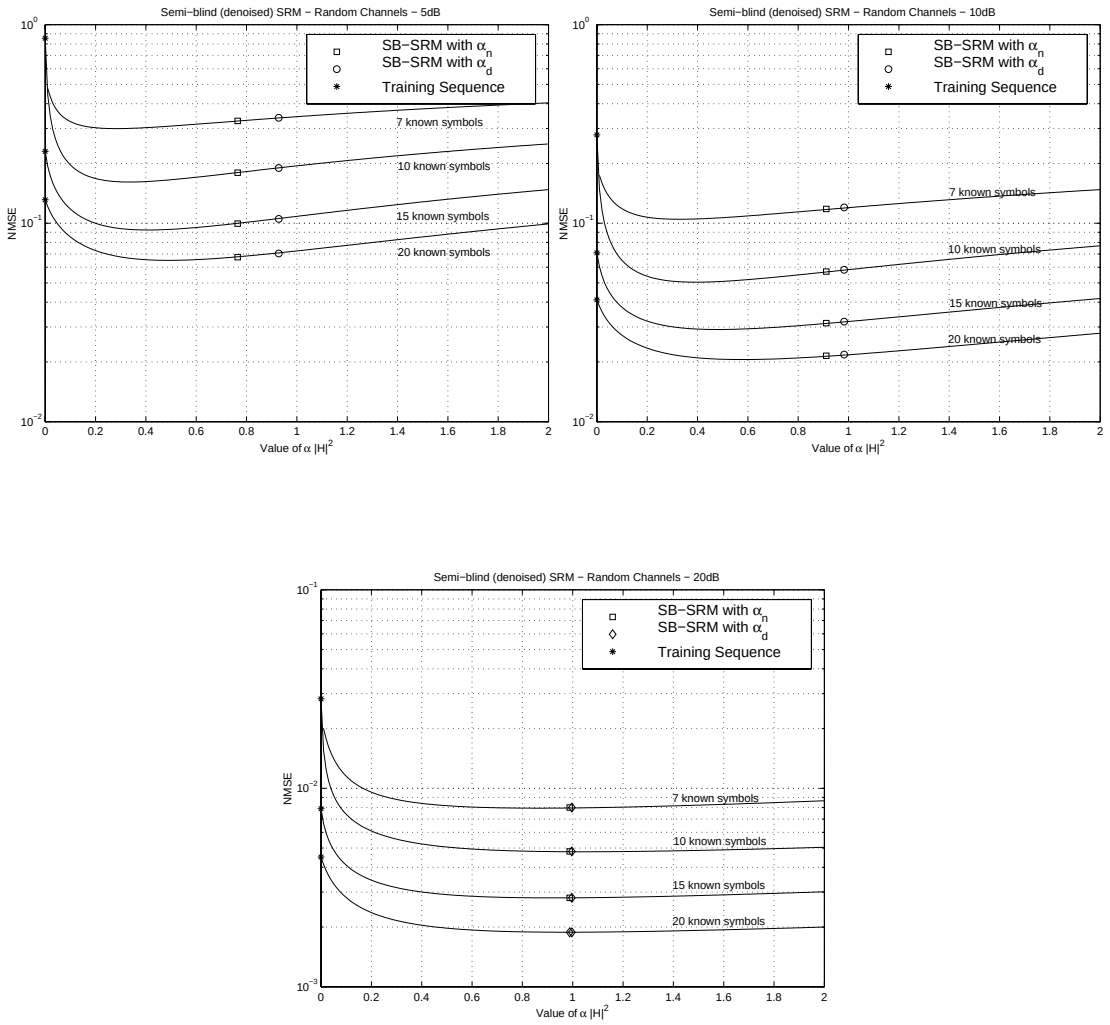



Figure 8.6: Performance of the weighted and denoised SRM with right channel order: semi–blind SRM is quite insensitive to the value of $\alpha$ (around 1) especially at high SNR.

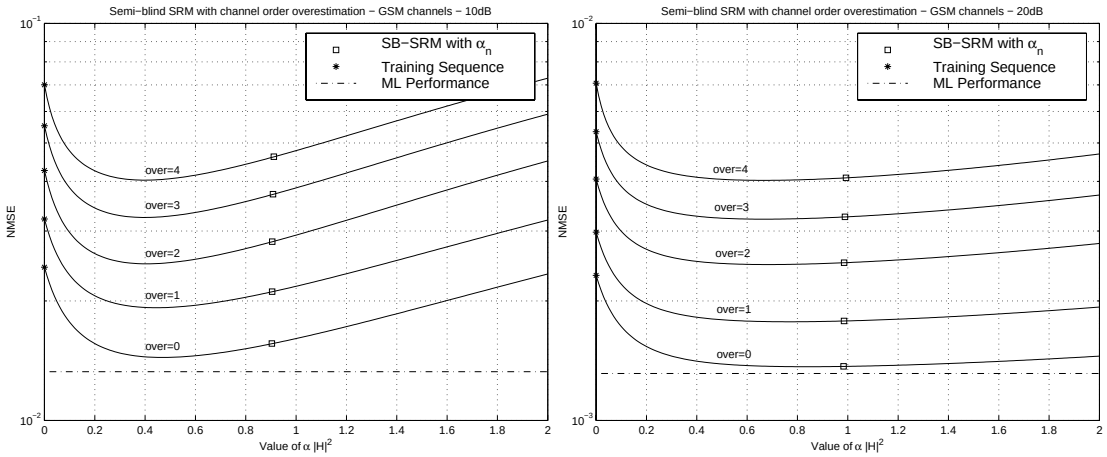



Figure 8.7: **Channel Order Overestimation:** performance of semi–blind SRM w.r.t. the scalar $\alpha$; over= $N' - N$, where $N'$ is the overestimated channel length.

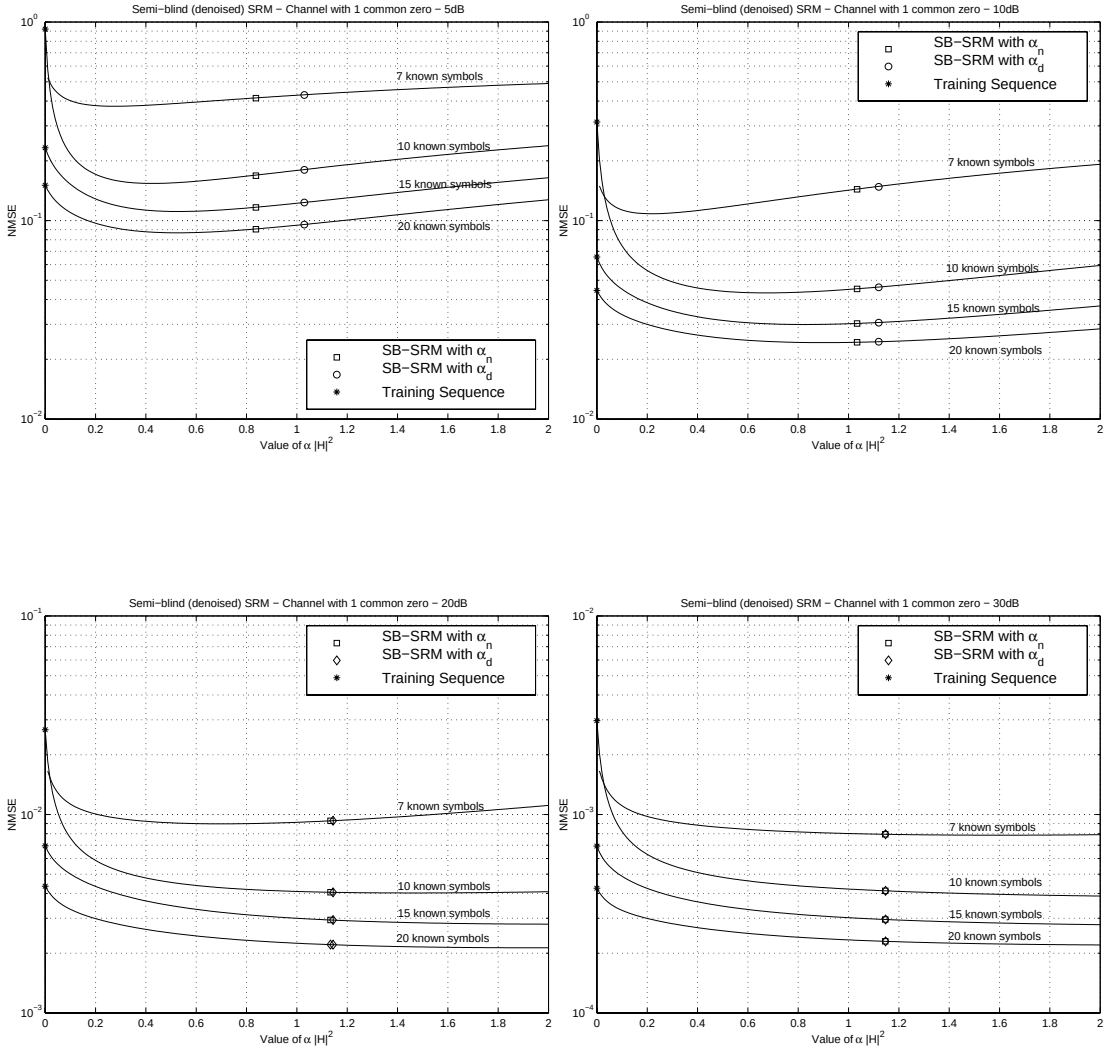



Figure 8.8: **Channel with a common zero:** performance of semi–blind SRM w.r.t. the scalar $\alpha$.

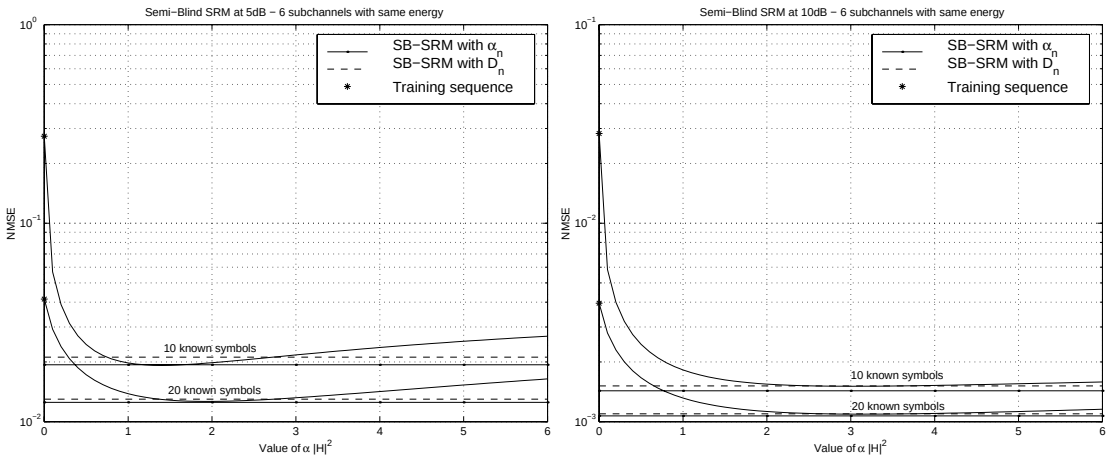



Figure 8.9: **Six Subchannels with same energy:** performance of semi–blind SRM w.r.t the scalar $\alpha$.

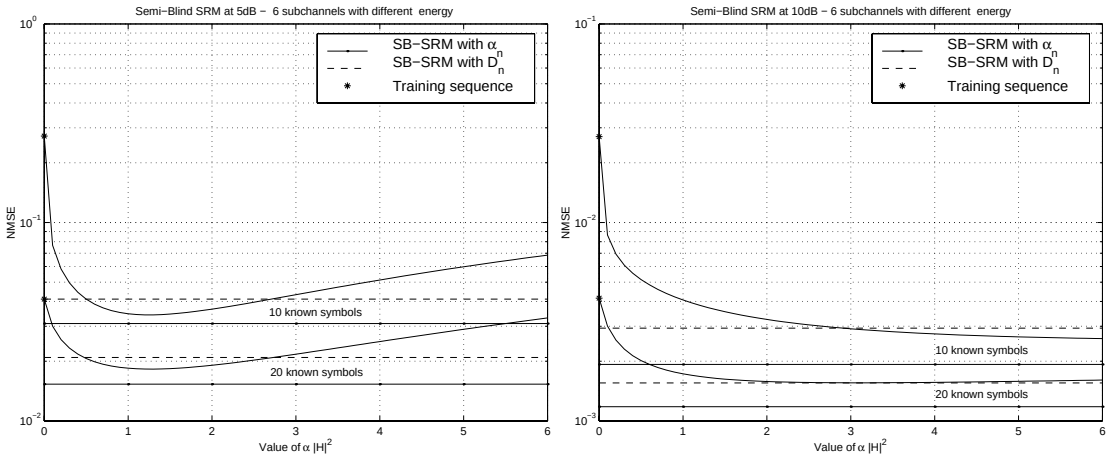



Figure 8.10: **Six Subchannels with different energy:** performance of w.r.t the scalar $\alpha$ w.r.t the scalar $\alpha$.

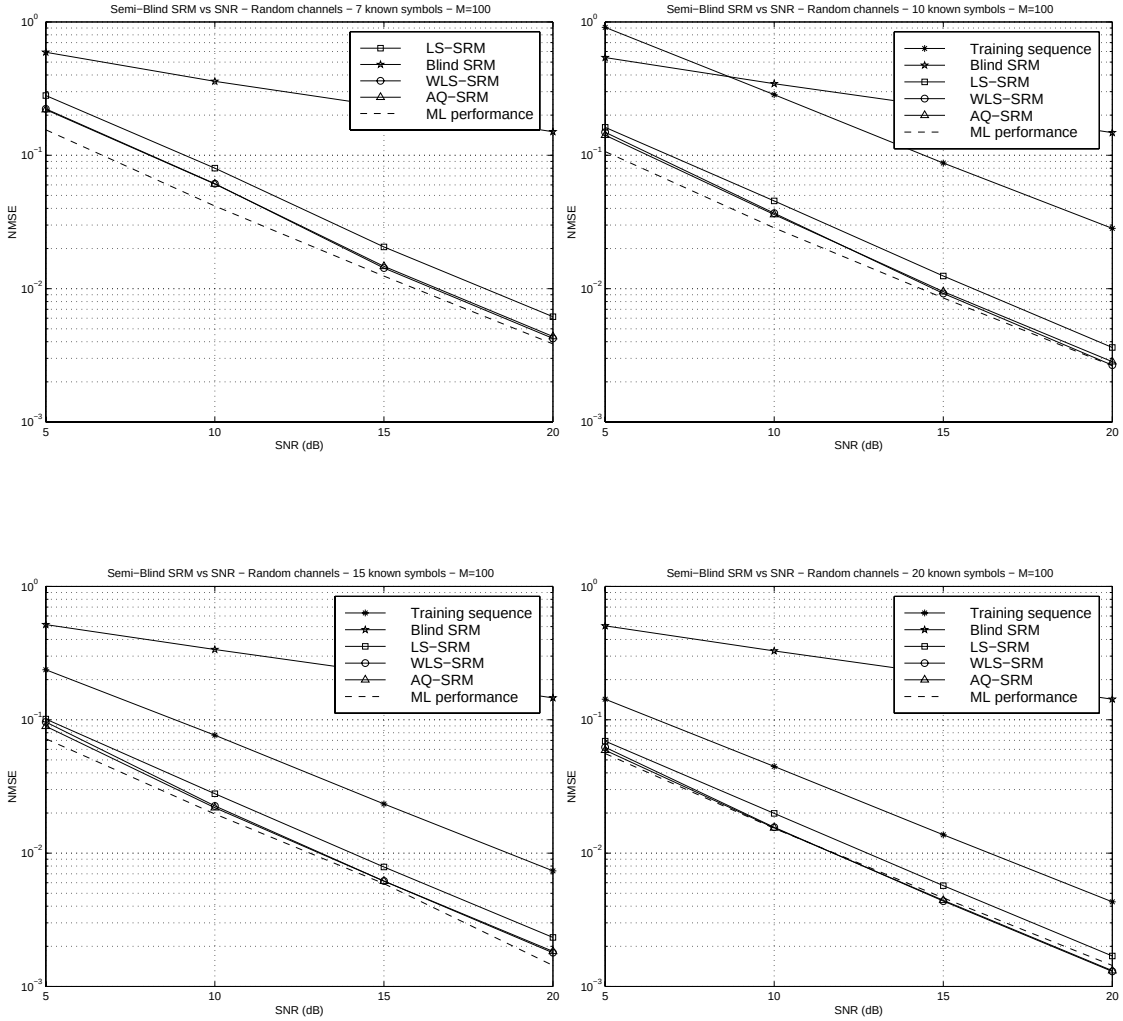



Figure 8.11: Performance of semi–blind SRM w.r.t the SNR.

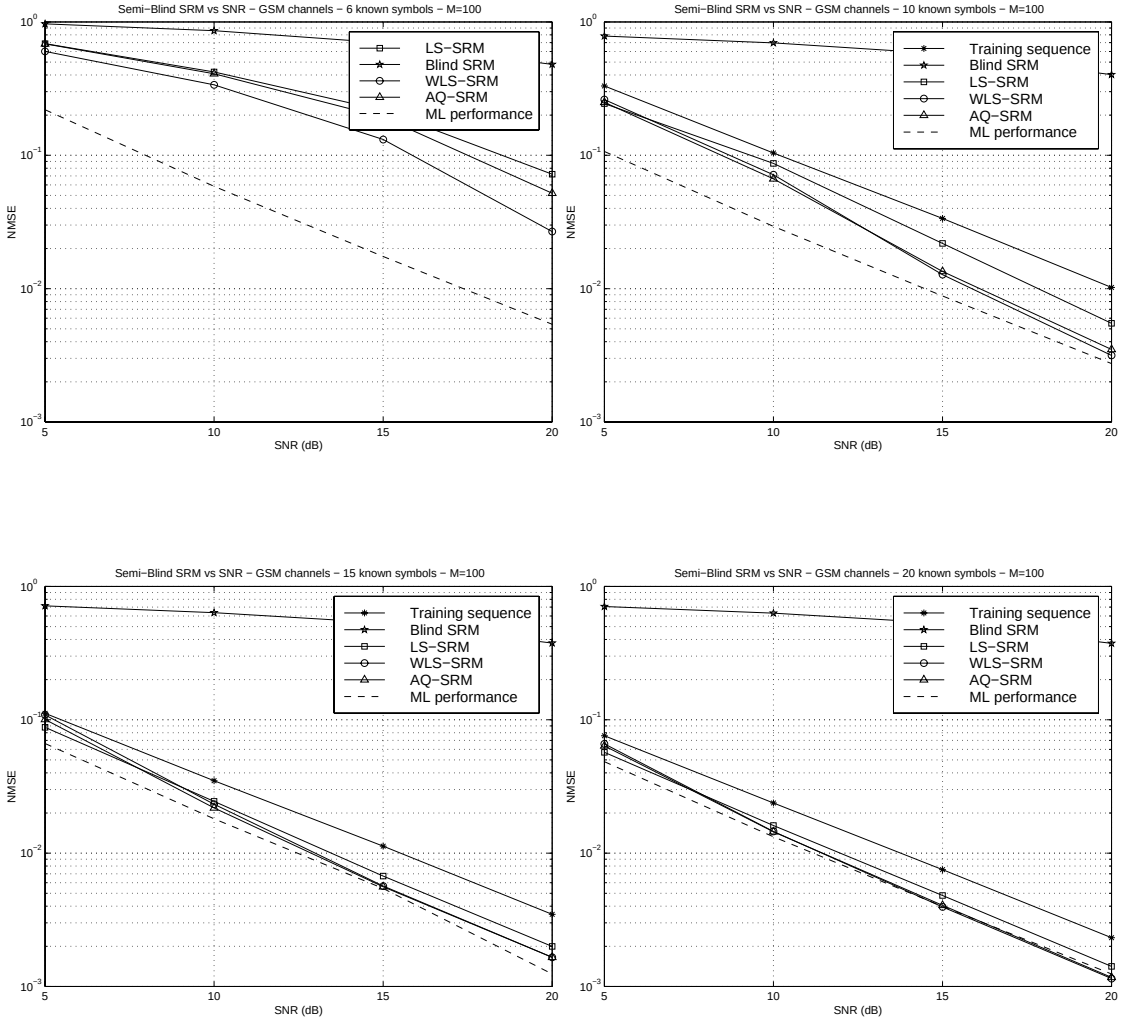



Figure 8.12: Performance of semi–blind SRM w.r.t the SNR.

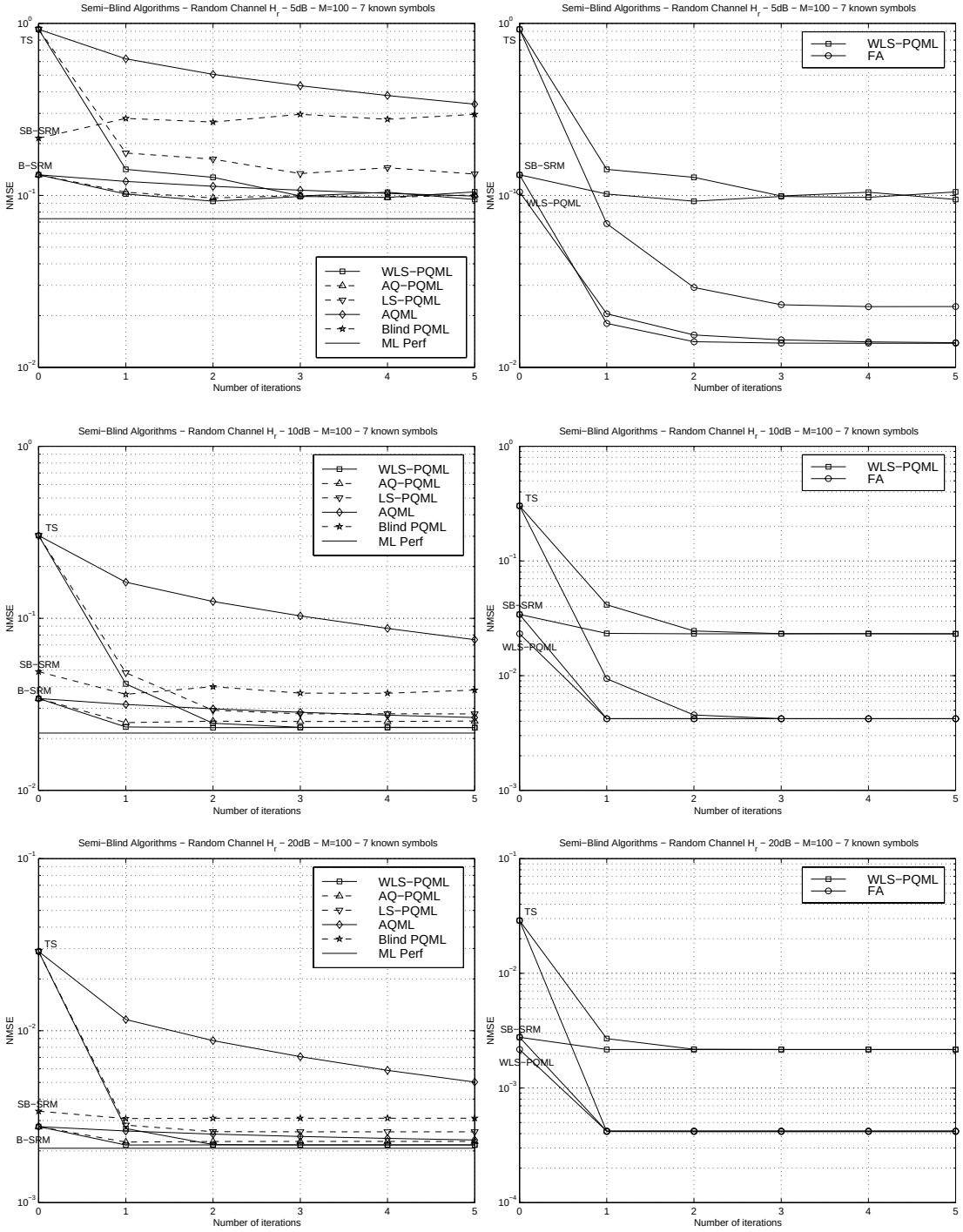



Figure 8.13: Semi–Blind Algorithms for a Random Channel of length 4 and a Burst Length of 100; 9 known symbols.

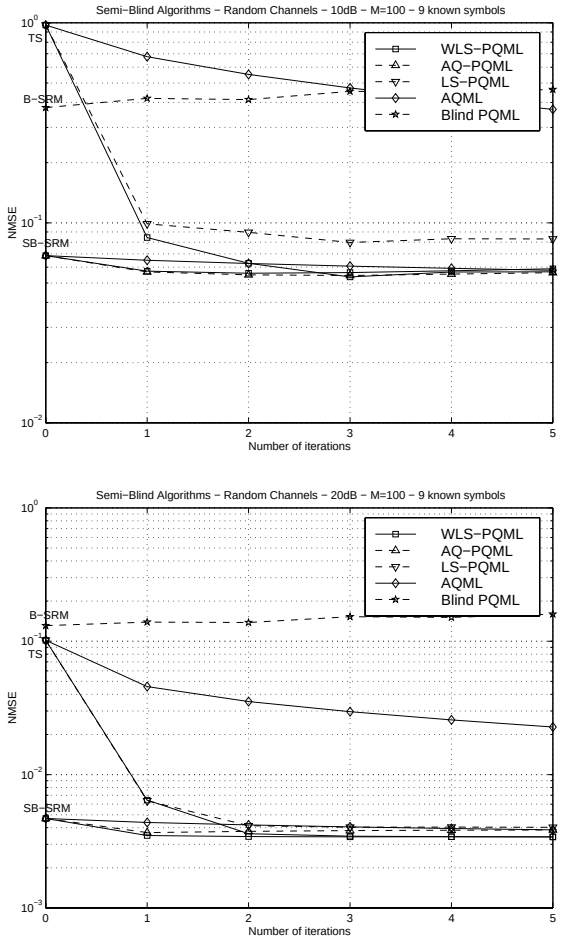



Figure 8.14: Semi–Blind Algorithms for Random Channels of length 5 and a Burst Length of 100: 1000 Monte–Carlo Runs of the channels, the input symbols and noise; 9 known symbols.

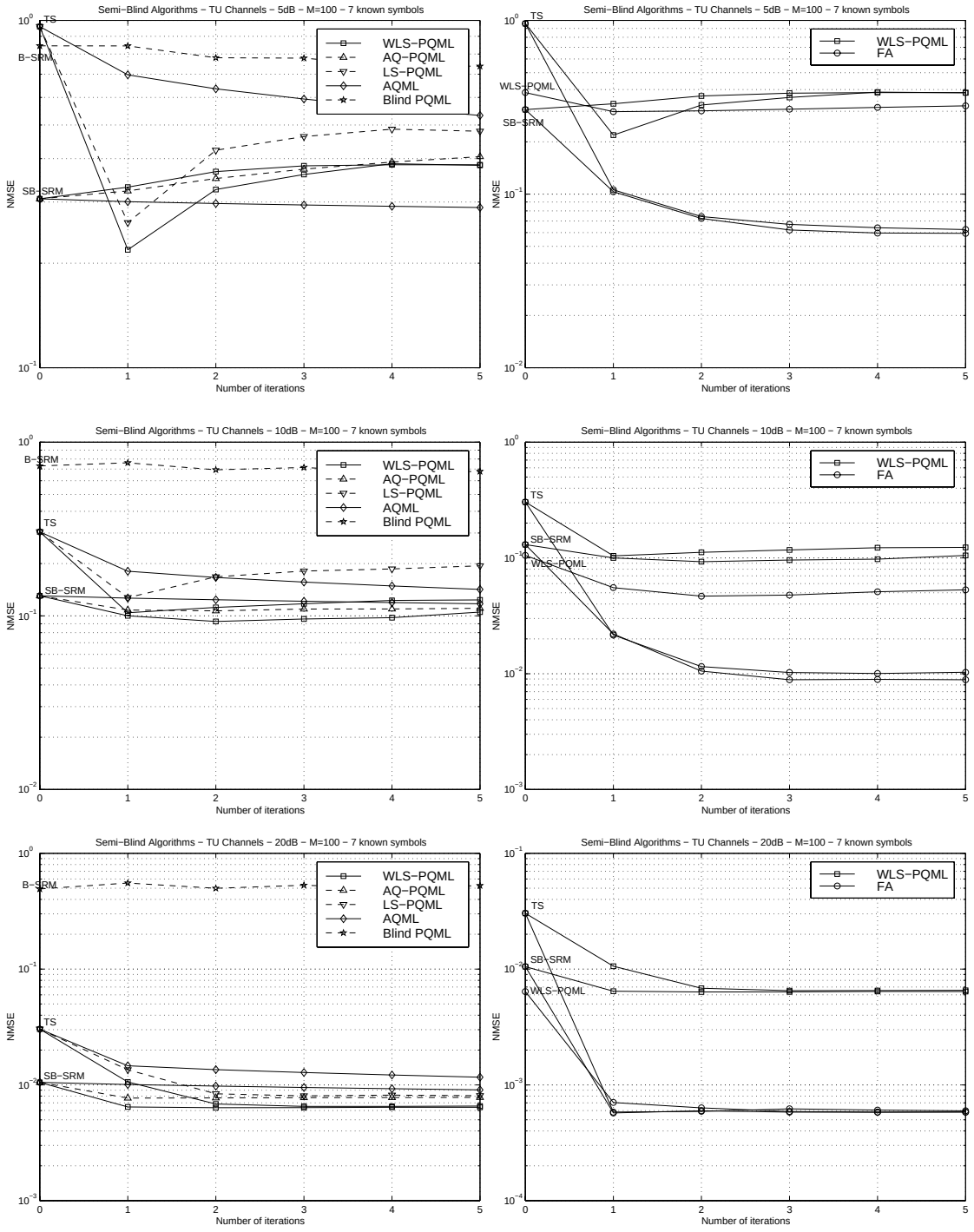



Figure 8.15: Semi–Blind Algorithms for GSM Channel ("Typically Urban") of length 4 and a Burst Length of 100: 1000 Monte–Carlo Runs of the channels, the input symbols and noise; 7 known symbols.

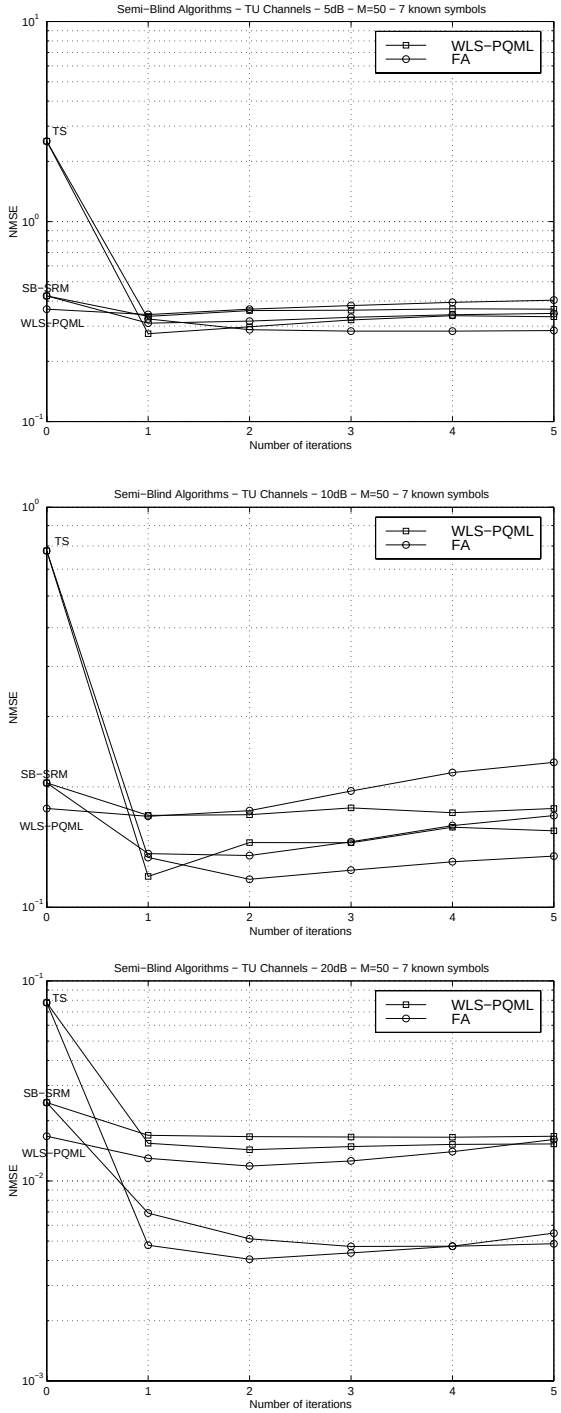



Figure 8.16: Semi–Blind Algorithms for GSM Channel ("Typically Urban") of length 4 and a Burst Length of 50: 1000 Monte–Carlo Runs of the channels, the input symbols and noise; 7 known symbols.

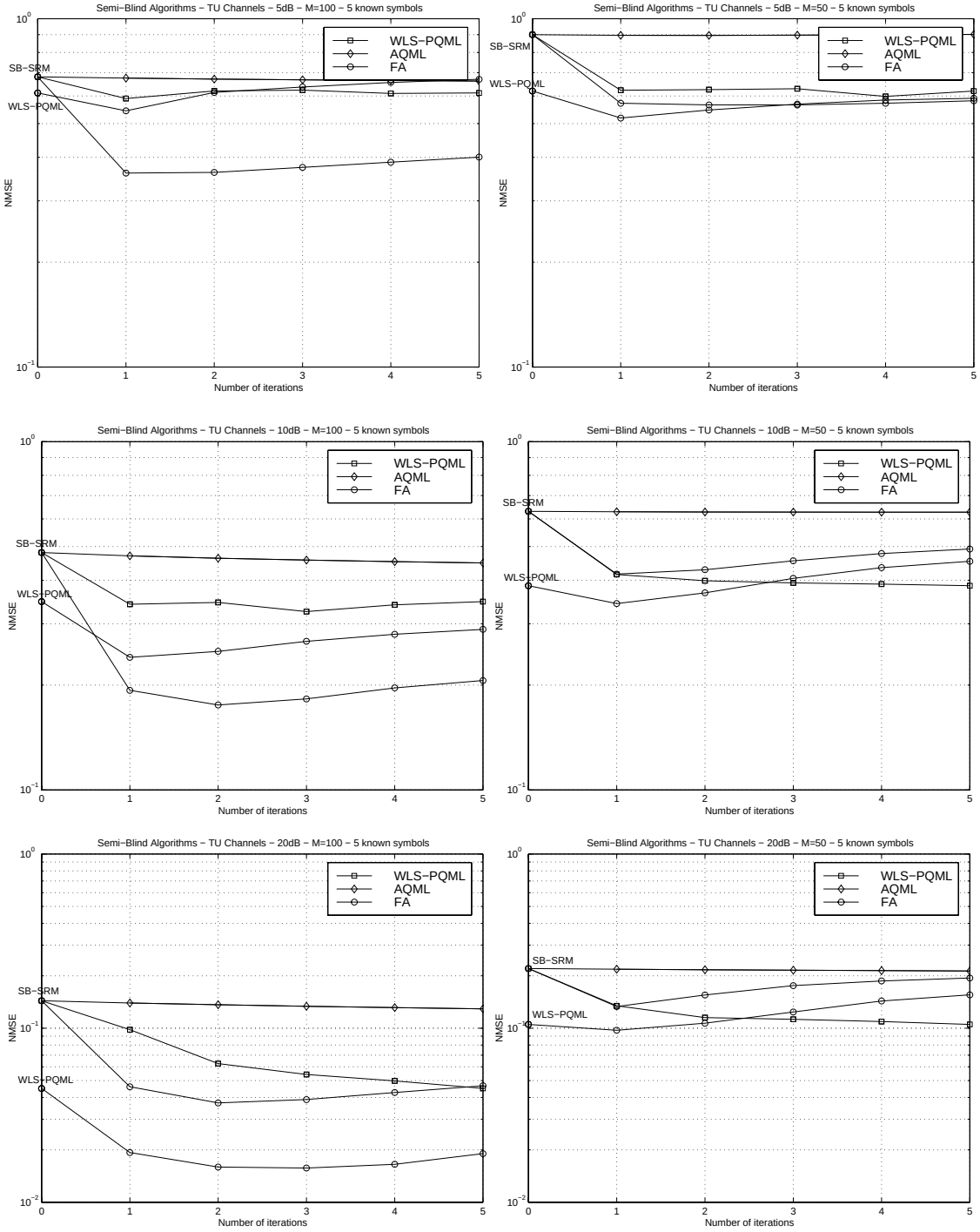



Figure 8.17: Semi–Blind Algorithms for GSM Channels ("Typically Urban") of length 4 and a Burst Length of 100 and 50: 1000 Monte–Carlo Runs of the channels, the input symbols and noise; 5 known symbols.

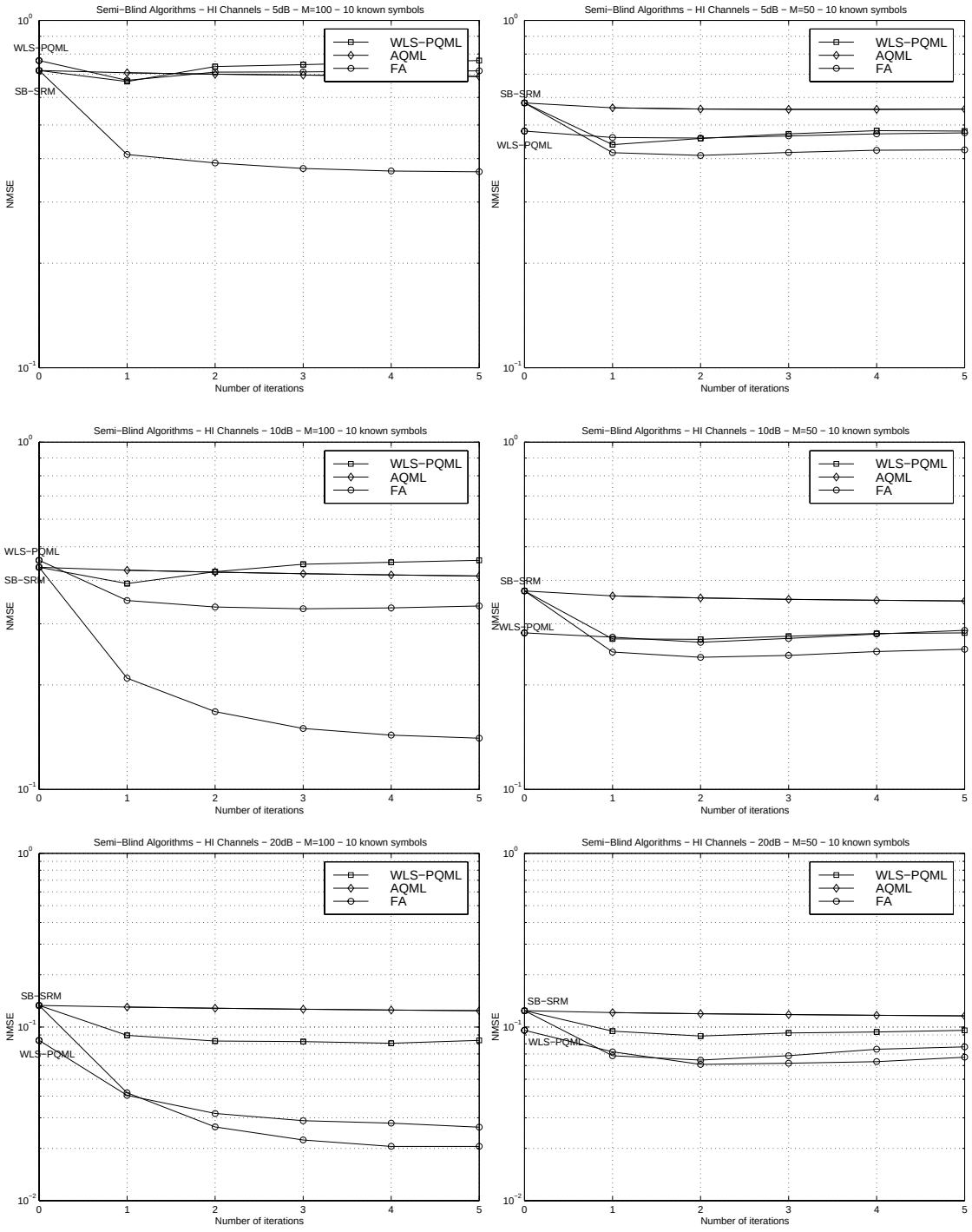



Figure 8.18: Semi–Blind Algorithms for GSM Channels ("Hilly Terrain" (HI)) of length 7 and a Burst Length of 100 and 50: 1000 Monte–Carlo Runs of the channels, the input symbols and noise; 10 known symbols.

Chapter 9

# BLIND AND SEMI–BLIND GAUSSIAN MAXIMUM–LIKELIHOOD

*In this chapter, we show that Gaussian ML (GML) can be seen as a form of covariance matching method and compare its performance to the classical covariance matching method based on weighted least-squares: both methods are equivalent for an asymptotic number of data and asymptotic size of covariance matrix. We use the scoring algorithm to solve GML and compare it to the blind and semi–blind deterministic methods of the previous chapter. Furthermore, we derive two fast algorithms which are approximations of the scoring method and the steepest descent algorithm, based on frequency domain approximation of the gradient of GML and the FIM. As for semi–blind deterministic methods, we give an example, based on covariance matching, to suggest how Gaussian semi–blind criteria could be built.*

## 9.1 Comparison of GML with the Covariance Matching Method

We compare GML to the Optimally weighted Covariance Matching (OCM) method [37] which is commonly said to be the most powerful method based on the second–order moments of the data. In DOA estimation, it is well-known that the equivalent of OCM and GML have asymptotically the same performance. Here, we are in a temporal context which differ from the DOA spatial context, and the performance equivalence problem is somewhat more difficult to handle. A noticeable difference between the DOA problem and our problem is that the optimal performance of OCM is attained when the length of the correlation sequence considered is infinite. Only the case of a complex channel will be treated in this section.

### 9.1.1 GML as a Covariance Matching Method

By the law of large numbers, the GML criterion:

$$\min_{\theta=(h,\sigma_v^2)} \left\{ \ln(\det C_{YY}(\theta)) + \mathrm{tr}\left\{ C_{YY}^{-1}(\theta) \boldsymbol{Y}\boldsymbol{Y}^H \right\} \right\} \tag{9.1}$$

is equivalent to:

$$\min_{\theta=(h,\sigma_v^2)} \left\{ \ln(\det C_{YY}(\theta)) + \mathrm{tr}\left\{ C_{YY}^{-1}(\theta) \mathrm{E}\left( \boldsymbol{Y}\boldsymbol{Y}^H \right) \right\} \right\} . \tag{9.2}$$

Then, asymptotically, $\boldsymbol{Y}\boldsymbol{Y}^H$ can be considered in the criterion as an estimate of $C_{YY}$: $\widehat{C}_{YY} = \boldsymbol{Y}\boldsymbol{Y}^H$. The GML criterion can then also be written as:

$$\min_{\theta=(h,\sigma_v^2)} \left\{ \ln(\det C_{YY}(\theta)) + \mathrm{tr}\left\{ C_{YY}^{-1}(\theta) \widehat{C}_{YY} \right\} \right\} . \tag{9.3}$$

Equation (9.3) looks then like a criterion matching $C_{YY}(\theta)$ to its estimate $\widehat{C}_{YY} = \boldsymbol{Y}\boldsymbol{Y}^H$, and then can be seen as a form of covariance matching.

### 9.1.2 Covariance Matching Method

The covariance matching method proceeds to a weighted least-squares fit between:

- the model of the second order statistics of the received signal:

$$\mathcal{R}(\theta) = [r_{yy}(0),\ r_{yy}(1),\ \cdots,\ r_{yy}(L-1)]$$

$$\text{where} \begin{cases} r_{yy}(i) = \displaystyle\sum_k \sigma_a^2 \boldsymbol{h}(k)\boldsymbol{h}^H(k+i) + \delta_{i0}\sigma_v^2 I \\ L \text{ is the length of the correlation sequence considered} \end{cases} \tag{9.4}$$

- their sample estimate built from the data:

$$\widehat{\mathcal{R}} = [\hat{r}_{yy}(0),\ \hat{r}_{yy}(1),\ \cdots,\ \hat{r}_{yy}(L-1)]$$

$$\text{where} \begin{cases} \hat{r}_{yy}(i) = \dfrac{1}{M-i} \displaystyle\sum_{k} \boldsymbol{y}(k)\boldsymbol{y}^H(k+i) \\ \boldsymbol{y}(k) \text{ is the output of the multichannel at time } k \\ M \text{ is the number of multichannel outputs available} \end{cases} \tag{9.5}$$

As for all Gaussian methods, the parameter estimate is $\theta = \left[h^T \ \ \sigma_v^2\right]^T$. Let $r_R(\theta)$ and $\hat{r}_R$ be two vectors containing the real and imaginary parts of the elements of $\mathcal{R}(\theta)$ and $\widehat{\mathcal{R}}$ resp. (they contain only the real part of $r_{yy}(0)$ which is real and of $\widetilde{r}_{yy}(0)$). The covariance matrix criterion may be written as:

$$\min_{\theta} \left(\hat{r}_R - r_R(\theta)\right)^H \mathcal{W}_R \left(\hat{r}_R - r_R(\theta)\right) \tag{9.6}$$

where $\mathcal{W}_R$ is a weighting matrix. The optimal weighting matrix is:

$$\mathcal{W}_R^o = \left(\mathrm{E}\left[\hat{r}_R - r_R(\theta^o)\right]\left[\hat{r}_R - r_R(\theta^o)\right]^H\right)^{-1} . \tag{9.7}$$

$\theta^o$ is the true parameter value: in practice, $\theta^o$ is replaced by a consistent estimate, which does not change the asymptotic performance of the criterion.

Which elements should be considered in $\tilde{r}_R(\theta) = \hat{r}_R - r_R(\theta)$? The authors of [37, 38] consider only the (non-redundant) non zero coefficients and claim that they are sufficient to get the optimal performance. This is not true however as stated in [39]. The optimal performance is obtained when the number of covariance lags involved tends to $\infty$. Asymptotically, OCM corresponds then to the best method exploiting the second order moments of the data and is equivalent, from a performance point of view, to GML.

### 9.1.3 Alternative Formulation

In order to give a closed form and simple expression for $\mathcal{W}_R^o$ as well as for the performance of the OCM, it is convenient to express the covariance matching method as a fit between the model of the covariance matrix of the received signal:

$$R_L(\theta) = \sigma_a^2 \mathcal{T}_L(h)\mathcal{T}_L^H(h) + \sigma_v^2 I \tag{9.8}$$

and its sample estimate:

$$\widehat{R}_L = \frac{1}{M} \sum_{k=0}^{M} \boldsymbol{Y}_L(k)\boldsymbol{Y}_L^H(k) . \tag{9.9}$$

Let the vectors $r(\theta)$ and $\hat{r}$ be defined as:

$$r(\theta) = \text{vec}\{R_L(\theta)\} \quad \text{and} \quad \hat{r} = \text{vec}\{\widehat{R}\}\,. \tag{9.10}$$

OCM may be written as:

$$\min_{h,\sigma_v^2} \left(\hat{r} - r(\theta)\right)^H \mathcal{W}^{o+} \left(\hat{r} - r(\theta)\right)\,. \tag{9.11}$$

This criterion can be proved to be equivalent to the matching criterion on the real and imaginary parts of $r(\theta)$, basically because $r(\theta)$ contains one element and its conjugate. The optimal weighting matrix is singular because $r(\theta)$ contains redundant elements: the choice of the pseudo–inverse of the weighting matrix leads to the best performance.

Formulation (9.11) is also equivalent to matching only the non redundant elements of $r(\theta)$, *i.e.* the elements of the first block column and block line of the matrices $R_L(\theta)$: vec$\{\underline{r}_o\}$, vec$\{\mathcal{R}_1^*\}$, vec$\{\mathcal{R}_1\}$ as shown in figure 9.1. Note that this is true because we weight the criterion optimally: for a non–weighted formulation, this is not valid.

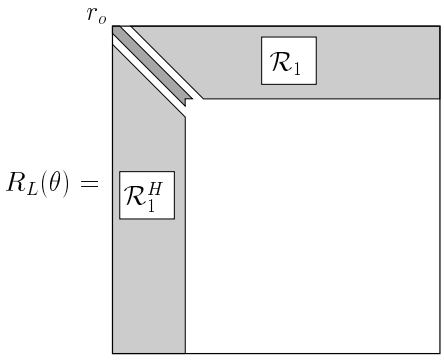



Figure 9.1: Elements selected in the covariance matrix to build the CM criterion.

The performance of the optimally weighted covariance matching method are:

$$C_{\tilde{\theta}_R\tilde{\theta}_R} = \left[\left(\frac{\partial r^H(\theta)}{\partial \theta_R}\right) \mathcal{W}^{o+} \left(\frac{\partial r^T(\theta)}{\partial \theta_R}\right)^T\right]^+ \tag{9.12}$$

where the performance is computed under our usual phase constraint (4.7). In appendix A, we study in more detail the performance of OCM, and especially the expression for the optimal weighting matrix which is:

$$\mathcal{W}^o = \frac{1}{M}\left[\sum_{u=-(L+N-1)}^{L+N-1} R_L^T(u) \otimes R_L(-u) - \text{vec}\left\{\sigma_a^2 \mathcal{T}(h)\mathcal{T}^H(h)\right\} \text{vec}^H\left\{\sigma_a^2 \mathcal{T}(h)\mathcal{T}^H(h)\right\}\right] \tag{9.13}$$

where $R_L(u) = \text{E}\left(\boldsymbol{Y}_L(k)\boldsymbol{Y}_L^H(k+u)\right)$.

### 9.1.4 Simplified Covariance Matching Criterion

Consider the following simplification of the weighting matrix:

$$\mathcal{W}^s = \frac{1}{M}\left[R_L^T \otimes R_L\right] \tag{9.14}$$

where we keep only the central term of the sum in (9.13). This weighting matrix corresponds to the expression of the optimal weighting matrix in DOA. The weighted CM criterion becomes:

$$\widetilde{r}^H(\theta)\left\{R_L^{-T} \otimes R_L^{-1}\right\}\widetilde{r}(\theta) \quad \Leftrightarrow \mathrm{tr}\left\{\widetilde{R_L}(\theta)R_L^{-1}\widetilde{R_L}(\theta)R_L^{-1}\right\} \tag{9.15}$$

which is a well–known form of OCM in DOA. We will see in simulations that this weighted CM criterion has performance very close to the optimal performance; its performance is:

$$\begin{aligned} C_{\tilde{\theta}_R\tilde{\theta}_R} = &\left[\left(\frac{\partial r^H(\theta)}{\partial \theta_R}\right)\mathcal{W}^{s-1}\left(\frac{\partial r^T(\theta)}{\partial \theta_R}\right)^T\right]^+ \left(\frac{\partial r^H(\theta)}{\partial \theta_R}\right)\mathcal{W}^{s-1}\mathcal{W}^o\mathcal{W}^{s-1}\left(\frac{\partial r^T(\theta)}{\partial \theta_R}\right)^T \\ &\left[\left(\frac{\partial r^H(\theta)}{\partial \theta_R}\right)\mathcal{W}^{s-1}\left(\frac{\partial r^T(\theta)}{\partial \theta_R}\right)^T\right]^+ . \end{aligned} \tag{9.16}$$

This form of covariance matching offers interesting perspectives as it can be solved by the scoring method in a less complex way than GML (see section 9.3).

### 9.1.5 Numerical Evaluations

A proof that GML and OCM have the same asymptotic performance (using frequency domain expressions for the error covariance matrices) is for the moment still under investigation. And we will just illustrate the equivalence in performance by numerical evaluations.

In figure 9.2 (left), we show the performance of GML and OCM for channel estimation only ($\sigma_v^2$ is assumed known), *i.e.* $\|h_R^o - \hat{h}_R\|^2/\|h_R^o\|^2$, when the sample matrix is based on $M$ and $M-L$ data samples (the true performance of OCM is in fact between the two curves); as the burst length is of 100, we have not reach completely asymptotic conditions, which explains why the curves are distinct. The multichannel has 2 subchannels of length 4: when the noise is known, CM (and the Gaussian methods in general) is defined for a burst length of at least $\underline{M}$ for a multichannel and $N$ for a monochannel. Here $\underline{M} = N - 1 = 3$. In this figure, it can be noticed that the performance of OCM gets better as more and more correlation coefficients are included. A quasi steady–state is rapidly attained, but considering only the $N$ first moments is not optimal.

In figure 9.2 (right), we make a comparison between CM with the optimal weighting, the approximated weighting in (9.14), and the non–weighted CM (the form (1) is based on the non–redundant elements of $R_L(\theta)$ and the form (2) on all the elements). It can be noticed that the approximated weighting gives performance very closed to OCM. There is however a certain gap between non–weighted CM and OCM.

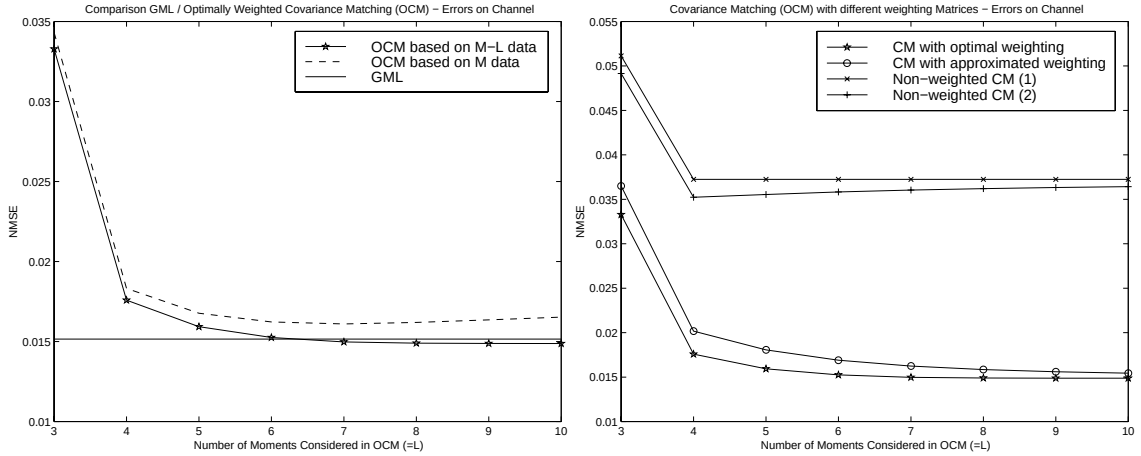



Figure 9.2: Comparison between GML and the optimally weighted covariance matching method w.r.t. the number of correlation coefficients considered in the OCM method (left) and comparison between CM methods differently weighted (right).

## 9.2 Method of Scoring

The purpose of this section is to compare by simulations the performance of blind, semi–blind DML and GML methods, and see the benefit of GML methods.

We propose to solve the GML criterion by the method of scoring. This method consists in an approximation of the Newton-Raphson algorithm which finds an estimate $\theta^{(i)}$ at iteration $i$ from $\theta^{(i-1)}$, the estimate at iteration $i-1$, as:

$$\theta^{(i)} = \theta^{(i-1)} - \left[ \frac{\partial}{\partial \theta^*} \left( \frac{\partial \mathcal{F}(\theta)}{\partial \theta^*} \right)^H \Bigg|_{\theta^{(i-1)}} \right]^{-1} \frac{\partial \mathcal{F}(\theta)}{\partial \theta^*} \Bigg|_{\theta^{(i-1)}} \tag{9.17}$$

where $\mathcal{F}(\theta)$ is the cost function and $\theta$ contains the parameters to estimate. The method of scoring approximates the Hessian by its expected value, which is here the Gaussian Fisher Information Matrix (FIM). This approximation is justified by the law of large numbers as the number of data is generally large. In the semi-blind case, the number of known symbols being finite, this approximation is not valid anymore but it will turn out to work very well in our simulations. We did not choose to apply directly the Newton method: indeed, the Hessian contains four terms one of which is the opposite of the FIM, so it would have represented an increase in complexity.

We detail only the case of a complex channel. In the blind case, the FIM is singular, so formula (9.17) cannot be applied directly: we take the Moore-Penrose pseudo–inverse of the FIM. At iteration (i), it corresponds to the constraint $\hat{h}_R^{(i-1)\,T} \hat{h}_R^{(i)} = 0$, where $\hat{h}_R^{(i)}$ is the

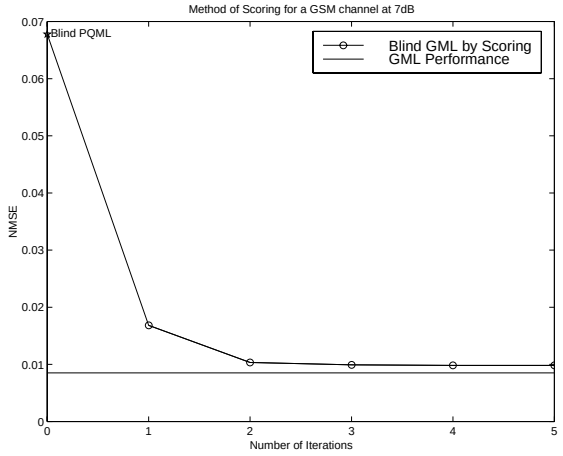



Figure 9.3: Blind GML optimized by the method of Scoring initialized by Blind PQML.

channel estimate at iteration $i$. When the algorithm converges correctly, asymptotically, this constraint is equivalent to our usual constraint $h_{S_2}^T h_R = 0$ (see 4.29).

The scoring algorithm is applied to the channel only, the noise variance is estimated apart: in the simulations, we estimate it by the SRM method (see Chapter 7, section 7.5.2). As pointed out in Chapter 5, at high SNR, the noise variance estimation is asymptotically decoupled from the channel estimation: at 10dB, this property is still verified, so the estimation quality of $\sigma_v^2$ is not really of importance.

In figure 9.3, we show the NMSE of the blind scoring algorithm initiated by blind PQML for a randomly chosen channel of length 4 and with 3 subchannels. The NMSE is averaged over 50 noise and input symbol realizations. Blind PQML gives an estimate of the channel up to a scale factor: in our simulations, we adjust the norm of the channel based on the covariance of a received signal sample and the phase factor by constraint (4.7). We notice the improvement brought by GML and performance closed to the theoretical blind GML performance.

In the course of this work, we became aware of [64],[84] in which a semi–blind Gaussian ML method and corresponding CRBs have been studied. The modeling of the training sequence information in [64] is inappropriate though: instead of the training sequence, the information considered is the training sequence times an unknown zero-mean unit-variance normal variable.

## 9.3 Method of Scoring for Covariance Matching

Element $(i,j)$ of the expected value of the Hessian of the cost function in (9.15) is:

$$2\text{tr}\left\{\frac{\partial R_L(\theta)}{\partial \theta_{R_i}} R_L^{-1} \left(\frac{\partial R_L(\theta)}{\partial \theta_{R_j}}\right)^H R_L^{-1}\right\} \tag{9.18}$$

which can also be expressed thanks to the quantities:

$$\mathcal{P}_{\theta\theta}(i,j) = \text{tr}\left\{\frac{\partial R_L(\theta)}{\partial \theta_i^*} R_L^{-1} \left(\frac{\partial R_L(\theta)}{\partial \theta_j^*}\right)^H R_L^{-1}\right\} \tag{9.19}$$

$$\mathcal{P}_{\theta\theta^*}(i,j) = \text{tr}\left\{\frac{\partial R_L(\theta)}{\partial \theta_i^*} R_L^{-1} \left(\frac{\partial R_L(\theta)}{\partial \theta_j^*}\right) R_L^{-1}\right\} . \tag{9.20}$$

The expected value of the Hessian is then 2 times the FIM. In the scoring algorithm, the difference with GML is that the FIM is of reduced size $L$ and that iterations are not done on each quantity (the matrix $R_L$ is the sample covariance matrix).

Furthermore, the gradient w.r.t. $\theta^*$ is:

$$-2\text{tr}\left\{\frac{\partial R_L(\theta)}{\partial \theta_i^*} R_L^{-1} \widetilde{R_L}(\theta) R_L^{-1}\right\} = -2\text{tr}\left\{R_L^{-1}\frac{\partial R_L(\theta)}{\partial \theta_i^*}\right\} + 2\text{tr}\left\{R_L^{-1}\frac{\partial R_L(\theta)}{\partial \theta_i^*} R_L^{-1} R(\theta)\right\} . \tag{9.21}$$

The gradient of the GML cost function w.r.t. $\theta^*$ is:

$$\text{tr}\left\{R_M^{-1}(\theta)\frac{\partial R_M(\theta)}{\partial \theta_i^*}\right\} - \text{tr}\left\{R_M^{-1}(\theta)\frac{\partial R_M(\theta)}{\partial \theta_i^*} R_M^{-1}(\theta)\boldsymbol{Y}\boldsymbol{Y}^H\right\} . \tag{9.22}$$

So we see that the gradient of CM is also related to the gradient of GML in which $\boldsymbol{Y}\boldsymbol{Y}^H$ plays the role of $R(\theta)$. We have not tested yet the scoring method on this form of CM.

## 9.4 Semi–Blind GML: Suboptimal Approaches

We consider here the case of grouped known symbols and as for DML, we determine suboptimal semi–blind GML criteria that keep the blind problem structure. Again, we consider 2 ways of splitting the data. The splitting is however not the same as for DML, where 2 observation vectors are uncorrelated if the corresponding noise vectors are uncorrelated. For GML, 2 observation vectors are uncorrelated if the corresponding noise and input symbol vectors are uncorrelated.

### 9.4.1 Least–Squared GML (LS–GML)

For Least–Squared GML (LS–GML), a proper way of splitting the data would be the following, as indicated in figure 9.4:

- $\boldsymbol{Y}_{TS} = \mathcal{T}_{TS}(h)A_{TS} + \boldsymbol{V}_{TS}$ groups all the observations containing only known symbols.
- $\boldsymbol{Y}_B = \mathcal{T}_B(h)A_B + \boldsymbol{V}_B$ groups all the observations containing only unknown symbols.

Semi–blind GML is applied to the vector:

$$\begin{bmatrix} \boldsymbol{Y}_{TS} \\ \boldsymbol{Y}_B \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} \mathcal{T}_{TS}(h)A_{TS} \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_v^2 I & 0 \\ 0 & \mathcal{T}_B(h)\mathcal{T}_B^H(h) + \sigma_v^2 I \end{bmatrix} \right) \tag{9.23}$$

to give the semi–blind criterion ($\sigma_v^2$ is assumed known):

$$\min_{h,\sigma_v^2} \left\{ \frac{1}{\sigma_v^2} \| \boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h)A_{TS} \|^2 + \ln\det C_{Y_B Y_B} + \boldsymbol{Y}_B^H C_{Y_B Y_B}^{-1} \boldsymbol{Y}_B \right\} . \tag{9.24}$$

The overlap zone (figure 9.4) containing at the same time known and unknown symbols is ignored in this first decomposition. We also tested criterion (9.24) when $\boldsymbol{Y}_B$ is replaced by $\boldsymbol{Y}'_B$ which contains $\boldsymbol{Y}_B$ and the overlap zone, and although the correlations between $\boldsymbol{Y}_{TS}$ and $\boldsymbol{Y}_B$ are neglected, this extended LS-GML criterion gives better performance than the LS-GML criterion based on the 2 uncorrelated observation vectors.

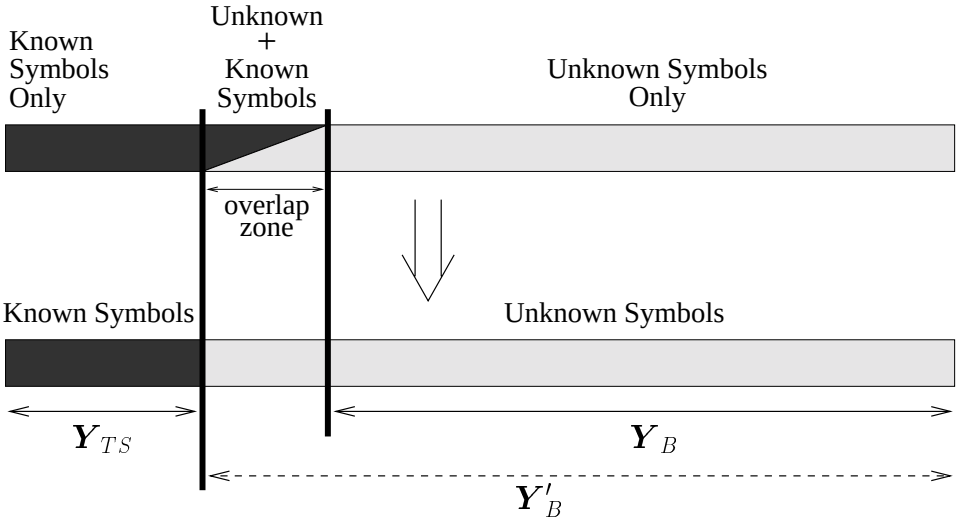



Figure 9.4: Output Burst: split of the data for LS–GML.

### 9.4.2 WLS–GML

An alternative splitting is as follows (figure 9.5):

- $\boldsymbol{Y}_{WLS} = \mathcal{T}_{WLS}(h)A_{WLS} + \boldsymbol{V}_{WLS}$ contains $\boldsymbol{Y}_{TS}$ and the overlap zone.
- $\boldsymbol{Y}_B = \mathcal{T}_B(h)A_B + \boldsymbol{V}_B$ groups all the observations containing only unknown symbols except the unknown symbols of the overlap zone

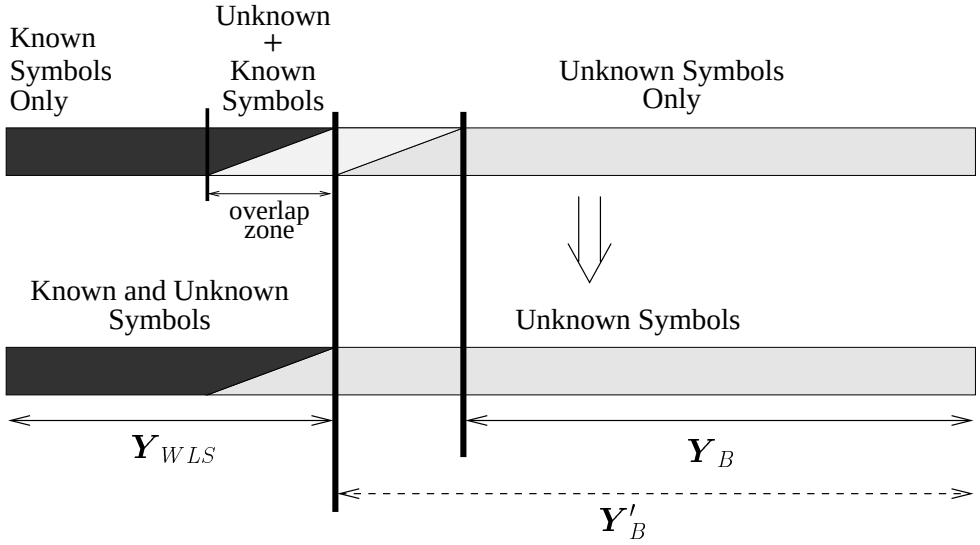



Figure 9.5: Output Burst: split of the data for WLS–GML.

Semi–blind GML is applied to the vector:

$$\begin{bmatrix} \boldsymbol{Y}_{WLS} \\ \boldsymbol{Y}_B \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} \mathcal{T}'_K(h)A_{TS} \\ 0 \end{bmatrix}, \begin{bmatrix} \mathcal{T}'_U(h)\mathcal{T}'^H_U(h) + \sigma_v^2 I & 0 \\ 0 & \mathcal{T}_B(h)\mathcal{T}_B^H(h) + \sigma_v^2 I \end{bmatrix} \right) . \quad (9.25)$$

As for LS–PQML, the GML part applied to $\boldsymbol{Y}_{WLS}$ is solved by weighted Least–Squares:

$$\min_{h,\sigma_v^2} \|\boldsymbol{Y}_{WLS} - \mathcal{T}'_K(h)A'_K\|^2_{C^{-1}_{Y_{WLS}Y_{WLS}}} + \ln\det C_{Y_BY_B} + \boldsymbol{Y}_B^H C^{-1}_{Y_BY_B}\boldsymbol{Y}_B . \quad (9.26)$$

Again, observations samples are ignored in this formulation. When $\boldsymbol{Y}_B$ is extended to $\boldsymbol{Y}'_B$ (see figure 9.5) and correlations between input symbols are ignored, the resulting WLS–GML gives better performance. An AQ–GML could also be built here.

## 9.5 Simulations for Semi–Blind GML

In figure 9.6, performance of optimal and suboptimal semi–blind GML initialized by WLS-PQML are presented for an SNR of 10dB. We show the NMSE for the channel averaged over 50 Monte-Carlo runs of the noise and the input symbols. We tested:

1. the optimal scoring (called semi–blind scoring in the figures),

2. the extended WLS–GML based on $\boldsymbol{Y}'_B$ (figure 9.4).

3. the extended LS-GML based on $\boldsymbol{Y}'_B$ (figure 9.5).

4. LS-GML based on uncorrelated vectors (ULS–GML in the curves).

For a randomly chosen channel ($m = 3$, $N = 4$), we see that GML does not offer much improvement. For a GSM channel (TU model), however, an improvement is more visible. We also tested a channel with a common zero: we also have an improvement with a slight anomaly at the first iteration. Furthermore, WLS–GML appears to be the best suboptimal semi–blind GML method and the extended versions appear better.

## 9.6 Two fast Solutions to Solve Blind GML

In this section, we shall focus on fast solutions to solve GML also presented in [85]. We remain here in a single–user context, however one of the reasons for examining, in a closer way, GML is its extension to the multiuser case where GML is of particular interest. Apart, from performance advantages, one of the great properties of GML, like all methods based on the second–order statistics of the data, is their robustness to channel length overestimation. Blind deterministic methods, like subspace fitting or DML, fail when the channel length has been overestimated: each channel length for each users has to be tested. On the contrary, the Gaussian approach can be shown not to suffer from this problem (as has been shown for linear prediction methods [33]). In multiuser communications, GML has also another advantage: deterministic methods can only identify the channel apart from a triangular dynamical multiplicative factor, whereas Gaussian methods can identify the channel up to a unitary static factor.

As initialization we use the Schur method briefly described in Chapter 1 and detailed in [35, 36] which is a low computational multi–user method.

### 9.6.1 Approximated Scoring Method

The problem parametrized in $h_R = \left[\mathrm{Re}(h^T)\ \ \mathrm{Im}(h^T)\right]^T$ can be equivalently parametrized in $h_C = \left[h^T\ h^{*T}\right]^T$. The FIM for $h_C$ is:

$$J_{h_C h_C} = \begin{bmatrix} J_{hh} & J_{hh^*} \\ J^*_{hh^*} & J^*_{hh} \end{bmatrix} \tag{9.27}$$

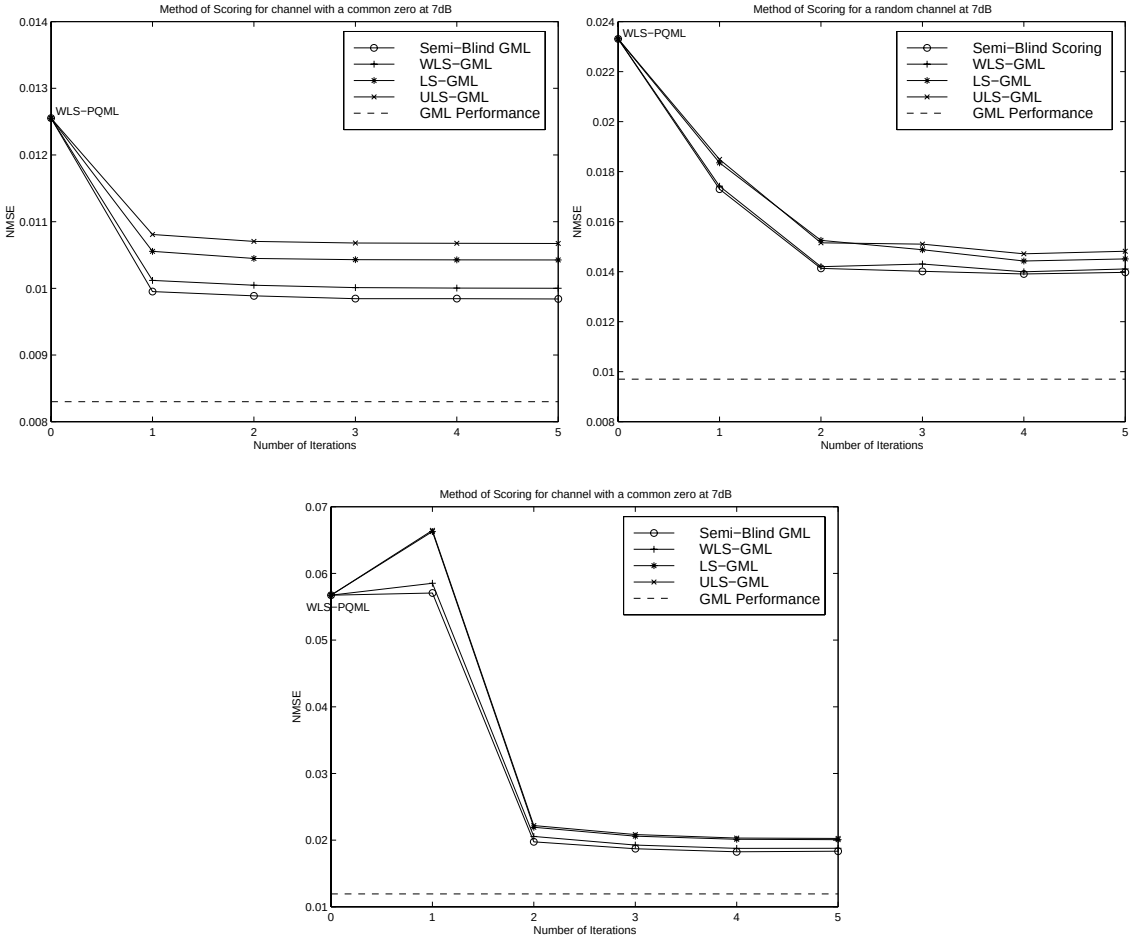



Figure 9.6: Semi–Blind GML by the method of Scoring initialized by WLS-PQML.

and the gradient is:

$$D_{h_C} = \begin{bmatrix} \dfrac{\partial \mathcal{F}(\theta)}{\partial h^*} \\ \dfrac{\partial \mathcal{F}(\theta)}{\partial h} \end{bmatrix} = \begin{bmatrix} D_h \\ D_h^* \end{bmatrix} . \tag{9.28}$$

Let us recall the expressions of $J_{hh}$ and $J_{hh^*}$:

$$J_{hh}(i,j) = \mathrm{tr}\left\{ C_{YY}^{-1} \frac{\partial C_{YY}}{\partial h_i^*} C_{YY}^{-1} \left( \frac{\partial C_{YY}}{\partial h_j^*} \right)^H \right\} \tag{9.29}$$

$$J_{hh^*}(i,j) = \mathrm{tr}\left\{ C_{YY}^{-1} \frac{\partial C_{YY}}{\partial h_i^*} C_{YY}^{-1} \frac{\partial C_{YY}}{\partial h_j^*} \right\} . \tag{9.30}$$

Our fast implementation of the method of scoring is based on a frequency–domain asymptotic (in the number of data) approximation of the FIM and of the gradient of the cost function.

**Approximation for $J_{hh}$** Let us consider first the term $J_{hh}$: it can be asymptotically approximated as [86]:

$$J_{hh}(i,j) = \frac{M}{2\pi j} \oint \mathrm{tr}\left\{ S_{yy}^{-1}(z) \frac{\partial S_{yy}(z)}{\partial h_i^*} S_{yy}^{-1}(z) \left( \frac{\partial S_{yy}(z)}{\partial h_j^*} \right)^H \right\} \frac{dz}{z} \tag{9.31}$$

where $S_{yy}(z) = \mathbf{h}(z)\mathbf{h}^\dagger(z) + \sigma_v^2 I$ is the spectral density of the received signal. From this expression, we see that $J_{hh}(i,j)$ can be approximated as a block Toeplitz matrix (which is also symmetric). The block $(1, j_b)$ of its first line is the coefficient of order $1 - j_b$ of the filter:

$$J_{hh}(1,j_b) = \frac{M}{2\pi j} \oint \frac{\mathbf{h}^\dagger(z)\mathbf{h}(z) \left[ \left( \mathbf{h}^\dagger(z)\mathbf{h}(z) + \sigma_v^2 \right) I - \mathbf{h}(z)\mathbf{h}^\dagger(z) \right]}{\sigma_v^2 \left( \mathbf{h}^\dagger(z)\mathbf{h}(z) + \sigma_v^2 \right)^2} z^{1-j_b} \frac{dz}{z} . \tag{9.32}$$

Using the Gohberg-Semencul formula: $\left( \mathbf{h}^\dagger(z)\mathbf{h}(z) + \sigma_v^2 \right)^{-2} = p(z)p^\dagger(z)/\tilde{\sigma}_p^2$, where $p(z)$ is the linear prediction filter associated to $\left( \mathbf{h}^\dagger(z)\mathbf{h}(z) + \sigma_v^2 \right)^2$, and $\tilde{\sigma}_p^2$, the prediction error.

$$J_{hh}(1,j_b) = \frac{M}{2\pi j} \oint \frac{1}{\tilde{\sigma}_p^2} p(z)p^\dagger(z)\mathbf{h}^\dagger(z)\mathbf{h}(z) \left[ \left( \mathbf{h}^\dagger(z)\mathbf{h}(z) + \sigma_v^2 \right) I - \mathbf{h}(z)\mathbf{h}^\dagger(z) \right] z^{1-j_b} \frac{dz}{z} . \tag{9.33}$$

$$J_{hh}(1,j_b) = \frac{1}{\tilde{\sigma}_p^2} \left[ p * p^\dagger * \mathbf{h}^\dagger * \mathbf{h} * \left[ \left( \mathbf{h}^\dagger * \mathbf{h} + \sigma_v^2 \right) I - \mathbf{h} * \mathbf{h}^\dagger \right] \right]_{1-j_b} . \tag{9.34}$$

$J_{hh}(1,j_b)$ is computed by truncating $p(z)$ (a truncation $\alpha N$, where $\alpha$ is 3 or 4 is in general sufficient), and involves then only FIR filtering operations. So computing the elements of $J_{hh}$ is of order $N$.

**Approximation for $J_{hh^*}$** The same kind of treatment holds for $J_{hh^*}$ which can be approximated as a block Hankel matrix. The block $(1, j_b)$ of its first line is the coefficient of order $j_b - 1$ of the filter:

$$J_{hh^*}(1, j_b) = \left[ \frac{\mathbf{h}(z)\mathbf{h}^T(z)}{\left(\mathbf{h}^\dagger(z)\mathbf{h}(z) + \sigma_v^2\right)^2} \right]_{j_b - 1} = \left[ \frac{1}{\sigma_p^2} p * \mathbf{h} * \mathbf{h}^T * p^\dagger \right]_{j_b - 1} . \tag{9.35}$$

The block $(i_b, 1)$ of the last column is its coefficient of order $N - 2 + i_b$.

**Approximation of the gradient** Using the band property of $C_{YY}$, a fast computation of the output of $C_{YY}^{-1}\boldsymbol{Y}$ is of order NM. $\frac{\partial C_{YY}}{\partial h_i^*} = \mathcal{T}(h)\mathcal{T}^H(\frac{\partial h}{\partial h_i})$ and both terms being banded, the computation of the second term is of order $MN$.

Using a frequency domain approximation, the block $i_b$ of the first term of $D_h$ can be approximated as the element $i_b - 1$ of the filter:

$$D_h(i_b) = \left[ \frac{\mathbf{h}(z)}{\mathbf{h}^\dagger(z)\mathbf{h}(z) + \sigma_v^2} \right]_{i_b - 1} = \frac{1}{\sigma_p^2} \left[ p * p^\dagger * \mathbf{h} \right]_{i_b - 1} . \tag{9.36}$$

$D_{h^*}$ can be computed using $D_{h^*} = (D_h)^*$.

At each step of the algorithm, equation (9.17) is solved using the Toeplitz and Hankel property of $J_{hh}$ and $J_{hh^*}$, which gives a complexity of order $N^2$.

### 9.6.2 Regularization of the FIM

The approximated FIM is nonsingular: it has an eigenvalue (negative or positive) closed to 0. The inverse of the approximated FIM could then be directly taken in the scoring algorithm: this solution makes the algorithm diverge however, as the step in the direction of the associated eigenvector is too large.

We use the Levenberg Marquardt method by regularizing the FIM by a factor $\lambda I$. Unfortunately, as seen in the simulations, the regularized approximated scoring algorithm looses the high convergence speed of the true scoring method, and in fact a simple steepest–descent algorithm:

$$\theta^{(i)} = \theta^{(i-1)} - \mu \left. \frac{\partial \mathcal{F}(\theta)}{\partial \theta^*} \right|_{\theta^{(i-1)}} \tag{9.37}$$

gives similar performance.

### 9.6.3 Simulations

We plot the averaged the NMSE over 50 noise and input symbol realizations for a randomly chosen channel ($N = 4$, $m = 3$): see figure 9.7. For the approximated scoring algorithm, the

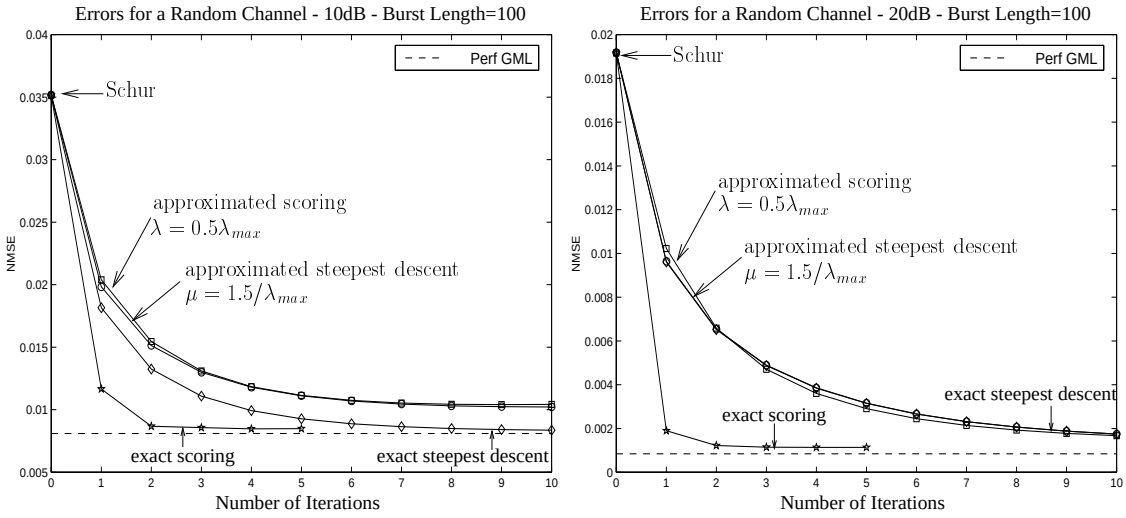



Figure 9.7: Approximated scoring algorithm for 10dB and 20dB.

best regularization factor is $\lambda = 0.5\lambda_{max}(FIM)$, ($\lambda_{max}(.)$ designate the larger eigenvalue) for the steepest descent, it is $\lambda = 1.5/\lambda_{max}(FIM)$. We notice that at 10dB, the approximation in the gradient makes the algorithms loose in performance.

## 9.7 Semi–Blind Criteria as a Combination of a Blind and a TS Based Criteria

As for the semi–blind deterministic model, we can build Gaussian semi–blind criteria by doing a weighted combination of a blind and a training sequence based algorithm (again, when it is possible). Let us take the example of CM (9.15). We consider the decomposition of figure 9.5 and choose the approximated weighted in (9.15). The semi–blind criterion is:

$$\begin{aligned} &\min_{h,\sigma_v^2} \left\{ \widetilde{r}^H \mathcal{W}^{s-1} \widetilde{r} + \frac{1}{\sigma_v^2} \| \boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_{TS} \|^2 \right\} \quad \Leftrightarrow \\ &\min_{h,\sigma_v^2} \left\{ \frac{1}{M_U} \widetilde{r}^H \left\{ R_L^{-T} \otimes R_L^{-1} \right\} \widetilde{r} + \frac{1}{\sigma_v^2} \| \boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_{TS} \|^2 \right\} . \end{aligned} \tag{9.38}$$

## 9.8 Conclusion

GML was compared to the optimally weighted covariance matching method, which was shown, through simulations, to have the same performance asymptotically (in the number of data but also in the number of moments considered). A covariance matching criterion built from an approximation of the weighting matrix was also proposed: this CM criterion offers

some computationally low solutions that are still under investigation. The method of scoring was used to solve GML and performs very well: we are now examining the computational load required to apply this method. We have also developed a fast implementation of the scoring algorithm and the steepest descent algorithm to solve GML. The fast GML should be next generalized to the multi–user case.

# A Asymptotic Performance Analysis of GML

The main purpose of this appendix is to compute the optimal weighting matrix $\mathcal{W}^o$. OCM gives consistent estimate, which we do not prove here and its performance can be expressed w.r.t. $\theta_C = [\theta^T \;\; \theta^{*T}]^T$:

$$\left(\frac{\partial r^H(\theta)}{\partial \theta_C}\right) \mathcal{W}^{o\,-1} \left(\frac{\partial r^H(\theta)}{\partial \theta_C}\right)^H \tag{9.39}$$

The error covariance matrix for $[\mathrm{Re}(h^T) \;\; \mathrm{Im}(h^T)]^T$ is recovered from (9.39) using equation (3.7).

$\mathcal{W}^o$ is computed in [39] but when the input symbols are really random variables. Here we compute $\mathcal{W}^o$ for the real distribution of the input symbols: discrete i.i.d. complex circular random variables.

**Result 11** *The complex optimal weighting matrix $\mathcal{W}^o$ is:*

$$\mathcal{W}^o = \frac{1}{M}\left[\sum_{u=-(L+N-1)}^{L+N-1} R_L^T(u) \otimes R_L(-u) - vec\left\{\sigma_a^2 \mathcal{T}(h)\mathcal{T}^H(h)\right\} vec\left\{\sigma_a^2 \mathcal{T}(h)\mathcal{T}^H(h)\right\}\right] \tag{9.40}$$

The first term in the sum would be the term obtained if the input symbols were really Gaussian.

$$\widehat{R}_L = \frac{1}{M}\sum_{k=1}^{M} \boldsymbol{Y}_L(k)\boldsymbol{Y}_L^H(k) \quad \Rightarrow \quad \hat{r} = \mathrm{vec}\left\{\widehat{R}_L\right\} = \frac{1}{M}\sum_{k=1}^{M} \mathrm{vec}\left\{\boldsymbol{Y}_L(k)\boldsymbol{Y}_L^H(k)\right\} \tag{9.41}$$

$$\mathrm{E}\hat{r}\hat{r}^H = \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M} \mathrm{E}\left\{\mathrm{vec}\left\{\boldsymbol{Y}_L(i)\boldsymbol{Y}_L^H(i)\right\} \mathrm{vec}^H\left\{\boldsymbol{Y}_L(j)\boldsymbol{Y}_L^H(j)\right\}\right\} \tag{9.42}$$

Using the following property:

$$\mathrm{vec}\{ABC\} = (C^T \otimes A)\ \mathrm{vec}\{B\} \tag{9.43}$$

$$\mathrm{vec}\{\boldsymbol{Y}_L(i)\boldsymbol{Y}_L^H(i)\} = (I \otimes \boldsymbol{Y}_L(i))\ \ \boldsymbol{Y}_L^*(i) = (\boldsymbol{Y}_L^*(i) \otimes I)\ \ \boldsymbol{Y}_L(i) \tag{9.44}$$

Then:

$$\frac{1}{M^2}\sum_{k=1}^{M}\sum_{k'=1}^{M} \mathrm{E}\left\{\underbrace{(\boldsymbol{Y}_L^*(k) \otimes I)}_{A}\ \underbrace{\boldsymbol{Y}_L(k)}_{B}\underbrace{\boldsymbol{Y}_L^T(k')}_{C}\underbrace{(I \otimes \boldsymbol{Y}_L^*(k'))}_{D}\right\} \tag{9.45}$$

In a first step, we consider the symbols as Gaussian random variables. Then the following formula can be applied:

$$\mathrm{E}\{ABCD\} = \mathrm{E}\{AB\}\ \mathrm{E}\{CD\} + \mathrm{E}\{C \otimes A\}\ \mathrm{E}\{D \otimes B\} + \mathrm{E}\{A\ \mathrm{E}\{BC\}D\} \tag{9.46}$$

$$\mathrm{E}\{AB\} = (\mathrm{E}\{CD\})^H = rr^H \tag{9.47}$$

$$\left\{ \begin{array}{rcccl} \mathrm{E}\{C \otimes A\} & = & \mathrm{E}\{\boldsymbol{Y}_L(k) \otimes \boldsymbol{Y}_L(k') \otimes I\} & = & R_L^T(k-k') \otimes I \\ \mathrm{E}\{D \otimes B\} & = & \mathrm{E}\{I \otimes \boldsymbol{Y}_L(k') \otimes \boldsymbol{Y}_L(k)\} & = & I \otimes R_L(k'-k) \end{array} \right. \tag{9.48}$$

$$\mathrm{E}\{C \otimes A\}\,\mathrm{E}\{D \otimes B\} = \left(R_L^T(k-k') \otimes I\right)\left(I \otimes R_L(k'-k)\right) = R_L^T(k-k') \otimes R_L(k'-k) \tag{9.49}$$

$$\mathrm{E}\left\{\hat{r}\hat{r}^H\right\} = \mathrm{E}\left\{rr^H\right\} + \frac{1}{M^2}\sum_{k,k'} R_L^T(k-k') \otimes R_L(k'-k) \tag{9.50}$$

$$\mathrm{E}\left\{\widetilde{r}\widetilde{r}^H\right\} = \frac{1}{M^2}\sum_{k,k'} R_L^T(k-k') \otimes R_L(k'-k) \tag{9.51}$$

$$\mathrm{E}\left\{\widetilde{\widetilde{r}}\widetilde{\widetilde{r}}^H\right\} = \frac{1}{M}\sum_{u=-(L+N-1)}^{L+N-1} R_L^T(u) \otimes R_L(u) \tag{9.52}$$

Now we consider the input symbols with their true distribution. The only term differing from the Gaussian case is the term where fourth order moments of the input symbols appear, with general term:

$$\begin{aligned} &\mathrm{vec}\left\{\mathcal{T}_L(h)A_L(k)A_L^H(k)\mathcal{T}_L^H(h)\right\}\mathrm{vec}\left\{\mathcal{T}_L(h)A_L(k')A_L^H(k')\mathcal{T}_L^H(h)\right\} \\ &= \left(\mathcal{T}_L^*(h) \otimes \mathcal{T}_L(h)\right) \underbrace{\mathrm{E}\left(I \otimes A_L(k)\right)A_L^*(k)A_L^T(k')\left(I \otimes A_L^H(k')\right)}_{F} \left(\mathcal{T}_L^T(h) \otimes \mathcal{T}_L^H(h)\right) \end{aligned} \tag{9.53}$$

In the Gaussian case, the matrix $F$ would be:

$$F_{Gaus} = r_a r_a^H + R_{A_L A_L}^T \otimes R_{A_L A_L} \tag{9.54}$$

where $R_{A_L A_L} = \mathrm{E}\left(A_L(k)A_L^H(k)\right) = \sigma_a^2 I$ and $r_a = \mathrm{vec}\{C_{A_L A_L}\}$.

For the true distribution, block $(i_b, j_b)$ of matrix $F$ is:

$$F_{i_b j_b} = A_{i_b}^*(k)A(k)A^H(k')A_{j_b}(k') \tag{9.55}$$

It can be verified that $F_{i_b j_b}$ differs from $F_{Gaus,(i_b,j_b)}$ onlt when $k = k'$:

$$F = F_{Gaus} - r_a r_a^H. \tag{9.56}$$

From that, we deduce expression (9.40).

Chapter 10

# SOFT DECISIONS APPLIED TO SEMI–BLIND CHANNEL ESTIMATION

*This chapter examines the difficulty of applying soft decision strategies to channel estimation. Starting from a semi–blind channel estimate, an equalizer is built that gives estimates of the unknown symbols. The most reliable symbols are selected and hard decisions on them are considered as known symbols: semi–blind channel estimation is reprocessed with the augmented number of known symbols. This idea seemed promising but contains some surprising difficulties.*

## 10.1 Principle of the Soft Decision Strategy

The soft decision strategy which is particularly well connected to the general semi-blind context, is as follows (see figure 10.1):

1. From an estimate of the channel, an equalizer is built that gives estimates of the unknown symbols. The most reliable estimates are selected and hard decisions on them are considered as known symbols. The non reliable symbols are still considered as unknown.

2. Semi-blind estimation is again applied with this augmented number of known symbols. Steps 1 and 2 can be reiterated.

This soft decision strategy is opposed to a hard decision strategy where all the equalizer outputs would be considered as error-free and then as known. Some algorithms exploiting the finite alphabet nature of the input symbols [45] follow that scheme: step 2 is replaced by a training sequence based channel estimation step where the training symbols are the hard decisions. This kind of algorithm requires a good channel initialization and because the decision step may not be error-free, fall easily in local minima. This could be avoided by the soft decision process.

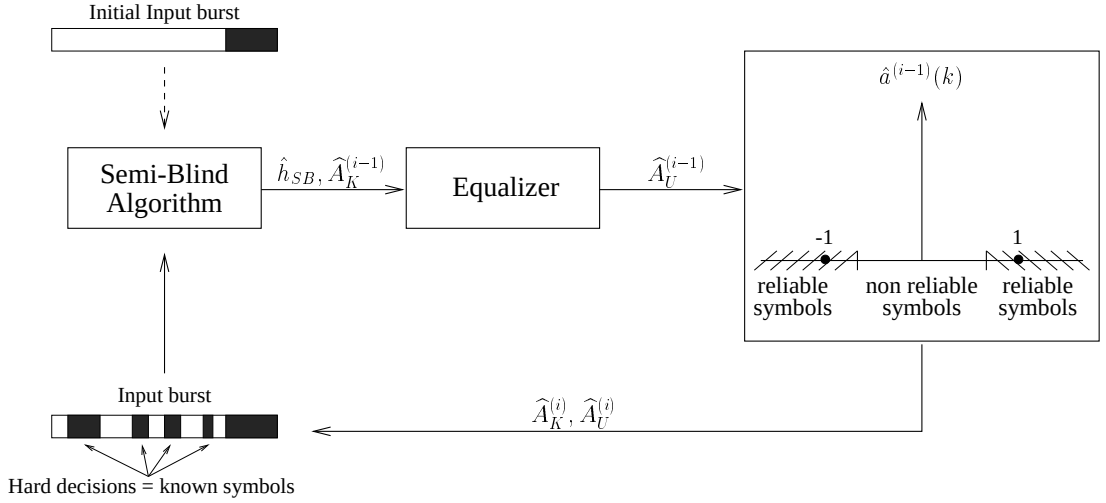



Figure 10.1: Soft Decisions for Semi–Blind Channel Estimation for a BPSK.

## 10.2 Reliable Symbols

Consider a MMSE-ZF equalizer based on the true channel. It gives as estimates for the unknown symbols:

$$\begin{aligned}\widehat{A}_U &= \left(\mathcal{T}_U^H(h)\mathcal{T}_U(h)\right)^{-1}\mathcal{T}_U^H(h)\left(\boldsymbol{Y} - T_K(h)A_K\right)\\ &= A_U + \left(\mathcal{T}_U^H(h)\mathcal{T}_U(h)\right)^{-1}\mathcal{T}_U^H(h)\boldsymbol{V}\,.\end{aligned} \tag{10.1}$$

Then for each unknown symbol:

$$\hat{a}(k) = a(k) + v'(k) \tag{10.2}$$

where $v'(k)$ is a centered Gaussian random variable, a linear combination of elements of $\boldsymbol{V}$.

We will consider here only the case of a BPSK; the principle of soft decisions could be extended to other constellations. Figure 10.2 shows the distribution of $\hat{a}(k)$.

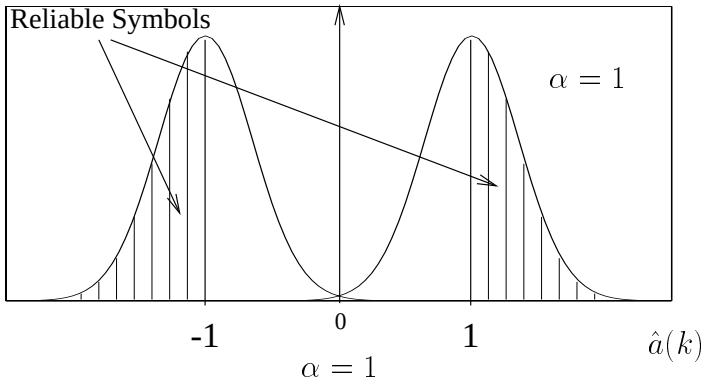



Figure 10.2: Distribution of the symbol estimates at the output of the MMSE ZF equalizer: reliable decisions are such that $|\hat{a}(k)| \geq \alpha$.

The reliable symbol estimates will verify $|\hat{a}(k)| \geq \alpha$; they will be all the more reliable as $\alpha$ is large: see figure 10.2 with $\alpha = 1$, in which case, as $v'(k)$ is centered, approximately half the symbol estimates would be considered as reliable.

## 10.3 The Difficulty of Applying Soft Decisions

This soft decision strategy introduces correlations between $a(k)$ $(= \mathrm{dec}(\hat{a}(k)))$ and $v'(k)$ and then between the noise $\boldsymbol{V}$ and the symbols $A$ originally independent. Figure 10.3 shows the joint distribution of $a(k)$ and $v'(k)$ for the reliable and non-reliable $\hat{a}(k)$: in both cases, $a(k)$ and $v'(k)$ are correlated (for $\alpha = 1$, the marginal distribution of $v'(k)$ remains approximately unchanged).

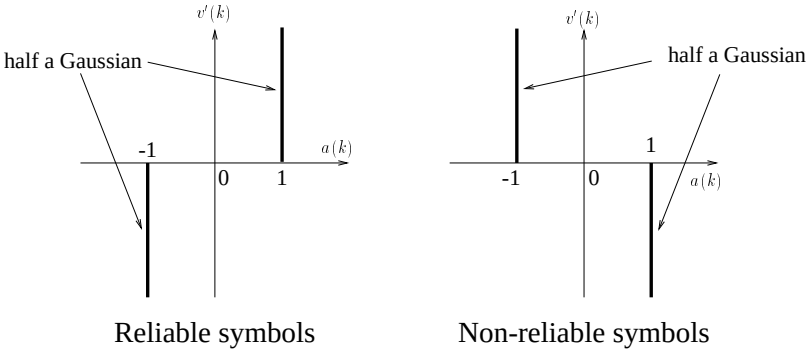



Figure 10.3: Joint distribution of $\boldsymbol{v}'(k)$ and $a(k)$ for the reliable (left) and non reliable symbols (right) for the asymmetric reliability intervals.

Simulations proved GML and DML to be very sensitive to these modifications in the correlations: we get better performance by not adding the hard decisions to the list of the known symbols. The repercussions of these correlations in the formulations of DML and GML are as follows:

- For DML, in $f_{V|A,h}(\boldsymbol{Y} - \mathcal{T}(h)A)$, correlations between $A$ and $V$ are to be taken into account: $v'(k)|a(k)$ does not have a Gaussian distribution anymore, but half a Gaussian.
- For GML, in a Gaussian approximation for $f_{Y|h}(\boldsymbol{Y}|h)$, the correlations between $A$ and $\boldsymbol{V}$ have to be taken into account.

Incorporation of these modifications have to be done in order to build properly the ML criterion. As an alternative approach, we tried another type of interval of reliability: symmetric intervals around the decision points.

With the choice of the symmetric intervals, where the symbols are considered as reliable if $|\hat{a}(k) - \mathrm{dec}(\hat{a}(k))| \leq \beta$ (figure 10.4), the correlation between symbols and noise disappears (as long as the interval is sufficiently small): see figure 10.5. The marginal distribution of the noise has changed though, and namely the variance of $v'(k)$ associated to the reliable or non-reliable symbols is different.

Note that the MMSE-ZF equalizer could be replaced by an MMSE equalizer which gives a higher output SNR. At the output of the equalizer, the ISI terms should now be taken into account, as well as its bias (see Chapter 11).

We illustrate the effect of the two types of intervals by an example. In the example, we plot the semi–blind cost function of DML and GML for a real multichannel with 2 coefficients: $[1 \quad -1.5]^T$, a burst of length $M = 100$, a BPSK and an SNR equal to 5dB. The threshold $\alpha$ is chosen equal to 0.3. We consider furthermore the ideal case where the hard decisions are considered as error–free.

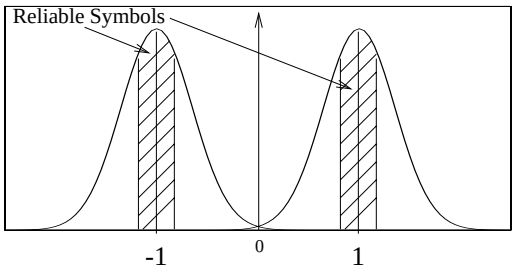



Figure 10.4: Distribution of the symbol estimates at the output of the MMSE ZF equalizer: reliable decisions such that $|\hat{a}(k) - \mathrm{dec}(\hat{a}(k))| \geq \beta$.

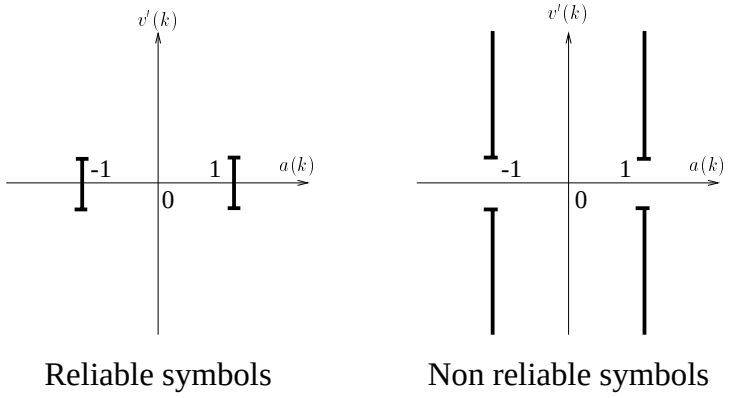



Figure 10.5: Joint distribution of $\boldsymbol{v}'(k)$ and $a(k)$ for the reliable (left) and non reliable decisions (right) for the symmetric reliability intervals.

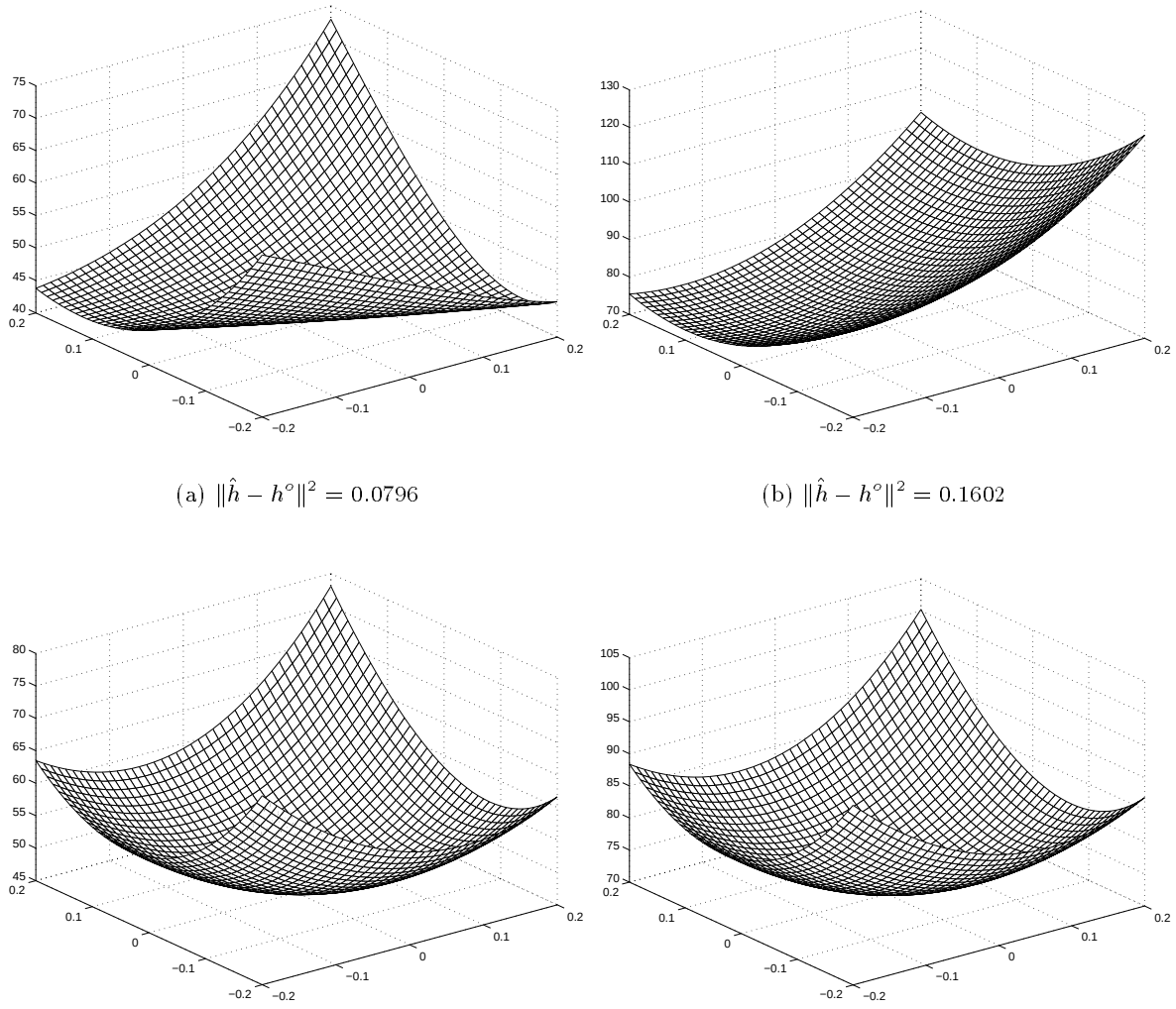
(a) Figures 10.6(a)–10.7(a) shows the pure semi–blind cost function, with 10 known symbols in the burst.

(b) Figures 10.6(b)–10.7(b) shows the cost function after considering the hard decisions given by the asymmetric reliability intervals as known. Only 17 symbols remain unknown after the soft decision step.

(c) Figures 10.6(c)–10.7(c) shows the cost function after considering the hard decisions given by the symmetric reliability intervals as known: 40 symbols remain unknown after the soft decision step.

(d) At last, in figures 10.6(d)–10.7(d), the soft decision have the same number as the previous ones (symmetric reliability interval), but this time the position of the known symbols is chosen arbitrarily in the burst.

In this example, we see that chosing non symmetric reliability intervals results in worse performance than the pure semi–blind case; the symmetric reliability intervals improve performance w.r.t. the pure semi–blind case.



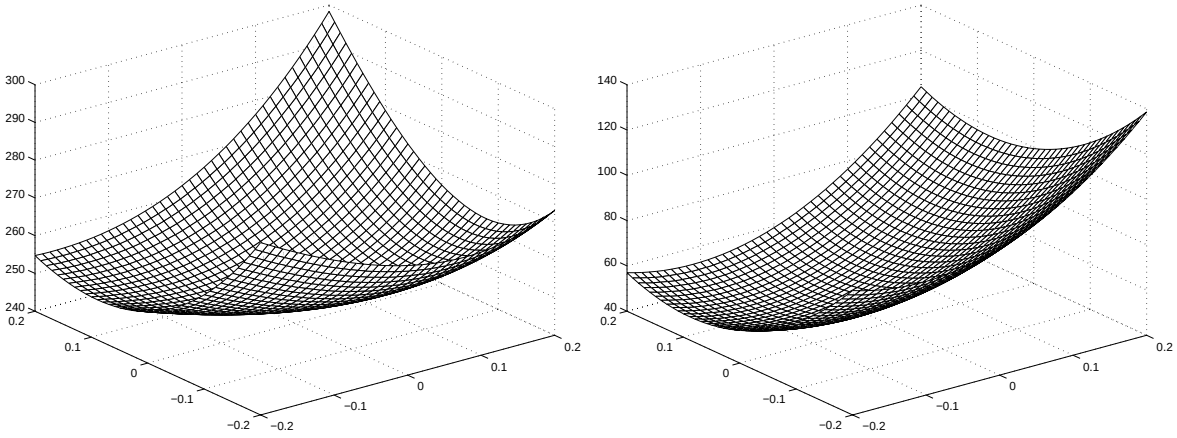

(a) $\|\hat{h} - h^o\|^2 = 0.0796$

(b) $\|\hat{h} - h^o\|^2 = 0.1602$

(c) $\|\hat{h} - h^o\|^2 = 0.0108$

(d) $\|\hat{h} - h^o\|^2 = 0.0141$

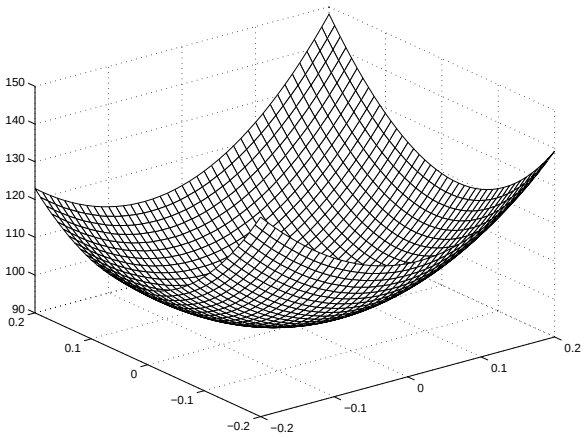
Figure 10.6: DML cost functions: (a) pure semi–blind DML cost function; (b) semi–blind DML based on hard decisions (asymmetric interval); (c) semi–blind DML based on hard decisions (symmetric interval); (d) semi–blind DML based on hard decisions (randomly chosen).



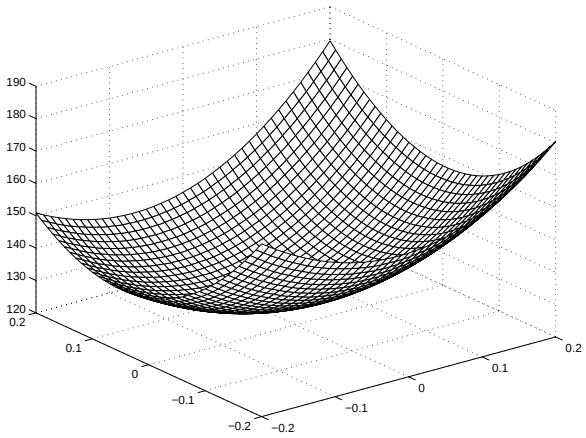

(a) $\|\hat{h} - h^o\|^2 = 0.0474$




(b) $\|\hat{h} - h^o\|^2 = 0.1388$




(c) $\|\hat{h} - h^o\|^2 = 0.0211$




(d) $\|\hat{h} - h^o\|^2 = 0.0130$

Figure 10.7: GML cost functions: (a) pure semi–blind GML cost function; (b) semi–blind GML based on hard decisions (asymmetric interval); (c) semi–blind GML based on hard decisions (symmetric interval); (d) semi–blind GML based on hard decisions (randomly chosen).

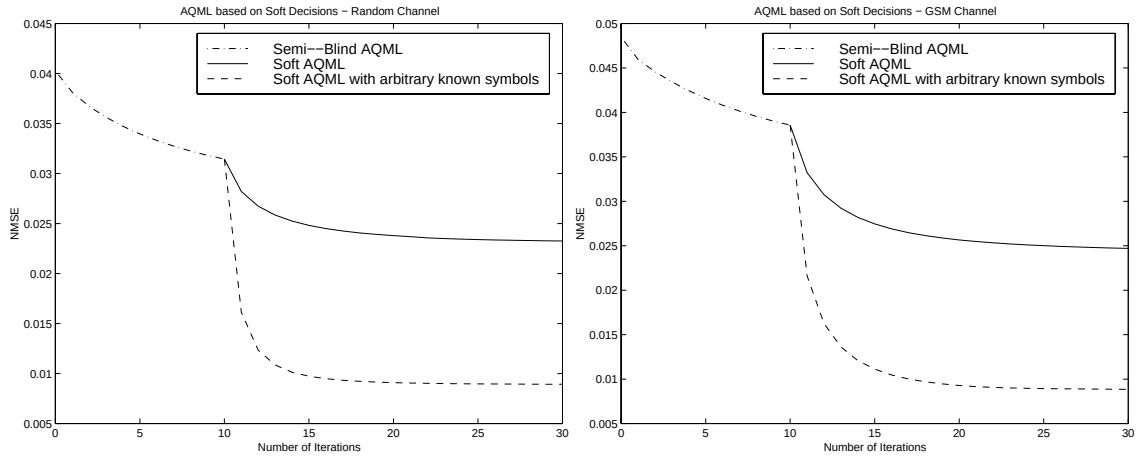



Figure 10.8: Soft AQML based on hard decisions compared soft AQML with the same number of hard decisions but arbitrarily dispersed.

## 10.4 Soft Decisions Applied to the Semi–Blind AQML

In practice, for more general channels, the symmetric reliability interval gives disappointing results. It improves performance w.r.t. purely semi–blind estimation but not significantly. As an example, we applied the soft strategy to semi–blind AQML. Results are shown in figure 10.4 for a randomly chosen channel ($N = 5$, $m = 2$) and a GSM channel ($N = 4$, $m = 2$), for a burst length of 100, and an SNR of 10dB. The soft AQML is compared to another soft AQML based on the same number of known symbols (including also the hard decisions) but with position randomly chosen in the burst.

## 10.5 Conclusion

In this chapter, we have seen that the channel estimation is sensitive to the soft decision strategy; we have proposed a symmetric reliability interval that decreases this sensitivity. As we will see in Chapter 12, soft decisions can also be used for the input symbol estimation and detection: in this case, correlations between the symbols and the noise are also present, and the structure of the equalizers should be modified accordingly. However, it appears that symbol detection is not sensitive at all to these changes in the correlations, and soft decisions can be very profitable in that case.

# Part III

# Receiver Structures

Chapter 11

# BURST MODE EQUALIZATION

*We consider a transmission by burst where the data is organized and sent by bursts. At each end of the burst of data, a sequence of symbols is assumed known, and the channel considered as constant over the burst duration. The optimal structure of the burst mode equalizers is derived. The class of linear and decision feedback equalizers is considered, as well the class of ISI cancelers that use past but also future decisions: for each class of equalizers the MMSE, the Unbiased MMSE and the MMSE Zero Forcing versions are derived. Unlike in the continuous processing mode, the optimal burst mode filters are time-varying. The performance of the different equalizers are evaluated and compared to each other in terms of SNR and probability of error: these measures depend on the position of the estimated symbol and on the presence of known symbols. Finally, we show that, by choosing correctly the number and position of the known symbols, (time-invariant) continuous processing filters applied to burst mode can be organized to give sufficiently good performance, so that optimal (time-varying) burst processing implementation can be avoided. This chapter extends the work of [87] and corresponds to the submitted paper [88].*

## 11.1 Introduction

In most of the present mobile communication systems, the data is divided and transmitted in bursts. In general, the bursts are separated by guard intervals, which avoid interburst interference, and contain known symbols, like synchronization bits or a training sequence to estimate the channel. This is typically the case of GSM, where the channel is assumed constant over the duration of a burst and is estimated by a middamble training sequence and the Viterbi algorithm is applied to estimate the transmitted data symbols.

We propose a scenario where a sequence of known symbols is attached to each end of the burst of information symbols. This scheme is proved to include the GSM case. The channel is assumed constant during the transmission of a burst. As we are operating with a finite amount of data, the usual time-invariant continuous processing equalizers are not optimal anymore. We propose a derivation of optimal burst mode equalizers, which are time-varying. Three classes of equalizers are considered: the usual linear and decision feedback equalizers, as well as the ISI canceler. This last equalizer uses past but also future decisions and was proposed in its continuous processing version in [89, 90], and in its burst mode version in [91, 92] where it is called Non Causal Decision Feedback Equalizer (NCDFE). The NCDFE is detailed in Chapter 12.

These three classes of equalizers are derived according to three different criteria: MMSE, Unbiased MMSE and MMSE Zero Forcing (MMSE-ZF) corresponding to increasingly strong constraints; the first criterion is unconstrained, the second one is the element–wise Best Linear Unbiased Estimate (BLUE), and the third one is the block–wise BLUE. These three criteria will then give increasing MSEs. The MMSE equalizer gives biased estimates of the symbols: the Unbiased MMSE equalizer is the best equalizer in the MMSE sense, giving unbiased estimates. Although possessing a higher SNR than its MMSE counterpart, the Unbiased MMSE equalizer gives a better error probability because the decision device is built for unbiased symbols estimates. The Unbiased MMSE DFE equalizer was introduced in [93]; we propose here a generalization to the other classes of equalizers.

All the equalizers are derived in the multichannel framework. We prove that the optimal processing consists in first removing the contribution of the known symbols, then applying the burst mode multichannel matched filter; the following filters depend on the specific equalizer considered. The performance of the different equalizers is evaluated in terms of SNR, studied according to the position of the unknown symbols in the burst and the presence of the known symbols.

In [94], burst mode MMSE and ZF equalizers are derived but for single channels: the ZF equalizer exists then only if there are at least a number of known symbols equal to the channel memory. In the multichannel context considered here, even with no known symbols, ZF equalizers exist and in fact a whole class of ZF equalizers: we will present the special class of MMSE-ZF versions of the equalizers. Furthermore, in [94], continuous–time matched

filtering is done, matched to the overall channel, which is unrealizable, followed by the symbol rate burst mode processing. We follow the more realistic fractionnally–spaced approach in which simple continuous–time lowpass filtering is followed by oversampling. Another interest of the multichannel model, which we will not detail here, is that it allows better performance as the number of subchannels increases.

[94] presents complexity computations of the burst mode filters, which appear more complex than the time–invariant filter continuous processing mode. We propose to compare the performance obtained by applying the optimal time–varying burst mode filters with the performance obtained by applying the time–invariant continuous processing filters to burst mode, which is not done in [94]. [95, 96] proposed to enable time–invariant processing (with cyclic convolution though) by introducing cyclic prefixes. We propose to minimize the suboptimality of continuous processing by considering the influence of the pre– and postamble lengths on the degradation between time–invariant filters and the optimal processing: the best situation happens when the lengths of these pre– and postambles equal the channel memory. In [97], N. Al–Dhahir presents such a comparison, but considering the single channel MMSE DFE only. His treatment of the known symbol is not correct however. He estimates the unknown symbols in terms of the received data only, whereas the correct treatment consists in estimating the unknown symbols in terms of the received data and also of the known symbols present in the burst. In his attempt to compare time–invariant and optimal processing fairly, he averages SNR in both cases over different amounts of symbols, estimating the known symbols also in the time–invariant processing, whereas the number of unknown symbols (to be equalized) is the same in both cases. So the comparison appears unfair. Furthermore, he summarizes the performance into one SNR average number over the burst: as will be seen in the paper, it appears important to consider on the contrary the performance as a function of symbol position.

## 11.2 Burst Transmission

We consider a transmission by burst in which detection is done burst by burst. We assume that the channel is time-invariant during the transmission of a burst. In the input burst $A$, a pre– and post–amble sequence of known symbols of variable length is attached to the burst of data symbols: $n_1$ known symbols at the beginning, grouped in the vector $A_{K_1}$, and $n_2$ at the end, grouped in the vector $A_{K_2}$: see figure 11.1. The total length of the burst is $L+n_1+n_2$; we want to detect the $L$ central unknown symbols, grouped in the vector $A_U$. For that purpose, we consider as observation data $\boldsymbol{Y}$, the channel outputs that contain only symbols of burst $A$ (the symbols to be detected or the known symbols of the burst), and not outputs containing symbols of neighboring bursts: see figure 11.1.

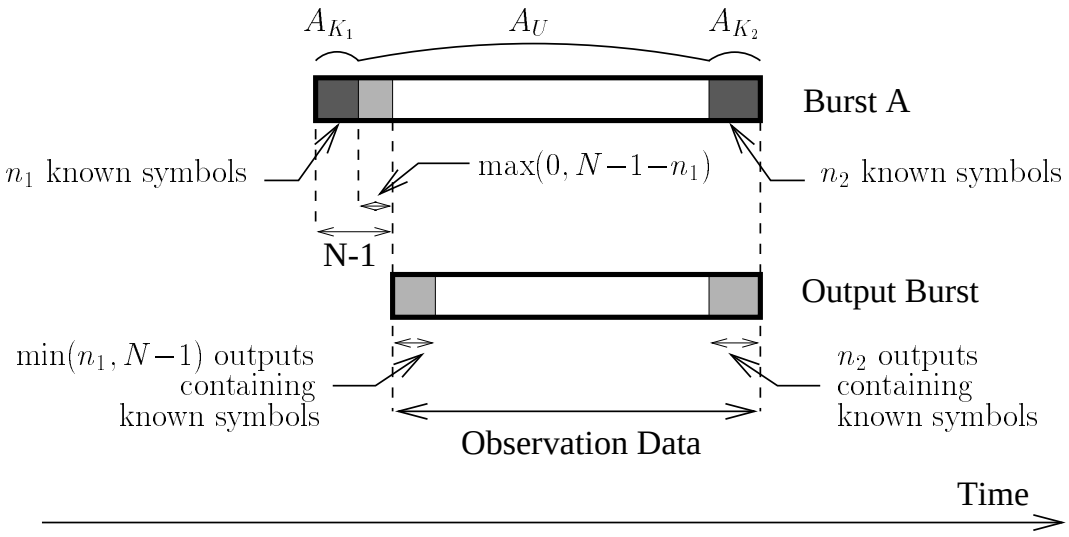



Figure 11.1: Burst Transmission.

In the following, we consider the decomposition:

$$\boldsymbol{Y} \;=\; \mathcal{T}A + \boldsymbol{V} \;=\; \mathcal{T}_{K_1}A_{K_1} + \mathcal{T}_U A_U + \mathcal{T}_{K_2}A_{K_2} + \boldsymbol{V} \;=\; \mathcal{T}_K A_K + \mathcal{T}_U A_U, \tag{11.1}$$

In this chapter, we will denote $\mathcal{T}$ the convolution matrix. $\mathcal{T}_i A_i$ represents the contribution of the symbols in $A_i$. $A_K = [A_{K_1}^T \;\; A_{K_2}^T]^T$ groups all the known symbols. As will be seen, the optimal process consists first in removing the contribution of the known symbols, all the filters will then be applied to the processing data $\boldsymbol{Y}_U$:

$$\boldsymbol{Y}_U \;=\; \mathcal{T}_U A_U + \boldsymbol{V} \;=\; \boldsymbol{Y} - \mathcal{T}_K A_K, \tag{11.2}$$

It should be noted that the derivations of the paper are valid for any position for the known symbols.

## 11.3 Burst-Mode Equalizers

In this section, we derive the expressions for the different equalizers in burst mode. Linear Equalizers (LE), classical Decision Feedback Equalizers (DFE) and the Non Causal DFE (NCDFE) are considered for the Minimum Mean Squared Error (MMSE), the Unbiased MMSE ($\mathcal{U}$MMSE) and the MMSE Zero-Forcing (MMSE ZF) criteria. The different equalizers are linear estimators of the input symbols:

- linear equalizers give linear estimates based on the received data $\boldsymbol{Y}$ and the known symbols $A_K$,

- DFEs give linear estimates based on $\boldsymbol{Y}$, $A_K$, as well as the decisions on the past input symbols,
- the NCDFE gives linear estimates given the $\boldsymbol{Y}$, $A_K$ and the decisions on the past and future input symbols.

We shall assume those past (and future) decisions to be error-free.

The different equalizers are solutions of the MSE criterion

$$\min_F \|A_U - F\boldsymbol{Y}'\|^2 \tag{11.3}$$

where $F$ is a matrix filled out with filter coefficients and $\boldsymbol{Y}'$ groups the whole observation set (e.g. $\boldsymbol{Y}$ and $A_K$ for the LE), under different constraints:

- MMSE: no constraints.
- $\mathcal{U}$MMSE: element–wise Best Linear Unbiased Estimate (BLUE).
- MMSE Zero-Forcing: burst–wise BLUE.

In burst mode, the equalizer filters are time-varying. We define the MSE of the ith symbol as:

$$MSE_i = \left(E(\widehat{A}_U - A_U)(\widehat{A}_U - A_U)^H\right)_{ii} \tag{11.4}$$

where $\widehat{A}_U$ is the vector estimate of the unknown input symbols and the Signal to Noise Ratio (SNR) of the ith symbol:

$$SNR_i = \frac{\sigma_a^2}{MSE_i}\,. \tag{11.5}$$

### 11.3.1 Linear Equalizers

#### The MMSE Linear Equalizer

The MMSE LE gives the unconstrained MMSE estimate of the unknown symbols $A_U$ based on the observations:

$$\boldsymbol{Y}' = \begin{bmatrix} \boldsymbol{Y}^T & A_K^T \end{bmatrix}^T\,. \tag{11.6}$$

The linear MMSE estimate of $A_U$ is:

$$\widehat{A}_{U,\,\text{MMSE LE}} = R_{A_U\boldsymbol{Y}'}R^{-1}_{\boldsymbol{Y}'\boldsymbol{Y}'}\boldsymbol{Y}' = R_{A_U\boldsymbol{Y}_U}R^{-1}_{\boldsymbol{Y}_U\boldsymbol{Y}_U}\boldsymbol{Y}_U\,. \tag{11.7}$$

The last equality, proved in Appendix A, shows that linear estimation in terms of $\boldsymbol{Y}'$ is the same as in terms of $\boldsymbol{Y}_U$: the optimal processing can be seen as eliminating first the contributions of known symbols from the observation data $\boldsymbol{Y}$ to get $\boldsymbol{Y}_U$ and then applying

the MMSE equalizer determined on the basis of $\boldsymbol{Y}_U$. For the other equalizers, the previous result is also true but will not be restated.

When a sequence of known symbols of length larger than the channel memory is present in the middle of the burst, like in GSM, the processing data $\boldsymbol{Y}_U$ can then be decomposed into two independent parts, and all the equalization process can be done on the two parts independently. For each part, the situation becomes equivalent to our proposed scenario of known symbols at each end of the burst.

From equation (11.7):

$$\widehat{A}_{U,\,\text{MMSE LE}} = \sigma_a^2 \mathcal{T}_U^H (\sigma_a^2 \mathcal{T}_U(h)\mathcal{T}_U^H(h) + \sigma_v^2 I)^{-1} \boldsymbol{Y}_U = \left(\mathcal{T}_U^H \mathcal{T}_U + \frac{\sigma_v^2}{\sigma_a^2} I\right)^{-1} \mathcal{T}_U^H \boldsymbol{Y}_U \,. \quad (11.8)$$

The last equality is obtained via the matrix inversion lemma. We will denote:

$$R = \mathcal{T}_U^H \mathcal{T}_U + \frac{\sigma_v^2}{\sigma_a^2} I \,. \quad (11.9)$$

Due to the presence of the regularizing term $\frac{\sigma_v^2}{\sigma_a^2} I$, the matrix $R$ is invertible and the MMSE LE is always defined.

In the continuous processing case, the MMSE equalizer gives the output:

$$\hat{a}_{\text{MMSE LE}}(k) = \left(\mathbf{H}^\dagger(q)\mathbf{H}(q) + \frac{\sigma_v^2}{\sigma_a^2} I\right)^{-1} \mathbf{H}^\dagger(q)\boldsymbol{y}(k) \quad (11.10)$$

where $\mathbf{H}^\dagger(z) = \mathbf{H}^H(1/z^*)$ and $q^{-1}\boldsymbol{y}(k) = \boldsymbol{y}(k-1)$. By analogy with the continuous processing case, we can find interpretations for the expression (11.8) in filtering terms:

- $\mathcal{T}_U^H$ represents the multichannel matched filter, matched to the filter $\mathcal{T}_U$. When the length of the two sequences of known symbols equals or is larger than the memory of the channel $N-1$, $\mathcal{T}_U^H$ is Toeplitz, banded and upper triangular, which implies that the filtering is time-invariant, FIR and anticausal. When the length of the sequences are shorter however, the filter is time-varying at the edges.

- $R$ is the FIR denominator of an IIR filter, $R^{-1}$ is non-causal.

Figure 11.2 shows the MMSE LE structure.

The LDU decomposition of $R = LDL^H$ can be used to do a fast implementation of the MMSE LE as mentioned in [94]. The Schur algorithm can indeed be used to compute these factors. $R^{-1} = L^{-H}D^{-1}L^{-1}$: the output of $L^{-1}$, $Z' = L^{-1}Z \Rightarrow LZ' = Z$, can be solved by backsubstitution. The same kind of remark is valid for the output of $L^{-H}$. So inverting $R$ becomes superfluous and the complexity is of order $MN$.

For the burst mode MMSE LE:

$$SNR_i(\text{MMSE LE}) = \frac{\sigma_a^2}{\sigma_v^2 (R^{-1})_{ii}}\,. \tag{11.11}$$

The SNR depends on the position of the symbol in the burst and we will see the influence of the known symbols on the SNR according to the position of the symbols to be estimated. This remark will also be valid for the other equalizers.

### The General Unbiased MMSE Problem

A MMSE equalizer produces a biased estimate of the symbol $a(i)$: the MMSE equalizer output can indeed be written as $\alpha(i)a(i) + n(i)$, where $n(i)$ and $a(i)$ are uncorrelated ($n(i)$ contains symbols different from $a(i)$ and noise terms). This bias increases the probability of error [93], as the decision devices are made for unbiased data. The purpose of the unbiased MMSE equalizer is to correct this bias. We then derive the best equalizer, in the MMSE sense, that gives unbiased symbol estimates: we will see that its SNR gets reduced w.r.t. the MMSE, but that the error probability increases. Note that ZF equalizers are unbiased equalizers: they minimize the MSE under the unbiasness constraint but also the zero ISI constraint; the $\mathcal{U}$MMSE are derived under the unbiasness constraint only. So ZF and $\mathcal{U}$MMSE equalizers are different except when there is no ISI at the output of the $\mathcal{U}$MMSE, which will be the case for the NCDFE.

In terms of estimation theory, the Unbiased MMSE equalizer is the element–wise BLUE. We give and prove here results that will be valid for all the Unbiased MMSE equalizers (LE, DFE, NCDFE).

Consider the estimation of symbol $a(i)$. $\boldsymbol{Y}'$ contains all the information available for estimation, $\boldsymbol{Y}_U$ and $\overline{A}$: $\overline{A}$ denotes here the past decisions w.r.t. $a(i)$ for the DFE, the past and future decisions for the NCDFE, and is zero for the LE. Let us decompose the processing data $\boldsymbol{Y}_U$ into the contribution of $a(i)$ and of the other symbols $\overline{A}_{U,i}$.

$$\boldsymbol{Y}_U = \mathcal{T}_U A_U + \boldsymbol{V} = \mathcal{T}_{U,i} a(i) + \overline{\mathcal{T}}_{U,i} \overline{A}_{U,i} + \boldsymbol{V} \tag{11.12}$$

$$\boldsymbol{Y}' = \begin{bmatrix} \mathcal{T}_{U,i} \\ 0 \end{bmatrix} a(i) + \begin{bmatrix} \overline{\mathcal{T}}_{U,i} \overline{A}_{U,i} + \boldsymbol{V} \\ \overline{A} \end{bmatrix} = \mathcal{T}'_{U,i} a(i) + \boldsymbol{V}'\,. \tag{11.13}$$

The BLUE theory for this linear model gives us as estimate for $a(i)$:

$$\hat{a}_{\text{BLUE}}(i) = \left( \mathcal{T}'^H_{U,i} R^{-1}_{\boldsymbol{Y}'\boldsymbol{Y}'} \mathcal{T}'_{U,i} \right)^{-1} \mathcal{T}'^H_{U,i} R^{-1}_{\boldsymbol{Y}'\boldsymbol{Y}'} \boldsymbol{Y}' \tag{11.14}$$

which can also be written as:

$$\hat{a}_{\text{BLUE}}(i) = \sigma_a^2 \left( R_{a(i)\boldsymbol{Y}'} R^{-1}_{\boldsymbol{Y}'\boldsymbol{Y}'} R_{a(i)\boldsymbol{Y}'} \right)^{-1} \underbrace{R_{a(i)\boldsymbol{Y}'} R^{-1}_{\boldsymbol{Y}'\boldsymbol{Y}'} \boldsymbol{Y}'}_{\hat{a}_{\text{MMSE}}(i)}\,. \tag{11.15}$$

The Unbiased estimate for the whole burst is then:

$$\widehat{A}_{U,\,\mathcal{U}\text{MMSE}} = \underbrace{\sigma_a^2 \left(\text{diag}\left(R_{A_U\boldsymbol{Y}'} R^{-1}_{\boldsymbol{Y}'\boldsymbol{Y}'} R_{\boldsymbol{Y}'A_U}\right)\right)^{-1}}_{\mathcal{D}} \underbrace{R_{A_U\boldsymbol{Y}'} R^{-1}_{\boldsymbol{Y}'\boldsymbol{Y}'}\boldsymbol{Y}'}_{\widehat{A}_{\text{U, MMSE}}} \tag{11.16}$$

$$\widehat{A}_{U,\,\mathcal{U}\text{MMSE}} = \mathcal{D}\ \widehat{A}_{U,\,\text{MMSE}}\,. \tag{11.17}$$

The Unbiased MMSE equalizer is simply a scaled version of the MMSE equalizer.

The SNR of the $\mathcal{U}$MMSE is related to the SNR of the MMSE:

$$SNR_i(\mathcal{U}\text{MMSE}) = SNR_i(\text{MMSE}) - 1\,. \tag{11.18}$$

The proof can be found in Appendix B.

### The Unbiased MMSE Linear Equalizer

For the specific case of the MMSE LE:

$$\begin{aligned} R_{A_U\boldsymbol{Y}'} R^{-1}_{\boldsymbol{Y}'\boldsymbol{Y}'} R_{\boldsymbol{Y}'A_U} &= \left(\mathcal{T}_U^H\mathcal{T}_U + \frac{\sigma_v^2}{\sigma_a^2} I\right)^{-1} \mathcal{T}_U^H\mathcal{T}_U \\ \Rightarrow \quad \mathcal{D} &= \left(\text{diag}\left(\left(\mathcal{T}_U^H\mathcal{T}_U + \frac{\sigma_v^2}{\sigma_a^2} I\right)^{-1} \mathcal{T}_U^H\mathcal{T}_U\right)\right)^{-1}\,. \end{aligned} \tag{11.19}$$

$\mathcal{D}$ can be further rearranged, and we get:

$$\widehat{A}_{U,\,\mathcal{U}\text{MMSE LE}} = \left(I - \frac{\sigma_v^2}{\sigma_a^2}\text{diag}\left[\left(\mathcal{T}_U^H\mathcal{T}_U + \frac{\sigma_v^2}{\sigma_a^2} I\right)^{-1}\right]\right)^{-1} \widehat{A}_{U,\,\text{MMSE LE}}\,. \tag{11.20}$$

As $\mathcal{D}$ is invertible, $\widehat{A}_{U,\,\mathcal{U}\text{MMSE LE}}$ is always defined. Figure 11.3 shows the $\mathcal{U}$MMSE structure.

In the continuous processing case, the output of the $\mathcal{U}$MMSE LE has for expression:

$$\hat{a}_{\mathcal{U}\text{MMSE LE}}(k) = \left(1 - \frac{\sigma_v^2}{\sigma_a^2}\oint \frac{dz}{z}\left(\mathbf{H}^\dagger(z)\mathbf{H}(z) + \frac{\sigma_a^2}{\sigma_v^2}\right)\right)^{-1} \hat{a}_{\text{MMSE LE}}(k)\,. \tag{11.21}$$

### The MMSE–ZF Linear Equalizer

A ZF equalizer has for property to leave the signal part of the received data undistorted: a block ZF equalizer $F$ verifies:

$$F\mathcal{T}_U = I\,. \tag{11.22}$$

In the monochannel case, the existence of the ZF equalizer is conditioned to the presence of known symbols. When there are no known symbols, $\mathcal{T}_U = \mathcal{T}$ admits no left inverse. For a number of known symbols of exactly $N-1$, the channel memory, $\mathcal{T}_U$ is square and there is a

unique ZF equalizer which is also the MMSE ZF equalizer. For a number of known symbols of more than $N-1$, $\mathcal{T}_U$ is strictly tall and full-column rank and ZF equalizers exist.

For a multichannel and also for a single channel, $\mathcal{T}_U$ has full column rank if $M \geq N$ and if there are as many known symbols as the number of zeros. These will be the conditions for a ZF equalizer to exist.

When $\mathcal{T}_U$ is strictly tall and has full column rank, it admits several left inverses. Indeed, let $\mathcal{T}_U^{\perp}$ be a matrix which columns are orthogonal to those of $\mathcal{T}_U$, then $\mathcal{T}_U^{\perp H}\mathcal{T}_U = 0$. $F = \left(\mathcal{T}_U^H\mathcal{T}_U\right)^{-1}\mathcal{T}_U^H$, the Moore–Penrose pseudo–inverse of $\mathcal{T}_U$ verifies $F\mathcal{T}_U = I$, but also $F = \left(\mathcal{T}_U^H\mathcal{T}_U\right)^{-1}\mathcal{T}_U^H + C\mathcal{T}_U^{\perp H}$, where $C$ is any $M \times M$ matrix.

We shall here concentrate on the MMSE–ZF LE, which give the lowest MSE among all the ZF LE equalizers. The MMSE–ZF LE corresponds to the block–wise BLUE based on $\boldsymbol{Y}_U$. Given the linear model: $\boldsymbol{Y}_U = \mathcal{T}_U A_U + \boldsymbol{V}$, the BLUE is given by:

$$\begin{aligned}\widehat{A}_{U,\,\text{BLUE}} &= (\mathcal{T}_U^H R^{-1}_{\boldsymbol{Y}_U\boldsymbol{Y}_U}\mathcal{T}_U)^{-1}\mathcal{T}_U^H R^{-1}_{\boldsymbol{Y}_U\boldsymbol{Y}_U}\boldsymbol{Y} \\ &= (\mathcal{T}_U^H R^{-1}_{\boldsymbol{V}\boldsymbol{V}}\mathcal{T}_U)^{-1}\mathcal{T}_U^H R^{-1}_{\boldsymbol{V}\boldsymbol{V}}\boldsymbol{Y}_U\,.\end{aligned} \tag{11.23}$$

So the MMSE–ZF LE is:

$$\widehat{A}_{U,\,\text{MMSE–ZF LE}} = (\mathcal{T}_U^H\mathcal{T}_U)^{-1}\mathcal{T}_U^H\boldsymbol{Y}_U\,. \tag{11.24}$$

Consider now the LDU decomposition of $\mathcal{T}_U^H\mathcal{T}_U = LDL^H$:

$$\left(\mathcal{T}_U^H\mathcal{T}_U\right)^{-1} = L^{-H}D^{-1}L^{-1}\,. \tag{11.25}$$

After the matched filter, the optimal process consists in whitening the noise by the filter $L^{-1}$. We will find these two optimal steps (matched filtering and noise whitening) for all the ZF equalizers. If we denote now $R = \mathcal{T}_U^H\mathcal{T}_U$, the process is the same as for the MMSE LE (see figure 11.2). The remarks on the fast implementation are also valid here.

The output burst mode SNR is:

$$SNR_i(\text{MMSE–ZF LE}) = \frac{\sigma_a^2}{\sigma_v^2((\mathcal{T}_U^H\mathcal{T}_U)^{-1})_{ii}}\,. \tag{11.26}$$

In the continuous processing case, the MMSE–ZF LE output is:

$$\hat{a}_{\text{MMSE–ZF LE}}(k) = \left(\mathbf{H}^{\dagger}(q)\mathbf{H}(q)\right)^{-1}\mathbf{H}^{\dagger}(q)\boldsymbol{y}(k)\,. \tag{11.27}$$

### 11.3.2 Decision Feedback Equalizers

#### The MMSE Decision Feedback Equalizer

The decision feedback equalizers consider the linear estimation of symbol $a(i)$ based on the processing data $\boldsymbol{Y}_U$ and the past decisions w.r.t. $a(i)$ assumed known that we denote $A_i^p$:

$$\hat{a}_{\text{MMSE DFE}}(i) = F_i\boldsymbol{Y}_U - B_iA_i^p \tag{11.28}$$

where $F_i$ is the forward filter and $B_i$ the feedback filter. Let $\boldsymbol{Y}' = \left[\boldsymbol{Y}_U^T \;\; A_i^{pT}\right]^T$, and let us decompose $\boldsymbol{Y}_U$ onto the contribution of $A_i^p$ the past symbols and $A_i^f$ grouping $a(i)$ and the future symbols.

$$\boldsymbol{Y}_U = \mathcal{T}_U^p A_i^p + \mathcal{T}_U^f A_i^f + \boldsymbol{V} \,. \tag{11.29}$$

$$[F_i \;\; -B_i] = R_{a(i)\boldsymbol{Y}'} R_{\boldsymbol{Y}'\boldsymbol{Y}'}^{-1} = \left[\sigma_a^2 \mathcal{T}_{U,i}^H \left(\sigma_a^2 \mathcal{T}_U^f \mathcal{T}_U^{fH} + \sigma_v^2 I\right)^{-1} \;\; -\sigma_a^2 \mathcal{T}_{U,i}^H \left(\sigma_a^2 \mathcal{T}_U^f \mathcal{T}_U^{fH} + \sigma_v^2 I\right)^{-1} \mathcal{T}_U^p\right] . \tag{11.30}$$

Consider the LDU factorization of $R = \mathcal{T}_U^H \mathcal{T}_U + \dfrac{\sigma_v^2}{\sigma_a^2} I = LDL^H$. After some manipulations, it can be proved that $F_i$ is the $i^{th}$ row of $D^{-1}L^{-1}\mathcal{T}_U^H$ and that $B_i$ the $i^{th}$ row of $L^H - I$. A proof for this result is provided in Appendix C.

The symbol estimate is then:

$$\widehat{A}_{U,\,\text{MMSE DFE}} = D^{-1}L^{-1}\mathcal{T}_U^H \boldsymbol{Y}_U - (L^H - I)\text{dec}(\widehat{A}_{U,\,\text{MMSE DFE}}) \,. \tag{11.31}$$

The MMSE DFE is always defined like the MMSE LE. The forward filter consists in the cascade of the multichannel matched filter and an anticausal filter $D^{-1}L^{-1}$. $L^H - I$ is a strictly causal filter, so that the feedback operation involves only past decisions. Figure 11.4 shows the structure of the MMSE DFE. As for the LEs, a fast implementation of the DFE using the LDU decomposition is also possible here [94]: the resulting complexity in of order $MN$.

The SNR is:

$$SNR_i(\text{MMSE DFE}) = \frac{\sigma_a^2}{\sigma_v^2 (D^{-1})_{ii}} \,. \tag{11.32}$$

In the continuous processing case:

$$\hat{a}_{\text{MMSE DFE}}(k) = \frac{\mathbf{H}^\dagger(q)}{dG^\dagger(q)} \boldsymbol{y}_k - (G(q) - 1)\text{dec}(\hat{a}_{\text{MMSE DFE}}(k)) \tag{11.33}$$

where $\mathbf{H}^\dagger(q)\mathbf{H}(q) + \frac{\sigma_v^2}{\sigma_a^2} = G^\dagger(q) d G(q)$, $G(q)$ is causal and $G(\infty) = 1$.

### The Unbiased MMSE Decision Feedback Equalizer

Using the results of section 11.3.1, we can prove that the output of the Unbiased MMSE DFE is:

$$\widehat{A}_{U,\,\mathcal{U}\text{MMSE DFE}} = (I - \frac{\sigma_v^2}{\sigma_a^2} D^{-1})^{-1} \widehat{A}_{U,\,\text{MMSE DFE}} \,. \tag{11.34}$$

Figure 11.5 shows this structure. The burst output SNR is decreased by 1 with respect to the MMSE DFE.

The continuous processing equalizer output is:

$$\hat{a}_{\mathcal{U}\text{MMSE DFE}}(k) = \left(1 - \frac{\sigma_v^2}{\sigma_a^2} exp\left(-\oint \frac{dz}{z} ln(\boldsymbol{H}^\dagger(z)\boldsymbol{H}(z) + \frac{\sigma_a^2}{\sigma_v^2})\right)\right)^{-1} \hat{a}_{\text{MMSE DFE}}(k) \,. \tag{11.35}$$

### The MMSE-ZF Decision Feedback Equalizer

As for the ZF LE, there is a whole class of ZF equalizers, and we derive here the ZF MMSE DFE equalizer. Consider the LDU factorization of $\mathcal{T}_U^H \mathcal{T}_U = LDL^H$. Then the forward and feedback filters are proven in Appendix D to be:

$$\begin{cases} F = L^{-1}D^{-1}\mathcal{T}_U^H \\ B = L^H - I \end{cases} \tag{11.36}$$

we have the same structure as the MMSE DFE. The same equalizability conditions as for MMSE–ZF LE hold here also.

We conclude this section by noting that the expression of the MMSE and MMSE–ZF DFEs can be recovered from the LEs: for the MMSE LE, we consider the LDU factorization of $R = LDL^H$ and for the MMSE ZF LE, the UDL factorization of $\mathcal{T}_U^H \mathcal{T}_U = LDL^H$. The ouput of these two equalizers can then be written as:

$$\widehat{A}_U = L^{-H}D^{-1}L^{-1}\mathcal{T}_U^H \boldsymbol{Y}_U = D^{-1}L^{-1}\mathcal{T}_U^H \boldsymbol{Y} - (L^H - I)\widehat{A}_U \,. \tag{11.37}$$

The DFE operation consists in taking $(L^H - I)\mathrm{dec}(\widehat{A}_U)$ instead of $(L^H - I)\widehat{A}_U$, where dec(.) is the decision operation.

## 11.3.3 Non Causal Decision Feedback Equalizers

### The MMSE NCDFE

The NCDFE considers the linear estimation of symbol $a(i)$ based on the processing data $\boldsymbol{Y}_U$ and the past and future decisions w.r.t. $a(i)$ assumed known that we denote $\overline{A}_{U,i}$. The burst mode equalizer is implemented in an iterative way. At the first iteration, the past and future decisions come from another classical LE or DFE. The output the NCDFE can then be used to reinitialized the NCDFE, and other iterations can be done. As for the DFE, we consider the past and future decisions as correct.

$$\hat{a}_{\text{MMSE NCDFE}}(i) = F_i \boldsymbol{Y}_U - B_i \overline{A}_{U,i} \tag{11.38}$$

where $F_i$ is the forward filter and $B_i$ the feedback filter. Let $\boldsymbol{Y'} = \left[\boldsymbol{Y}_U^T \;\; \overline{A}_{U,i}^T\right]^T$;

$$\boldsymbol{Y}_U = \mathcal{T}_{U,i} a(i) + \overline{\mathcal{T}}_{U,i} \overline{A}_{U,i} \tag{11.39}$$

$$[F_i \quad -B_i] = R_{a(i)\boldsymbol{Y}'} R^{-1}_{\boldsymbol{Y}'\boldsymbol{Y}'} \tag{11.40}$$

and we get:

$$[F_i \quad B_i] = \left[\sigma_a^2 \mathcal{T}_{U,i}^H \left(\sigma_a^2 \mathcal{T}_{U,i}\mathcal{T}_{U,i}^H + \sigma_v^2 I\right)^{-1} \;\; \sigma_a^2 \mathcal{T}_{U,i}^H \left(\sigma_a^2 \mathcal{T}_{U,i}\mathcal{T}_{U,i}^H + \sigma_v^2 I\right)^{-1} \overline{\mathcal{T}}_{U,i}\right] \tag{11.41}$$

$$[F_i \quad B_i] = \left[ \left( \mathcal{T}_{U,i}^H \mathcal{T}_{U,i} + \frac{\sigma_v^2}{\sigma_a^2} I \right)^{-1} \mathcal{T}_{U,i}^H \qquad \left( \mathcal{T}_{U,i}^H \mathcal{T}_{U,i} + \frac{\sigma_v^2}{\sigma_a^2} I \right)^{-1} \mathcal{T}_{U,i}^H \overline{\mathcal{T}}_{U,i} \right] . \tag{11.42}$$

Then,

$$\begin{cases} F = \left( \text{diag} \left( \mathcal{T}_U^H \mathcal{T}_U + \frac{\sigma_v^2}{\sigma_a^2} I \right) \right)^{-1} \mathcal{T}_U^H \\ B = \left( \text{diag} \left( \mathcal{T}_U^H \mathcal{T}_U + \frac{\sigma_v^2}{\sigma_a^2} I \right) \right)^{-1} \left( \mathcal{T}_U^H \mathcal{T}_U - \text{diag} \left( \mathcal{T}_U^H \mathcal{T}_U \right) \right) \end{cases} \tag{11.43}$$

The MMSE NCDFE has a very simple structure: the forward filter is proportional to the matched filter and the feedback filter to the cascade of the channel and the forward filter without the central coefficient. Figure 11.6 shows the structure of the NCDFE.

All the ISI is removed if there are no errors in the non causal feedback: the NCDFE attains then the matched filter bound. But, like the decision feedback equalizer, the NCDFE suffers from the error propagation phenomenon.

The burst mode SNR is:

$$SNR_i(\text{MMSE NCDFE}) = \frac{\sigma_a^2}{\sigma_v^2} \left( \mathcal{T}_U^H \mathcal{T}_U + \frac{\sigma_v^2}{\sigma_a^2} I \right)_{i,i} . \tag{11.44}$$

In the continuous processing case:

$$\hat{a}_{\text{MMSE NCDFE}}(k) = \left( \mathbf{H}^\dagger(q)\mathbf{H}(q) + \tfrac{\sigma_v^2}{\sigma_a^2} \right)^{-1} \left( \mathbf{H}^\dagger(q)\boldsymbol{y}(k) - (\mathbf{H}^\dagger(q)\mathbf{H}(q) - \|\mathbf{H}\|^2)\text{dec}(\hat{a}_{\text{MMSE NCDFE}}(k)) \right) . \tag{11.45}$$

### The Unbiased/ZF-MMSE NCDFE

As seen in section 11.3.1, the Unbiased MMSE estimate $\widehat{A}_{\mathcal{U}\text{MMSE NCDFE}}$ is a scale version of $\widehat{A}_{\text{MMSE NCDFE}}$. We find:

$$\begin{cases} F = \left( \text{diag} \left( \mathcal{T}_U^H \mathcal{T}_U \right) \right)^{-1} \mathcal{T}_U^H \\ G = \left( \text{diag} \left( \mathcal{T}_U^H \mathcal{T}_U \right) \right)^{-1} \left( \mathcal{T}_U^H \mathcal{T}_U - \text{diag} \left( \mathcal{T}_U^H \mathcal{T}_U \right) \right) \end{cases} \tag{11.46}$$

As all the ISI is removed by the NCDFE, the ZF NDFE and the $\mathcal{U}$MMSE NCDFE are the same. If $n_1 = n_2 = N - 1$, the burst mode filters of the NCDFE are time–invariant. When $n_1 < N-1$ and $n_2 < N-1$, the filters vary only at the edges and are otherwise time–invariant. It appears that another interest of the burst mode NCDFE is that it is as easy to implement than its continuous processing version.

The burst mode SNR is:

$$SNR_i(\mathcal{U}/\text{ZF-MMSE NCDFE}) = \frac{\sigma_a^2}{\sigma_v^2} \left( \mathcal{T}_U^H \mathcal{T}_U \right)_{i,i} . \tag{11.47}$$

No special conditions are required for the ZF and MMSE NCDFEs to be defined as $\text{diag}\left(\mathcal{T}_U^H\mathcal{T}_U\right)$ and $\text{diag}\left(\mathcal{T}_U^H\mathcal{T}_U + \frac{\sigma_v^2}{\sigma_a^2}I\right)$ are invertible.

## 11.4 Performance Comparisons

In this section, we discuss the performance of the equalizers in terms of SNR and probability of error. In figure 11.7, the SNR curves are drawn for a channel $\boldsymbol{H}_1$ of length 7 with 3 subchannels which coefficients were randomly chosen ($\boldsymbol{H}_1$ is given in Appendix E). The SNR per channel is 10dB. The input symbols are drawn from a BPSK ($\sigma_a^2 = 1$) and the number of unknown input symbols in the burst is $L = 30$.

### 11.4.1 Case of no Known Symbols

In figure 11.7 (left), the case of no known symbols is shown. We notice that degradations appear at the ends of the burst. The middle symbols appear in $N$ outputs. When no symbols are known, the first and last unknown symbols of the burst appear in strictly less than $N$ outputs, so that there is less information about those symbols in the observations.

The SNR in the middle of the burst converges to the continuous processing level as the burst length increases.

### 11.4.2 Case of $N-1$ Known Symbols at Each End of the Burst

We assume now that $n_1 = n_2 = N-1$. The SNR curves are drawn in figure 11.7 (right). This time, burst processing performs better than continuous processing. The middle observations contain N symbols. After eliminating the contributions of the known symbols the outputs at the edges contain strictly less than $N$ symbols, so that there is more information on those symbols. This explains why the symbols are better estimated at both ends for the LEs.

For the DFEs, things are slightly different at the beginning of the burst: the situation is as if the feedback filter had been correctly initialized and the contribution of the past decisions removed, and as the forward filter is anticausal we tend to the continuous processing case as

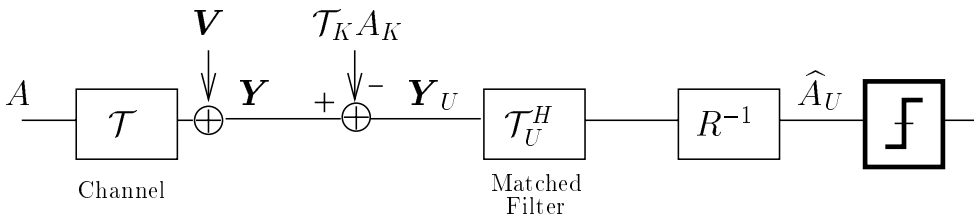



Figure 11.2: Structure of the MMSE and the MMSE-ZF LE.

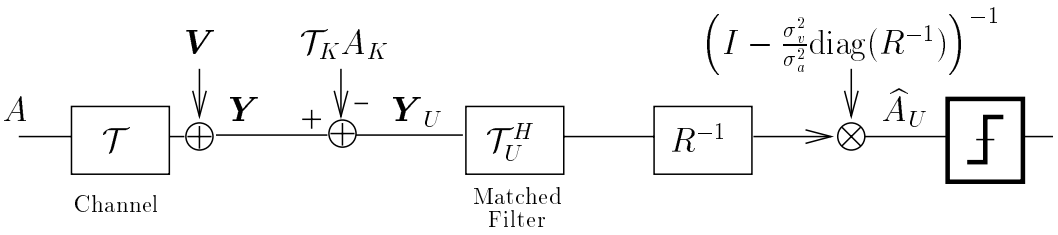



Figure 11.3: Structure of the $\mathcal{U}$MMSE LE.

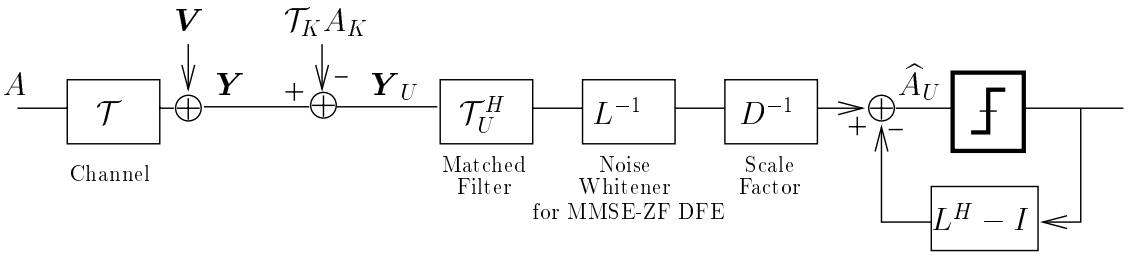



Figure 11.4: Structure of the MMSE DFE and MMSE-ZF DFE.

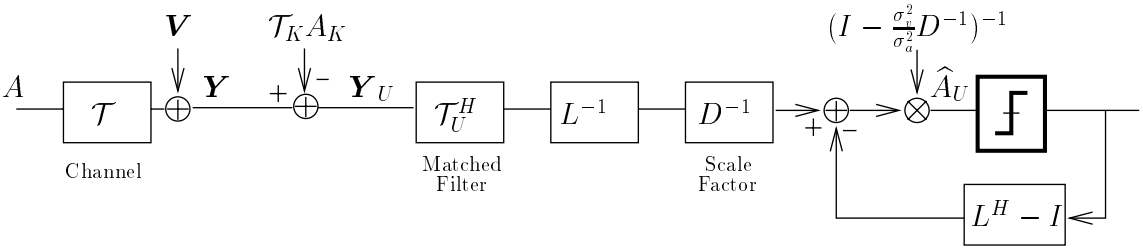



Figure 11.5: Structure of the $\mathcal{U}$MMSE DFE.

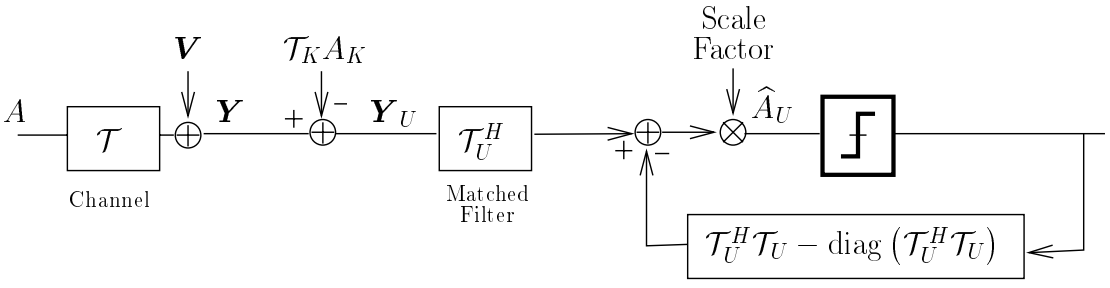



Figure 11.6: Structure of the NCDFE.

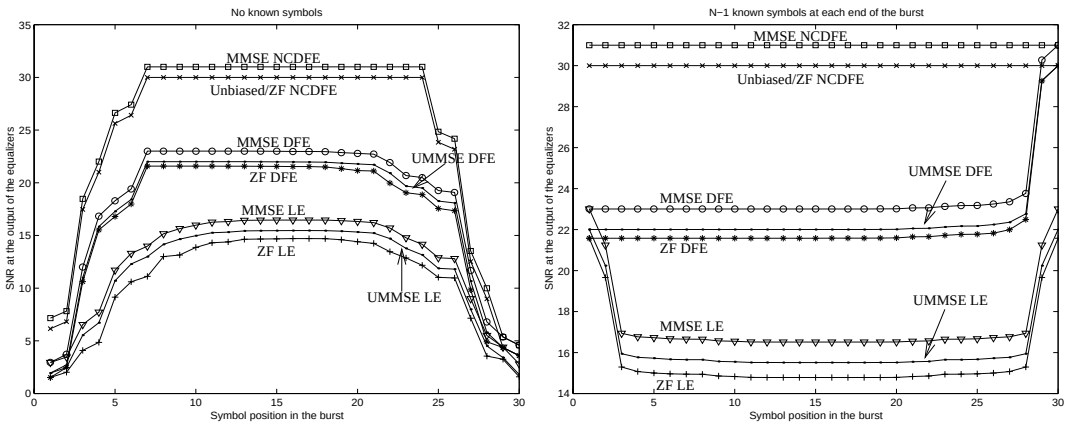



Figure 11.7: SNRs at the output of the different equalizers when no symbols are known (left) and when N-1 symbols at each end of the burst are known (right).

the number of data tends to infinity. For the last symbol of the burst, the estimation process is the same as that of the NCDFE. We notice that the NCDFE has a constant SNR over the burst equal to the one of continuous processing.

### 11.4.3 Equalizers Comparisons

#### In terms of SNR

The following comparisons are deduced from the amount of a priori information used for estimating the unknown symbols.

- Within each class of equalizers, LE, DFE, NCDFE:

$$SNR_i(\text{MMSE}) \geq SNR_i(\mathcal{U}\text{MMSE}) \geq SNR_i(\text{ZF-MMSE}) \tag{11.48}$$

- For each criterion, MMSE, $\mathcal{U}$MMSE and ZF-MMSE:

$$SNR_i(\text{NCDFE}) \geq SNR_i(\text{DFE}) \geq SNR_i(\text{LE}) \tag{11.49}$$

#### In terms of Probabilities of Error

For unbiased equalizers, a higher SNR implies a lower probability of error: MMSE ZF equalizers will then have a higher probability of error than the corresponding Unbiased MMSE equalizers. However, it is not obvious to rank the MMSE equalizers w.r.t. the ZF equalizers because they are biased. In fact people would tend to believe that a MMSE equalizer performs better than the corresponding MMSE-ZF equalizer.

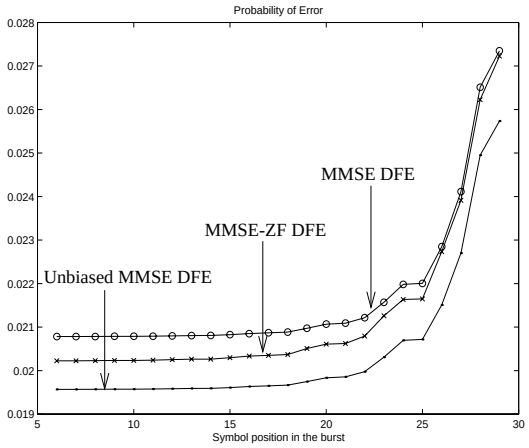



Figure 11.8: Probability of Error for the ZF-DFE, the MMSE-DFE and the Unbiased MMSE-DFE.

In the case of constant modulus modulations, MMSE equalizers have the same performance as the corresponding unbiased MMSE equalizers and so a higher performance than MMSE ZF equalizers. For non constant-modulus constellations, the bias in MMSE equalizers may have a stronger effect than its higher SNR compared to MMSE-ZF equalizers. This is all the more true as the difference in SNRs between the different equalizers tends to be lower as subchannels are added.

Figure 11.8 treats of the DFE case. We plot the probabilities of error for the channel $\boldsymbol{H}_2$ (see Appendix E) for the different DFEs. In the error probability computations of the MMSE and $\mathcal{U}$MMSE, the symbols other than the current symbol of interest are approximated as Gaussian random variables. The input symbols belong to a 4-PAM constellation. No symbols are assumed known; the number of known symbols is equal to $L = 3N$. In order to see better the difference between the different curves, we only plot the probability of error for the central coefficients. We notice here that the MMSE equalizer has poorer performance than the MMSE–ZF equalizer, and that the $\mathcal{U}$MMSE performs the best.

## 11.5 Applying Continuous Processing Equalizers to the Burst Case

As already mentioned, burst processing involves time-varying filters. We may wonder if it is worth implementing these time-varying filters, because of complexity reasons, and if simply applying the time-invariant filters corresponding to continuous processing in burst mode could give acceptable performance.

For that purpose we will consider the case of N-1 known symbols at each end of the input burst. We will show that the continuous processing filters also give better SNR at the ends

of the burst than in the middle and always give strictly better SNR than in the continuous processing case.

For the LEs, the contribution of the known symbols is removed at the end of the observation data. For the DFEs, the initialization is done by putting the N-1 leading known symbols in the memory of the feedback filter. Only the trailing known symbols are removed from the processing data.

In both cases, we put the channel outputs before and after the data to be processed equal to zero. The only difference with the continuous processing case is that we have a finite input symbol sequence, but also a finite noise sequence. As will be seen in the simulations of the next section, for the DFE, the way we proceed is equivalent to the continuous processing case at the beginning of the burst.

For the LEs, the different reasonings will be held for zero delay non-causal continuous processing filters. For the DFEs, the forward filter is assumed to be anticausal (zero delay) the feedback filter is causal and FIR (of the same length as the channel). As the channel output is zero outside the time interval of the processing data, these filters will involve only a finite number of data.

In the MMSE ZF case, the MSE contains only the noise contributions. Since the noise is only finite length, the MSE is smaller at the edges. The MSE of MMSE (unbiased or not) equalizers outputs contains residual ISI also. This variance gets also reduced as the input sequence becomes finite length.

For the NCDFE, the leading and trailing symbols are both put in the memory of the feedback filter. In this case, the optimal burst mode feedforward and feedback filters are time-invariant and are the same as the continuous mode filters. This fact reinforces the interest of the NCDFE.

### 11.5.1 MSE Calculations

The outputs of the different linear equalizers based on the continuous processing filters may be written as:

$$\widehat{A}_U = F\boldsymbol{Y}_U \tag{11.50}$$

where $F$ is a structured matrix containing the coefficients of the continuous processing filter.

In general:

$$MSE_i = (\sigma_a^2(F\mathcal{T}_U - I)(F\mathcal{T}_U - I)^H + \sigma_v^2 FF^H)_{ii} \tag{11.51}$$

where $F\mathcal{T}_U = I$ in the ZF case.

The outputs of the different DFEs be may written as:

$$\widehat{A}_U = F\boldsymbol{Y}_U - (B - I')A' \tag{11.52}$$

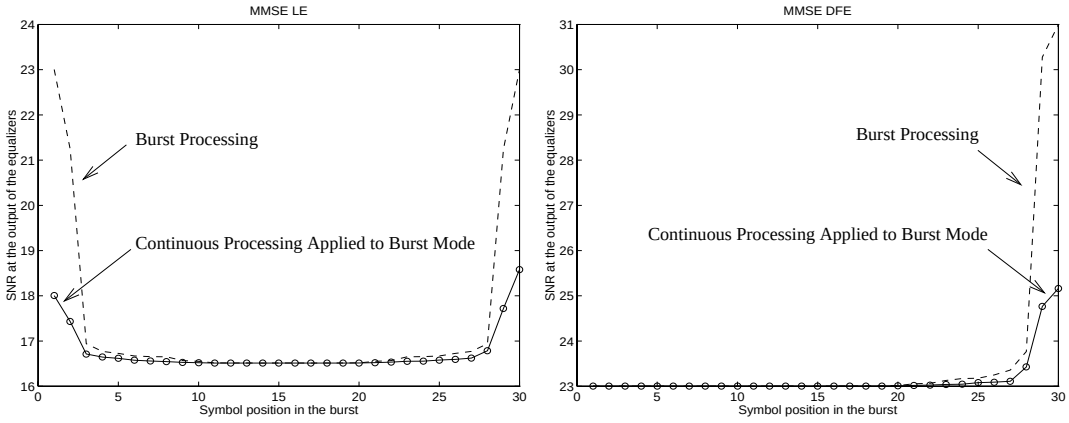



Figure 11.9: SNR Curves: optimal burst processing compared to continuous processing applied to burst mode for the MMSE LE (left) and MMSE DFE (right).

$A'$ contains $A_U$ and the leading symbols, $I' = [I \quad 0]$, $F$ contains the coefficients of the continuous processing forward filter. $B$ the coefficients of the continuous processing feedback filter.

In general:

$$MSE_i = (\sigma_a^2(F\mathcal{T}_U - B)(F\mathcal{T}_U - B)^H + \sigma_v^2 FF^H)_{ii} \tag{11.53}$$

where $F\mathcal{T}_U = B$ in the ZF case.

In figure 11.9, we present the case of the MMSE LE and MMSE DFE: we compare performances for channel $\boldsymbol{H}_1$. The input symbols are drawn from a BPSK. The length of the filters for the LEs and for the feedforward filter for the DFE is equal to $3N$. The number of unknown input symbols is $L = 30$.

## 11.6 Conclusion

We have derived the optimal structure of the burst mode equalizers for three classes of equalizers: linear, decision feedback and non causal decision feedback equalizers. Three different criteria have been considered: the MMSE, the unbiased MMSE, and Zero Forcing. The problem of finding the equalizer filters have been formulated in terms of linear estimation based on the data and certain a priori information, which allows a simple classification of the equalizers in terms of performance. The SNR degradations have been studied as a function of the position of the unknown symbols in the bursts and as a function of the presence of known symbols. The more favorable situation for burst mode is when pre– and postamble sequence of known symbols are attached at each end of the burst: in this case burst mode equalization performs better than continuous processing. At last we have shown how time–

varying burst mode filters can be approximated by time–invariant filters in the situation where the pre and postamble sequences have the same length as the channel memory: time–invariant filters still have better performance than the continuous processing level and allows a lower complexity for implementation than the time–varying optimal burst mode filters. The case of the NCDFE appears also of particular interest: it is potentially the most powerful equalizer as it can eliminate all the ISI, and has a particularly simple structure. In particular, when the pre and postamble sequences have the same length as the channel memory, the NCDFE filters are time–invariant.

## A Linear MMSE Estimation in terms of $\boldsymbol{Y}_U$

$$R_{A_u\boldsymbol{Y}'}R^{-1}_{\boldsymbol{Y}'\boldsymbol{Y}'}\boldsymbol{Y}' = R_{A_u\boldsymbol{Y}'}R^{-1}_{\boldsymbol{Y}'\boldsymbol{Y}'}\left(\begin{bmatrix}\boldsymbol{Y}_u\\0\end{bmatrix}+\begin{bmatrix}\mathcal{T}_K A_K\\A_K\end{bmatrix}\right) \tag{11.54}$$

$$R_{A_U\boldsymbol{Y}'} = \begin{bmatrix}\sigma_a^2\mathcal{T}_U^H & 0\end{bmatrix} = \begin{bmatrix}R_{A_U\boldsymbol{Y}_U} & 0\end{bmatrix} \tag{11.55}$$

$$R_{\boldsymbol{Y}'\boldsymbol{Y}'} = \begin{bmatrix}\sigma_a^2\mathcal{T}\mathcal{T}^H+\sigma_v^2 I & \sigma_a^2\mathcal{T}_K\\ \sigma_a^2\mathcal{T}_K^H & \sigma_a^2 I\end{bmatrix} \tag{11.56}$$

The element (1,1) of $R^{-1}_{\boldsymbol{Y}'\boldsymbol{Y}'}$ is:

$$R^{-1}_{\boldsymbol{Y}'\boldsymbol{Y}'}(1,1) = \left(\sigma_a^2\mathcal{T}_U\mathcal{T}_U^H+\sigma_a^2\mathcal{T}_K\mathcal{T}_K^H+\sigma_v^2 I-\sigma_a^2\mathcal{T}_K\mathcal{T}_K^H\right)^{-1} = R^{-1}_{\boldsymbol{Y}_U\boldsymbol{Y}_U} \tag{11.57}$$

The element (1,2) of $R^{-1}_{\boldsymbol{Y}'\boldsymbol{Y}'}$ is:

$$R^{-1}_{\boldsymbol{Y}'\boldsymbol{Y}'}(1,2) = -R^{-1}_{\boldsymbol{Y}'\boldsymbol{Y}'}(1,1)\mathcal{T}_K = -R^{-1}_{\boldsymbol{Y}_U\boldsymbol{Y}_U}\mathcal{T}_K \tag{11.58}$$

Then:

$$R_{A_u\boldsymbol{Y}'}R^{-1}_{\boldsymbol{Y}'\boldsymbol{Y}'}\begin{bmatrix}\boldsymbol{Y}_u\\0\end{bmatrix} = R_{A_U\boldsymbol{Y}_U}R^{-1}_{\boldsymbol{Y}'\boldsymbol{Y}'}(1,1)\boldsymbol{Y}_U = R_{A_U\boldsymbol{Y}_U}R^{-1}_{\boldsymbol{Y}_U\boldsymbol{Y}_U}\boldsymbol{Y}_U \tag{11.59}$$

and

$$R_{A_u\boldsymbol{Y}'}R^{-1}_{\boldsymbol{Y}'\boldsymbol{Y}'}\begin{bmatrix}\mathcal{T}_K A_K\\A_K\end{bmatrix} = R_{A_u\boldsymbol{Y}'}R^{-1}_{\boldsymbol{Y}_U\boldsymbol{Y}_U}\begin{bmatrix}I & -\mathcal{T}_K\end{bmatrix}\begin{bmatrix}\mathcal{T}_K A_K\\A_K\end{bmatrix} = 0 \tag{11.60}$$

Then, the have the result:

$$R_{A_u\boldsymbol{Y}'}R^{-1}_{\boldsymbol{Y}'\boldsymbol{Y}'}\boldsymbol{Y}' = R_{A_u\boldsymbol{Y}_U}R^{-1}_{\boldsymbol{Y}_U\boldsymbol{Y}_U}\boldsymbol{Y}_U \tag{11.61}$$

## B SNR of the Unbiased MMSE

We prove that:

$$SNR_i(\mathcal{U}\text{MMSE}) = SNR_i(\text{MMSE}) - 1. \tag{11.62}$$

From expression (11.16), we deduce:

$$\left(R_{\widehat{A}_{U,\,\mathcal{U}MMSE}\widehat{A}_{U,\,\mathcal{U}MMSE}}\right)_{i,i} = \sigma_a^2\mathcal{D}_{i,i} \qquad \text{and} \qquad \left(R_{\widehat{A}_{U,\,\mathcal{U}MMSE}A_U}\right)_{i,i} = \sigma_a^2 \tag{11.63}$$

$$\begin{aligned} R_{\widetilde{A}_{U,\,\mathcal{U}MMSE}\widetilde{A}_{U,\,\mathcal{U}MMSE}} &= \mathrm{E}\left(A_U-\widehat{A}_{U,\mathcal{U}MMSE}\right)\left(A_U-\widehat{A}_{U,\mathcal{U}MMSE}\right)^H\\ &= R_{A_UA_U}-R_{A_U\widehat{A}_{U,\,\mathcal{U}MMSE}}-R_{\widehat{A}_{U,\,\mathcal{U}MMSE}A_U}+R_{\widehat{A}_{U,\,\mathcal{U}MMSE}\widehat{A}_{U,\,\mathcal{U}MMSE}} \end{aligned} \tag{11.64}$$

and we find:

$$MSE_i(\mathcal{U}\text{MMSE}) = \sigma_a^2 \left(\mathcal{D}_{i,i} - 1\right) \tag{11.65}$$

Noting that:

$$\sigma_a^2 \mathcal{D}_{i,i}^{-1} = \sigma_a^2 - MSE_i(\text{MMSE}) \tag{11.66}$$

$$MSE_i(\mathcal{U}\text{MMSE}) = \sigma_a^2 \frac{MSE_i(\text{MMSE})}{\sigma_a^2 - MSE_i(\text{MMSE})} \tag{11.67}$$

from which we get (11.62).

## C Filters of the MMSE DFE

We derive the expressions of the feedforward filter $F$ and feedback filter $B$ of the MMSE-DFE. We do not use directly expression (11.30), but rather a slightly more elegant way.

The expression of the estimate of the unknown symbols by the MMSE-DFE is:

$$\widehat{A}_{U,\,\text{MMSE DFE}} = F\boldsymbol{Y}_U - BA_U \tag{11.68}$$

with $B$ strictly triangular inferior. The MMSE criterion writes as:

$$\min_{F,B} \| A_U - (F\boldsymbol{Y}_U - BA_U) \|^2 \tag{11.69}$$

and

$$A_U - \widehat{A}_{U,\,\text{MMSE DFE}} = (I+B)A_U - F\boldsymbol{Y}_U \tag{11.70}$$

Using the orthogonality principle, which states that the error on $A_U$ should be orthogonal to $\boldsymbol{Y}_U$, we find:

$$(I+B)\sigma_a^2 \mathcal{T}_U^H - F R_{\boldsymbol{Y}_U \boldsymbol{Y}_U} = 0 \tag{11.71}$$

from which we get:

$$F = \sigma_a^2 (I+B)\mathcal{T}_U^H R_{\boldsymbol{Y}_U \boldsymbol{Y}_U}^{-1} \Rightarrow F = (I+B)(\mathcal{T}_U^H \mathcal{T}_U + \frac{\sigma_v^2}{\sigma_a^2} I)^{-1} \mathcal{T}_U^H \tag{11.72}$$

and

$$A_U - \widehat{A}_{U,\,\text{MMSE DFE}} = (I+B)(A_U - \widehat{A}_{U,\,\text{MMSE LE}}) \tag{11.73}$$

Then:

$$\text{E}\|A_U - A_{U,\,\text{MMSE DFE}}\|^2 = \sigma_v^2 \text{tr}\left\{(I+B)R^{-1}(I+B)^H\right\} \tag{11.74}$$

(we recall that $R = \mathcal{T}_U^H \mathcal{T}_U + \frac{\sigma_v^2}{\sigma_a^2} I$). Consider the LDU decomposition of $R$:

$$R = LDL^H, \tag{11.75}$$

$$\mathrm{E}\|A_U - A_{U,\ \mathrm{MMSE\ DFE}}\|^2 = \sigma_v^2 \mathrm{tr}\left\{(I+B)L^{-H}D^{-1}L^{-1}(I+B)^H\right\} \tag{11.76}$$

The minimization problem:

$$\min_{\{C:\ \mathrm{diag}(C)=I\}} \mathrm{tr}\left\{C^H D^{-1} C\right\} \tag{11.77}$$

as for solution $C = I$.

Then (11.76) is minimized when $I + B = L^H$. Then:

$$\begin{cases} F = D^{-1}L^{-1}\mathcal{T}_U^H \\ B = L^H - I \end{cases} \tag{11.78}$$

# D Filters of the MMSE–ZF DFE

We derive the expressions of the feedforward filter $F$ and feedback filter $B$ of the MMSE–ZF DFE. We want to solve:

$$\begin{array}{c} \min \\ F, B \\ F\mathcal{T}_U - B = I \end{array} \quad \mathrm{E}\|A_U - (F\boldsymbol{Y}_U - BA_U)\|^2 \tag{11.79}$$

and with constraint that $B$ be strictly triangular inferior. Let the following decomposition of $F$ onto the rows of $\mathcal{T}_U^H$ and its orthogonal complement $\mathcal{T}_U^{H\perp}$ (the rows of $\mathcal{T}_U^{H\perp}$ span the orthogonal complement of the rows of $\mathcal{T}_U^H$):

$$F = F_1\mathcal{T}_U^H + F_2\mathcal{T}_U^{H\perp}, F_2\mathcal{T}_U^{H\perp}\mathcal{T}_U^H = 0 \Rightarrow F_1\mathcal{T}_U^H\mathcal{T}_U = I + B \Rightarrow F_1 = (I+B)\left(\mathcal{T}_U^H\mathcal{T}_U\right)^{-1} \tag{11.80}$$

$$\Leftrightarrow \quad \begin{array}{c} \min \\ F, B \\ F\mathcal{T}_U - B = I \end{array} \quad \mathrm{E}\|F\boldsymbol{V}\|^2 \Leftrightarrow \quad \begin{array}{c} \min \\ F, B \\ F\mathcal{T}_U - B = I \end{array} \quad \sigma_v^2\|FF^H\|^2 \tag{11.81}$$

$$\|FF^H\|^2 = \mathrm{tr}\left\{F_1\mathcal{T}_U^H\mathcal{T}_U F_1^H\right\} + \mathrm{tr}\left\{F_2\mathcal{T}_U^{H\perp}\mathcal{T}_U^{H\perp H}F_2^H\right\} \tag{11.82}$$

(11.82) gives $F_2 = 0$:

$$(11.79) \Leftrightarrow \min_B (I+B)\left(\mathcal{T}_U^H\mathcal{T}_U\right)^{-1}(I+B)^H \tag{11.83}$$

Considering now the LDU factorization of $\mathcal{T}_U^H\mathcal{T}_U$, we obtain as in appendix 3:

$$\begin{cases} F = D^{-1}L^{-1}\mathcal{T}_U^H \\ B = L^H - I \end{cases} \tag{11.84}$$

# E Channels used in the Simulations

The channels used in the simulations are the following:

$$\boldsymbol{H}_1 = \begin{bmatrix} -0.7989 & -0.0562 & 0.7562 & 0.3750 & -2.3775 & 0.3180 & 1.6065 \\ -0.7652 & 0.5135 & 0.4005 & 1.1252 & -0.2738 & -0.5112 & 0.8476 \\ 0.8617 & 0.3967 & -1.3414 & 0.7286 & -0.3229 & -0.0020 & 0.2681 \end{bmatrix} \tag{11.85}$$

$$\boldsymbol{H}_2 = \begin{bmatrix} -0.6776 & -0.4710 & 0.4992 & 0.1558 & -0.7209 \\ 0.4617 & -0.5649 & 0.3827 & -0.5692 & 0.3998 \\ -1.1939 & -0.4239 & -0.0136 & -0.7488 & -1.3747 \end{bmatrix}$$

Chapter 12

# BURST MODE NON-CAUSAL DFE BASED ON SOFT DECISIONS

*The Non-Causal Decision-Feedback Equalizer (NCDFE) is a decision-aided equalizer that uses not only past decisions, like DFEs, but also future decisions, which usually come from another, classical equalizer. When there are no errors on the decisions, the NCDFE output contains no ISI and the Matched Filter bound (MFB) is attained. In practice, it suffers from the propagation of errors. We propose a multiple stage implementation of the NCDFE based on soft decisions. At each stage, reliability intervals are defined based on the SNR, the position of the symbol in the burst and the presence of known symbols in the burst: only the reliable decisions are fed back. The other, non–reliable symbols are classified as unknown and again estimated in the next iteration. The augmented presence of known symbols coming from the previous decision stage results in a better estimation quality of the unknown symbols. This soft strategy scheme is compared to other schemes.*

## 12.1 Introduction

The principle of the Non-Causal Decision-Feedback Equalizer (NCDFE) was first proposed by Proakis [89]: this equalizer uses past and future decisions in order to cancel all the ISI present in the signal. Gersho and Lim [90] introduced its MMSE design: the forward filter is proportional to the matched filter and the feedback filter applied to the past and future symbol decisions w.r.t. the symbol to be detected, is the cascade of the channel and the forward filter, without the central coefficient. These past and future symbol decisions come from another classical equalizer, linear or DFE (note that the past decisions may come from the NCDFE itself). A burst mode unbiased MMSE version based on MLSE was also proposed in [91].

When no errors on the past and future decisions are made, we reach the ideal zero ISI situation and the equalizer attain the Matched Filter Bound (MFB): it is then potentially more powerful than the other equalizers, linear or DFEs. It possesses furthermore a particularly simple structure with only FIR feedforward and feedback filters. As highlighted in the previous chapter, in its burst mode implementation, the optimal filters are almost time-invariant (they are only time-varying at the edges of the burst) when no known symbols are present in the burst and are time–invariant when known symbols of a number of at least the channel memory are present at each end of the burst. The error propagation phenomenon can cause some degradations, however, like for the classical DFE, and the NCDFE may offer only a marginal improvement or even degradations.

Our purpose is to build a non-causal decision-feedback equalizer by replacing the hard decision scheme by a soft decision one, allowing one then to reduce the feedback errors. Our soft decision scheme consists in taking a decision on the most reliable symbols and leaving the other non reliable symbols undecided. We will see that this scheme is equivalent to feeding back only the most reliable symbols. The NCDFE is implemented in multiple stages. Each iteration finds a linear estimate of the unknown symbols based on the received signal, the known symbols (*i.e.* , the hard decisions) computed in the previous iterations, and the past hard decisions of the current iteration. Due to the augmented presence of hard decisions/known symbols at each iteration, the reliability of the unknown estimate increases. The first stage corresponds then to a (classical) decision feedback equalizer in which only the most reliable past decisions are fed back. The following stages mix a decision and non-causal decision feedback strategy. We will consider here only a BPSK; the principle of soft decisions as explained in this chapter can be extended to other constellations though. We will furthermore consider only Unbiased linear symbols estimation, *i.e.* Unbiased MMSE and ZF-MMSE.

We proposed this soft decision scheme in [92]; it is similar to the "clipped" soft decisions [98, 99] applied to multi–user detection but was independently developed. In multi–user detection, soft decision schemes are becoming very popular and the benefit of soft decisions

can be quite visible. In the single–user context considered here its effects are probably less dramatic but remain of interest.

## 12.2 Burst Mode NCDFE

Let us recall the structure of the burst mode NCDFE given in figure 12.1. The forward filter is the multichannel matched filter $\mathcal{T}^H(h)$ followed by a scaling operation. $D$ is a diagonal matrix which differs according to the nature of the equalizer, MMSE or $\mathcal{U}$MMSE: for the $\mathcal{U}$MMSE NCDFE [91], $D = \left(\text{diag}(\mathcal{T}^H(h)\mathcal{T}(h))\right)^{-1}$. The non-causal feedback filter consists of the forward filter without the central coefficient. $\widehat{A}$ may be the output of another equalizer or the output of the burst mode NCDFE at a previous iteration. If $\widehat{A}$ contains no errors, the performance of the NCDFE attains the MFB.

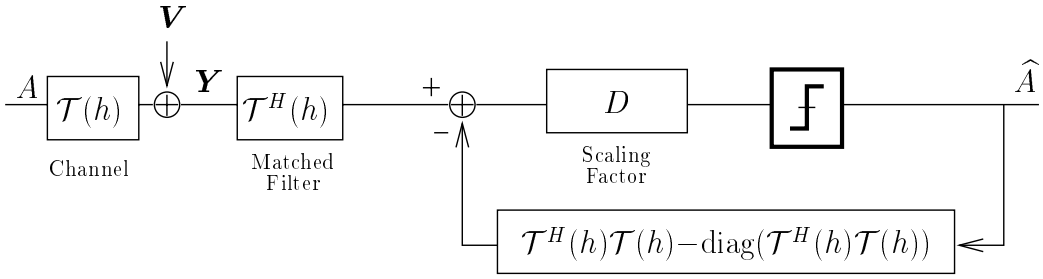



Figure 12.1: Burst Mode Non Causal DFE.

## 12.3 Non Causal Soft Feedback

### 12.3.1 Linear Equalizers seen as Non–Causal DFEs

Consider the class of Linear Equalizers. The output symbol estimates can be written in the general form:

$$\widehat{A}_{LE} = \mathcal{R}^{-1}\mathcal{T}^H(h)\boldsymbol{Y}\,. \tag{12.1}$$

For a ZF–MMSE, $\mathcal{R} = \mathcal{T}^H(h)\mathcal{T}(h)$, for a MMSE, $\mathcal{R} = \mathcal{T}^H(h)\mathcal{T}(h) + \dfrac{\sigma_v^2}{\sigma_a^2}I$ and for a $\mathcal{U}$MMSE,

$$\mathcal{R} = \left(I - \frac{\sigma_v^2}{\sigma_a^2}\text{diag}\left(\mathcal{T}^H(h)\mathcal{T}(h) + \frac{\sigma_v^2}{\sigma_a^2}I\right)^{-1}\right)^{-1}\left(\mathcal{T}^H(h)\mathcal{T}(h) + \frac{\sigma_v^2}{\sigma_a^2}I\right).$$

$$(12.1) \quad \Leftrightarrow \quad \mathcal{R}\,\widehat{A}_{LE} = \mathcal{T}^H(h)\boldsymbol{Y}\,. \tag{12.2}$$

Assume we want to estimate the symbol $a(k)$ and assume the hypothetic situation where the linear estimates of the past and future symbols w.r.t. $a(k)$ are known. Denoting $\mathcal{R}\widehat{A}_{LE} =$

$\mathcal{R}_k a(k) - \overline{\mathcal{R}}_k \overline{\widehat{A}}_k$:

$$\hat{a}(k) = \left(\mathcal{R}_k^H \mathcal{R}_k\right)^{-1} \mathcal{R}_k^H \mathcal{T}^H(h) \boldsymbol{Y} - \left(\mathcal{R}_k^H \mathcal{R}_k\right)^{-1} \mathcal{R}_k^H \overline{\mathcal{R}}_k \overline{\widehat{A}}_k \,. \tag{12.3}$$

A linear equalizer can then be seen as a NCDFE where the past and future linear estimates are fed back. The structure of a DFE can be deduced from (12.2) by considering, for each symbol to estimate, the past symbol decisions as known (as well as the future symbol estimates, *i.e.* the estimates given by the DFE before decision, but we will see that these estimates do not intervene in the estimation of the symbol). In the NCDFE, the past and future decisions are assumed known.

Our soft strategy consists in replacing the linear estimates $\overline{\widehat{A}}_k$ by soft decisions that will be a mixture of linear and hard decisions.

### 12.3.2 Soft Decision Scheme

We propose an iterative scheme with each iteration composed of two steps.

1. In the first step, we perform linear estimation of the symbols based on the received data and the symbol estimates from the previous iteration. This first step would correspond to the NCDFE if the symbol estimates were perfect.

2. The second step performs element-wise nonlinear estimation, and iteration 1 is repeated.

The linear estimation step will be given by an unbiased equalizer, *i.e.* , a ZF or $\mathcal{U}$MMSE equalizer: each output can be put in the form $\hat{a}(k) = a(k) + \tilde{a}(k)$: for a ZF equalizer, $\hat{a}(k)$ contains only noise terms and for a $\mathcal{U}$MMSE noise and interference terms that we will approximate as a zero–mean Gaussian variable.

The optimal nonlinearity to be used in the second step is the hyperbolic tangent function $\tanh(.)$. Indeed, the MMSE estimate of $\hat{a} = a + \tilde{a}$, with $a$ taking with equal probability the values $+1$ and $-1$ and $\tilde{a}$ a zero–mean Gaussian random variable independent of $a$, hypotheses verified (with the Gaussian approximation) in our problem, is:

$$\hat{\hat{a}} = \tanh\left(\frac{\hat{a}}{\sigma^2_{\tilde{a}}}\right) \,. \tag{12.4}$$

However, with such nonlinear symbol estimates, the design of the linear estimator for step 1 in the next iteration becomes nontrivial. Therefore, we propose the following simplified nonlinearity:

$$\hat{\hat{a}} = f_\alpha(\hat{a}) = \begin{cases} \hat{a} & |\hat{a}| < \alpha \\ sign(\hat{a}) & |\hat{a}| \geq \alpha \,. \end{cases} \tag{12.5}$$

The question is to know where the linearity should be (*i.e.* what should the value of $\alpha$ be) in order to get an improvement w.r.t. the linear estimates of step 1 and a hard decision step 2. We do not answer this question here. Our guess would be that the linearity should be in between the linear and the hard decision curve, as indicated in figure 12.2.

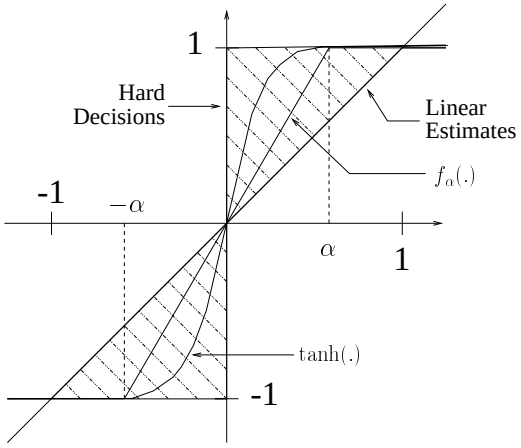



Figure 12.2: Soft Decision Curves.

## 12.4 Soft Scheme as an Approximation of the Optimal Scheme tanh(.)

The variable $\alpha$ gives the reliability of the symbol estimate and depends on $\sigma^2_{\hat{a}}$, which depends itself on the SNR conditions and other characteristics detailed later. Usually the soft decision schemes that are based on the reliability principle (see section 12.7) use a constant threshold that does not depend on the experimental conditions; for example, in [98, 99], $\alpha = 1$. Here, the threshold will be adapted to the experimental conditions: this will allow us to incorporate more hard decisions in the feedback, with high confidence, resulting in better performance.

We determine $\alpha$ by seeking the best MMSE estimate of $\hat{a}$ of the form $f_\alpha(\hat{a})$ shown in figure 12.3 (left):

$$\min_{\alpha} \mathrm{E}\left(a - f_\alpha(\hat{a})\right)^2 \ . \tag{12.6}$$

A closed form expression for $\alpha$ could not be found. However a linear approximation w.r.t. $\sigma^2_{\hat{a}}$ seemed to match well, especially for low $\sigma^2_{\hat{a}}$: $\alpha = 1.33\sigma^2_{\hat{a}}$ (see figure 12.3 (right)).

The complete iterative scheme is depicted in figure 12.4. $\widehat{A}_{soft,i}$ denotes the $\hat{a}$ for which $|\hat{a}| < \alpha$, whereas $\widehat{A}_{hard,i}$ denotes the $\hat{a}$ for which $|\hat{a}| = 1$. $\widehat{A}_{hard}(i)$ denotes the accumulation of $\widehat{A}_{hard,1}, \ldots, \widehat{A}_{hard,i}$. $\widehat{A}_{lin,i}$ is a linear combination of $\widehat{A}_i = \left\{\widehat{A}_{soft,i}, \widehat{A}_{hard}(i)\right\}$ and $\boldsymbol{Y}$, i.e. $\widehat{A}_{lin,i}$ is a linear estimate of the remaining undecided symbols in terms of the received data and soft decisions for all symbols. One can observe that $\widehat{A}_{lin,i}$ is in fact also a linear combination of only $\widehat{A}_{hard}(i)$ and $\boldsymbol{Y}$ and since the $\widehat{A}_{hard}(i)$ are assumed to be error-free, the MMSE design of $\widehat{A}_{lin,i}$ becomes tractable.

This explanation means also that when you feed back the linear symbol estimates of step 1, it does change the linear estimates in the following step.

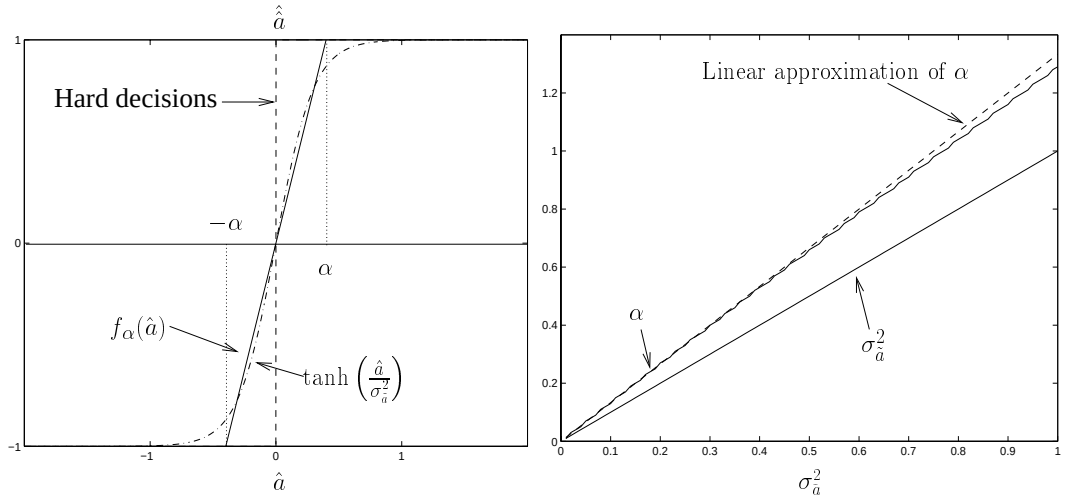



Figure 12.3: Soft Decision Curves (left) and Linear approximation of $\alpha$ (right).

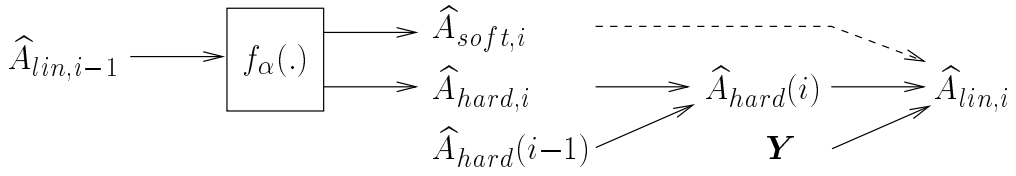



Figure 12.4: Iterative Soft Decision Scheme.

## 12.5 NCDFE Based on Soft Decisions

The soft implementation of the NCDFE will mix a causal and a non–causal soft decision–feedback strategy. The initialization is provided by a decision–feedback equalizer with feedback of the reliable decisions only.

For each symbol $a(k)$, the implementation of the NCDFE based on the soft decisions is as follows:

1. Linear estimation of the unknown symbols $a(k)$ based on:
   - the observations $\boldsymbol{Y}$
   - the "known" symbols made of the true known symbols, the hard decisions of the previous iteration, the past hard decisions of the actual iteration.
2. The reliability measure $\alpha(k)$ is computed and the soft decision strategy (12.5) is applied to $\hat{a}(k)$. The hard decisions are treated as known symbols.

The initial DFE step is the same except that the linear estimate of step 1 is computed from $\boldsymbol{Y}$ and the past decisions of the actual iteration. Steps 1-2 are reiterated until $\widehat{A}_{hard,i}$ is empty. At the end of this process, the symbols that remain non reliable even when using the feedback from known symbols are decided upon. Few iterations of the algorithm are necessary in general as will be seen in the simulations.

The performance of this soft scheme relies on 2 ideas:

- Feeding back only the most reliable symbols helps to avoid the phenomenon of error propagation.
- The presence of known symbols allows to increase the estimation quality of the unknown symbols.

### 12.5.1 Computation of $\alpha(k)$

Recalling results from the previous chapter, the error variance for each symbols is computed as, for the $\mathcal{U}$MMSE:

$$\frac{1}{\sigma_{\tilde{a}}^2(k)} = \frac{1}{\sigma_v^2 \left( \left( \mathcal{T}_U^H(h)\mathcal{T}_U(h) + \frac{\sigma_v^2}{\sigma_a^2} I \right)^{-1} \right)_{k,k}} - 1 \tag{12.7}$$

and for the ZF–MMSE:

$$\sigma_{\tilde{a}}^2(k) = \sigma_v^2 \left( \left( \mathcal{T}_U^H(h)\mathcal{T}_U(h) \right)^{-1} \right)_{k,k}. \tag{12.8}$$

For simplicity reasons, we do not take into account the past hard decisions of the current iterations and used in the causal feedback; we only take into account the hard decisions of the previous iteration.

When no known symbols are present in the burst, fast algorithms allow us to compute the vectors of the $\sigma^2_{\tilde{a}}(k)$'s linearly in $M$. A simplification consists in computing the continuous–processing error variance, neglecting then edge phenomena. After a first soft decision step, in general, few symbols remain unknown: $\mathcal{T}_U$ has a small column dimension and $\sigma^2_{\tilde{a}}(k)$ can be computed using the previous expressions without requiring an intensive computational effort.

### 12.5.2 Influence of the Known Symbols

In the previous chapter, the influence of the known symbols in a burst on the estimation of the unknown symbols was already studied. As an example, in figure 12.5, we show the SNR at the output of the MMSE Linear Equalizer (LE). Our soft design uses 2 important properties of the burst mode equalization:

- The SNR depends on the position of the symbol on the burst.
- For a given symbol, the SNR is higher when there are known symbols in the burst and especially when the symbol is surrounded by known symbols.

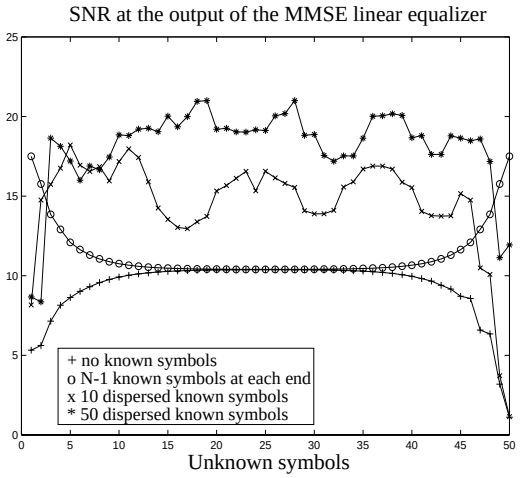



Figure 12.5: SNR at the output of a MMSE LE as a function of unknown symbol position in the burst: influence of the presence of known symbols on the estimation of the unknown symbols.

In figure 12.5, we show the case of no known symbols, grouped known symbols and arbitrarily dispersed known symbols for a number of unknown symbols of 50: we can see the advantage

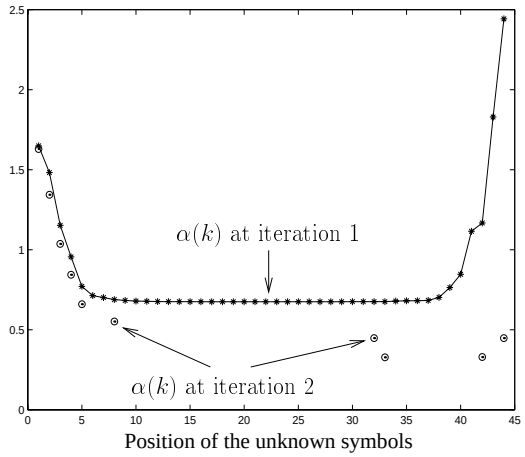



Figure 12.6: Evolution of the reliability intervals.

of taking into account the presence of known symbols in the burst to estimate the unknown symbols.

### 12.5.3 Adaptation of the Reliability Intervals

This strategy allows one to automatically adapt the reliability intervals to the experimental conditions:

- The noise level: $\alpha(k)$ is all the larger as the noise level is large.
- The presence of known symbols will be reflected in the value of $\alpha(k)$. Figure 12.6 shows the evolution of the reliability intervals from one iteration to the next (for a randomly chosen channel at 5 dB): most of the symbols that remain unknown at the second iteration are located at the edges where indeed performance is lower. At the second iteration, the reliability of those symbols increases due to the feedback of the known symbols.

## 12.6 Fast Implementation of the Soft DFE and NCDFE

### 12.6.1 Soft DFE

We start from the equation:

$$\mathcal{R}\widehat{A} = \mathcal{T}^H(h)\boldsymbol{Y} \tag{12.9}$$

and consider the $LDU$ decomposition of $\mathcal{R} = LDL^H$, then:

$$(12.9) \quad \Leftrightarrow \quad L^H\widehat{A} = D^{-1}L^{-1}\mathcal{T}^H(h)\boldsymbol{Y} = Y'. \tag{12.10}$$

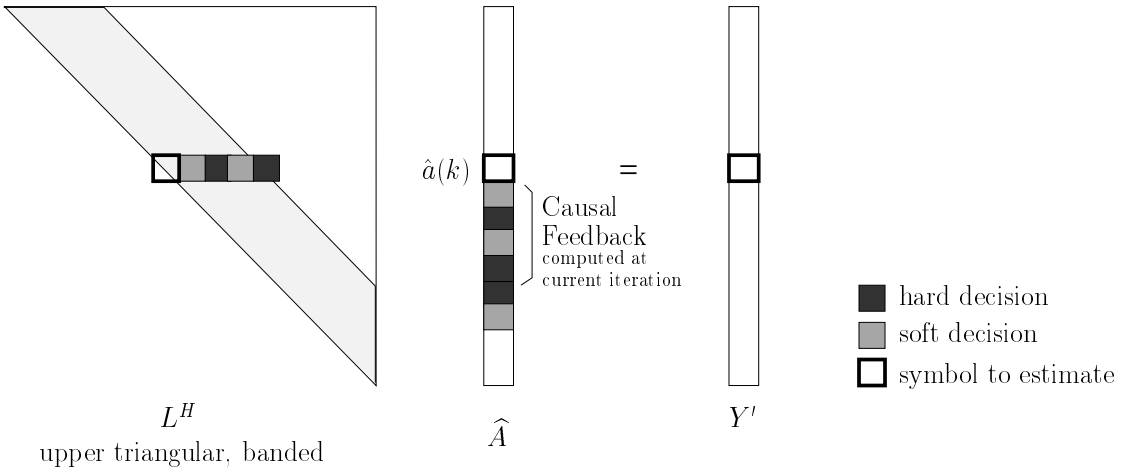



Figure 12.7: Computation of the soft DFE outputs.

This system is solved recursively by back substitution as illustrated in figure 12.7: $\hat{a}(k) = Y' - \sum_{i>k} L_{k+i,k}\hat{a}(i)$. Unlike in the classical implementation, the unreliable symbols are not decided upon and their soft value is fed back.

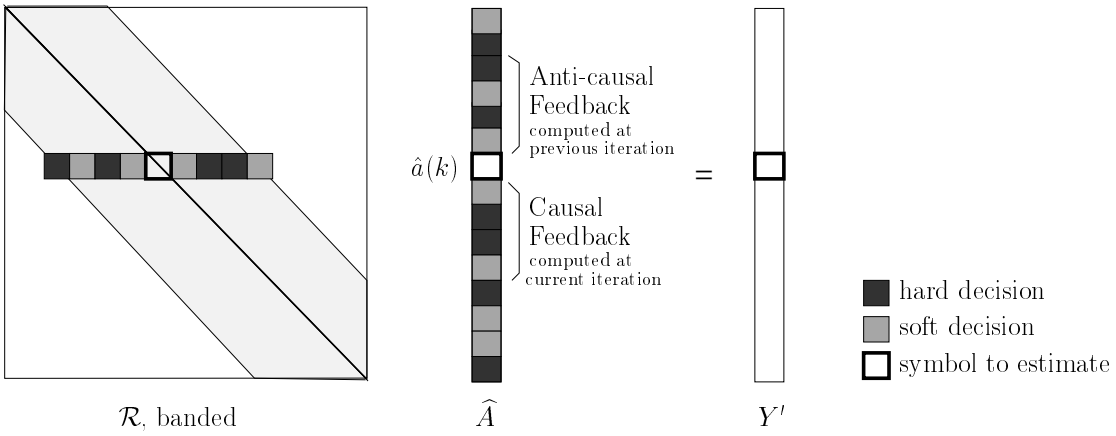



Figure 12.8: Computation of the soft NCDFE outputs.

### 12.6.2 Soft NCDFE

For the computation of the soft NCDFE output, we consider the equation:

$$\mathcal{R}\widehat{A} = \mathcal{T}^H(h)\boldsymbol{Y} \tag{12.11}$$

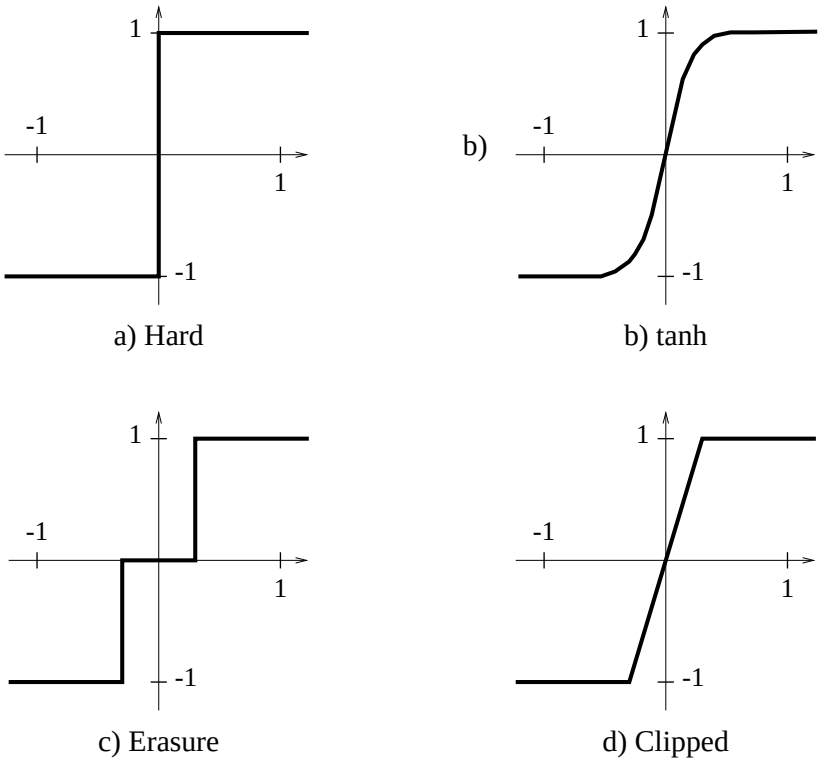



Figure 12.9: Different Soft Decisions Scheme.

where $\widehat{A}$ is initiated by the soft DFE. Each $a(k)$ is solved as: $\hat{a}(k) = Y' - \sum_{i<k} \mathcal{R}_{k-i,k}\hat{a}(i) - \sum_{i>k} \mathcal{R}_{k+i,k}\hat{a}(i)$, as in figure 12.8.

## 12.7 Different Soft Strategies

In figure 12.9 are depicted the different soft decision schemes found in the literature. Most of the schemes are also based on the idea of reliable and unreliable symbols.

a) Hard Decision: hard decisions are taken on all the symbol estimates.

b) *tanh* soft decisions.

c) Erasure soft decisions: in this scheme, hard decisions are taken on the reliable symbol estimates and the non–reliable symbols are put to 0.

d) Clipped soft decisions: this scheme corresponds to the one we propose.

The Erasure technique is used to reduce the error propagation in a decision-feedback equalizers in [100] for a BPSK and proposed for a general 1–D constellation in [101]. [102, 103] presents a dual feedback equalizer where two DFEs run independently. When a symbol estimate is judged non reliable, one DFE chooses $+1$ and the other $-1$, so that both alternative decisions $+1$ and $-1$ are tested. The output error energy over the feedback memory is then compared for both DFEs; the DFE with the lowest error energy is selected and its settings (previous decisions) are transferred to the other equalizer.

Non linear multi–user detection techniques, like Parallel Interference Cancellation (PIC) and Successive Interference Cancellation (SIC) use also soft decisions [104, 105]. PIC is based in fact on the same idea as the NCDFE: the past and the future decisions of the user of interest and the decisions of the other users are fed back. In [98, 99, 106, 107], some of the soft techniques of figure 12.9 are tested. In [108], we can also find a variable threshold of reliability based on the symbols already detected.

## 12.8 Simulations

For a well–conditioned channel, the effect of the soft decisions will not be very visible and, often, hard decisions will give better results, even if some decision errors are fed back. To illustrate the benefits of the proposed soft scheme (that we call, as in [98], clipped soft decision scheme), we test here a particular channel: the DFE built from this channel gives poor performance w.r.t. the MFB, which is the performance bound of the NCDFE. A channel with a zero closed to the unit circle verifies this property. We chose the channel:

$$\boldsymbol{H} = \begin{bmatrix} 1 & 1 \\ 1 & 1.01 \end{bmatrix} . \tag{12.12}$$

For this channel, the mean burst SNR value is 1.87 for the MMSE DFE, and the MFB is 0.026.

Below, we show the averaged error probability over 5000 Monte–Carlo runs of the input symbols and of the noise. We also test different reliability thresholds: the threshold $1.3\sigma^2_{\tilde{a}(k)}$ is based on mean quantities and for ill–conditioned channels, there are some advantages in choosing a larger value. We test the value $1.7\sigma^2_{\tilde{a}(k)}$ and $2\sigma^2_{\tilde{a}(k)}$. We only show the results for the MMSE equalizers: the ZF equalizers perform poorly as the subchannels of (12.12) possess a nearly common zero.

### 12.8.1 Soft DFEs

In figure 12.10, we show the performance of the soft DFEs based on the clipped, the erasure and the *tanh* soft decisions, for 10dB and 15dB and with different threshold levels. The clipped scheme appears to give the best performance. We also plot the performance of the

clipped soft scheme with a constant threshold of 1: the advantage of adapting the reliability intervals to the experimental conditions can be noticed.

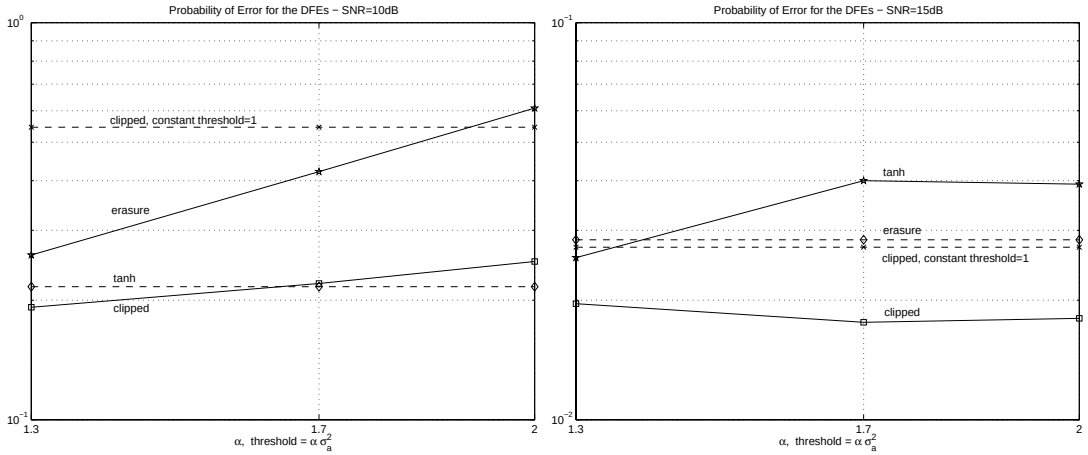



Figure 12.10: Probability of Error of the Soft DFEs.

### 12.8.2 Soft NCDFEs

Figure 12.11 shows the error probabilities of the different soft NCDFEs initialized by the soft DFEs. As the DFE based on the *tanh* curve gives good performance, we also test a NCDFE mixing the clipped and the *tanh* scheme: at a certain iteration, symbol estimation is based on the hard decisions of the previous iteration and the *tanh* scheme for the unknown symbols of the current iteration; the reliability intervals are computed based on the clipped scheme. This NCDFE is initialized by the *tanh* based DFE. Again, we see that the clipped scheme performs the best, with an improvement w.r.t. to the soft DFE that can reach a factor 10. Choosing a threshold larger than $1.3\sigma^2_{\tilde{a}(k)}$ can be seen to be advantageous also.

In figure 12.12, the clipped soft NCDFE is compared to the hard NCDFE initialized by the MMSE Linear Equalizer (LE) and the MMSE DFE. The benefit of the clipped soft decisions can also be noticed.

We do not plot the simulated MFBs which are here much lower than the practical performance of the proposed equalizers: in the example treated here, the purpose was to improve the performance of the equalizers by using soft decisions; it turns out that the problem is ill–conditioned and that the MFB cannot be easily attained.

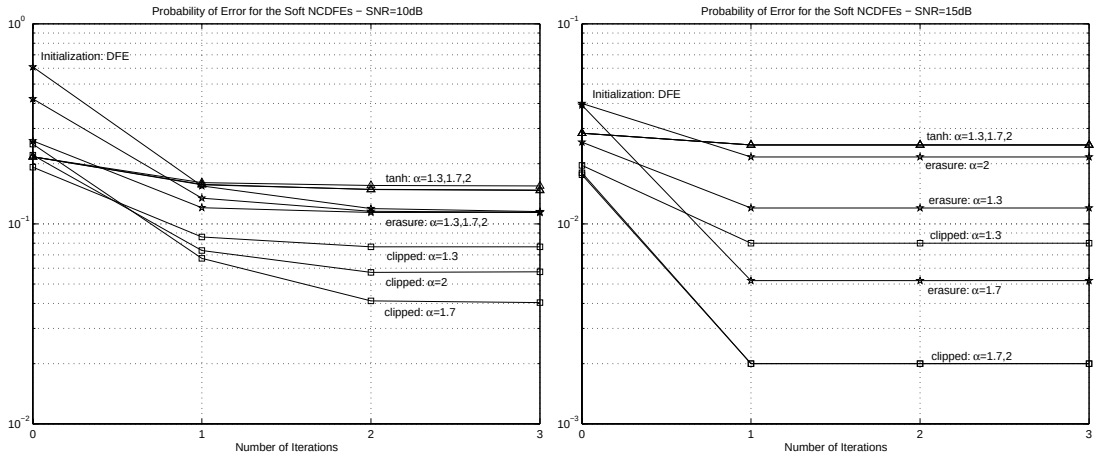



Figure 12.11: Probability of Error of the Soft NCDFEs.

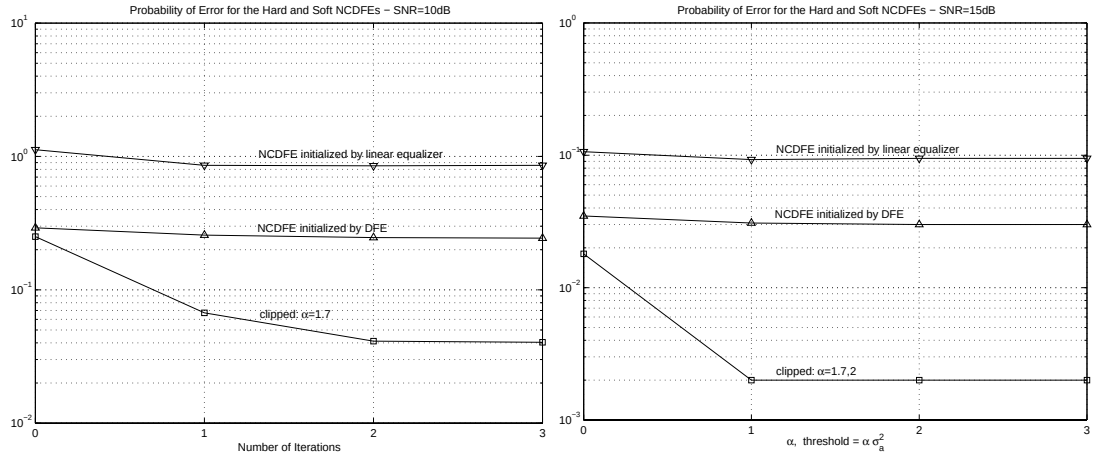



Figure 12.12: Probability of Error of the Hard and Soft NCDFEs.

## 12.9 Conclusion

We have described an implementation of the NCDFE using soft decisions instead of hard decisions, which cause error propagation. In this soft implementation, only the most reliable symbols are fed back and the unreliable ones are left undecided. A reliability interval has been defined which is adapted at each iteration of the NCDFE according to the SNR, the position of the symbol to be estimated in the burst, the known symbols and the reliable symbols on which a decision was taken at the previous iteration. Simulations have shown that the soft implementation can improve dramatically performance w.r.t. a hard implementation. The proposed soft decision scheme has been compared to other soft schemes and has given the best simulation results.

Chapter 13

# MATCHED FILTER BOUNDS FOR REDUCED-ORDER MULTICHANNEL MODELS

*We propose two Matched Filter Bounds (MFBs) to characterize the performance of receivers using reduced-order channel models. The first one (WMFB) uses the channel model to perform the spatio-temporal matched filtering that yields data reduction from multichannel to single-channel form. The rest of the processing remains optimal. The second one (ICMFB) on the other hand bounds the performance of the Viterbi algorithm with the reduced channel model. Two methods for obtaining reduced-order channel models are discussed to illustrate these measures: blind channel estimation by Deterministic Maximum Likelihood (which maximizes WMFB) and channel estimation by training sequence. This work was presented in [109]*

## 13.1 Different Matched Filter Bound Definitions

### 13.1.1 Continuous Processing MFB

We present here four different ways of computing the MFB in the case of continuous transmission, for the multichannel $\mathbf{H}(z)$, shown in figure 13.1, where the input symbols $a(k)$ are white and the additive noise $\boldsymbol{v}(k)$ is temporally and spatially white. We introduce the following notation for the matched filter: $\boldsymbol{H}^{\dagger}(z) = \boldsymbol{H}^{H}(1/z^{*})$.

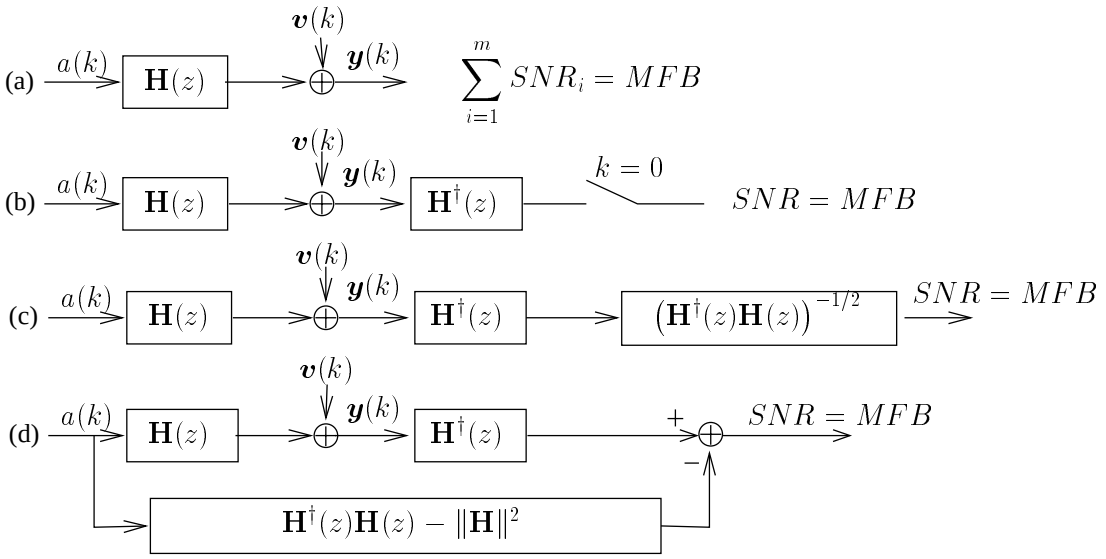



Figure 13.1: Four Interpretations for the Continuous Processing MFB from SNRs.

The MFB can alternatively be calculated as the sum of the SNRs in the individual channels in (a), as the SNR of the appropriate output sample of the matched filter (MF) when transmitting only one symbol in (b), as the SNR of the output of the whitened MF (WMF) in (c) or finally as the SNR at the output of the MF from which past and future symbol contributions (ISI) are eliminated. The MFB, calculated from (a), (b), (c) or (d), is equal to $\|\mathbf{H}\|^2\sigma_a^2/\sigma_v^2$.

### 13.1.2 Burst Processing MFB

The MFB becomes symbol-dependent in the case of burst (packet) transmission. In the different structures presented in figure 13.1, the multichannel $\mathbf{H}(z)$ is replaced by the filtering matrix $\mathcal{T}(h)$ in the time-domain, and the burst multichannel matched filter becomes $\mathcal{T}^{H}(h)$. Since the MFB is symbol-dependent, we shall in fact consider the average MFB over the symbols in the burst. In a burst context, figure 13.2(b) is no longer of interest, which is why we will not consider this configuration anymore.

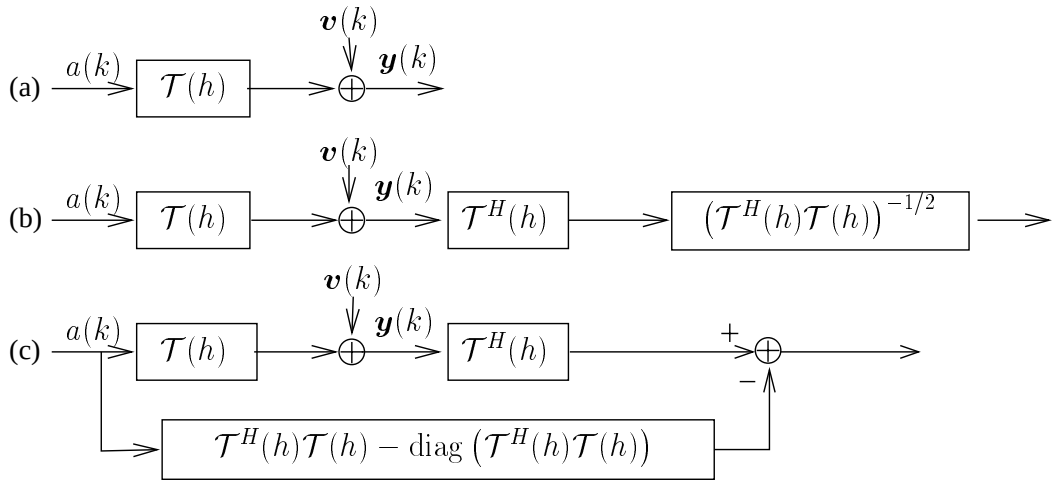



Figure 13.2: Four ways to get the Burst MFB from SNRs.

In figure 13.2(a), the burst signal covariance matrix at the channel output is: $\sigma_a^2 \mathcal{T}(h)\mathcal{T}^H(h)$. The noise variance is $\sigma_v^2 I$. The SNR for the $i^{th}$ element of the output is then:

$$SNR_i^{(a)} = \frac{\sigma_a^2 \left[\mathcal{T}(h)\mathcal{T}^H(h)\right]_{i,i}}{\sigma_v^2}. \tag{13.1}$$

We find for (b):

$$SNR_i^{(b)} = \frac{\sigma_a^2 \left[\left(\mathcal{T}^H(h)\mathcal{T}(h)\right)^{H/2}\left(\mathcal{T}^H(h)\mathcal{T}(h)\right)^{1/2}\right]_{i,i}}{\sigma_v^2}. \tag{13.2}$$

The signal component is $\mathrm{diag}(\mathcal{T}^H(h)\mathcal{T}(h))A$ in (c) (where diag(.) denotes a diagonal matrix containing of the main diagonal of its argument), hence its variance, $\sigma_a^2 \left(\mathcal{T}^H(h)\mathcal{T}(h)\right)_{i,i}^2$, for the $i^{th}$ element, for which the noise variance is $\sigma_v^2 \left[\mathcal{T}^H(h)\mathcal{T}(h)\right]_{i,i}$. Thus we find:

$$SNR_i^{(c)} = \frac{\sigma_a^2 \left[\mathcal{T}^H(h)\mathcal{T}(h)\right]_{i,i}}{\sigma_v^2}. \tag{13.3}$$

Hence, we find the following equivalent expressions:

$$\sum_{i=1}^{Mm} SNR_i^{(a)} = \sum_{i=1}^{M+N-1} SNR_i^{(b)} = \sum_{i=1}^{M+N-1} SNR_i^{(c)}. \tag{13.4}$$

The structure in figure 13.2(a) represents in fact a different point of view from (b) or (c). Indeed, the $M+N-1$ outputs of (b) and (c) are directly related to the the $M+N-1$ input

samples; in (a) we get $Mm$ output samples. The measure in equation (13.4) can then be taken as a measure of the burst MFB. This leads to the following average MFB per symbol:

$$MFB = \frac{1}{M+N-1}\sum_{i=1}^{Mm} SNR_i^{(a)} = \frac{1}{M+N-1}\sum_{i=1}^{M+N-1} SNR_i^{(b)} = \frac{1}{M+N-1}\sum_{i=1}^{M+N-1} SNR_i^{(c)}\,. \tag{13.5}$$

Note that, as the length of the burst grows to infinity, the average MFB over the burst converges to the continuous processing MFB. For structures (b) and (c) this follows from the fact that the MFB in the middle of the burst converges to the continuous processing MFB.

## 13.2 Matched Filter Bounds for Reduced-Order Models

The MFB computation considered in the previous section requires knowledge of the channel. However, in channel estimation, a channel order misestimation may happen. Since physical channel impulse responses tend to be of infinite length, this misestimation will often mean an underestimation. Furthermore, the channel length assumed in the channel estimation is often limited due to complexity considerations for the estimation procedure and/or the symbol detection procedure. We now discuss appropriate MFBs when a reduced-order channel model is used. Two levels of suboptimality ensue in that case. These correspond to the two ways of implementing ML sequence estimation (MLSE) in the multichannel case: either use a vectorial matched filter and work with a scalar signal, or work with the vector received signal directly. These two strategies are only equivalent if the further processing of the scalar signal obtained in the first case is done in a specific way. Two measures corresponding to these two strategies are proposed.

### 13.2.1 Whitened Matched Filter Bound (WMFB)

We denote by $\mathbf{H}_N(z)$ the full-order multichannel. Assume we have a reduced-order model $z^{-d}\mathbf{H}_{N'}(z)$ of $\mathbf{H}_N(z)$ ($d \in \{0,1,\ldots,N-N'\}$, $1 \le N' \le N$). In a first step of suboptimality, we can consider that in the data reduction step from multichannel to single channel, we use the MF matched to the reduced model $z^{-d}\mathbf{H}_{N'}(z)$. However, after this suboptimal data reduction, we shall allow optimal processing of the resulting single channel (this requires knowledge of $\mathbf{H}_{N'}^{\dagger}(z)\mathbf{H}_N(z)$ which represents less information than $\mathbf{H}_N(z)$ itself). In order to find the conventional MFB for the processing of the resulting scalar channel, it suffices to whiten the noise after the vector MF. The resulting scalar channel then indeed becomes one of additive white noise $n(k)$ as indicated in figure 13.3 and becomes similar to figure 13.1(a) so that the MFB can be calculated as in figure 13.1(a). We get for the continuous processing MFB:

$$\mathrm{WMFB} = \frac{\sigma_a^2}{\sigma_v^2}\frac{1}{2\pi j}\oint \mathbf{H}_N^{\dagger}(z)\mathrm{P}_{\mathbf{H}_{N'}(z)}\mathbf{H}_N(z)\,\frac{dz}{z} \tag{13.6}$$

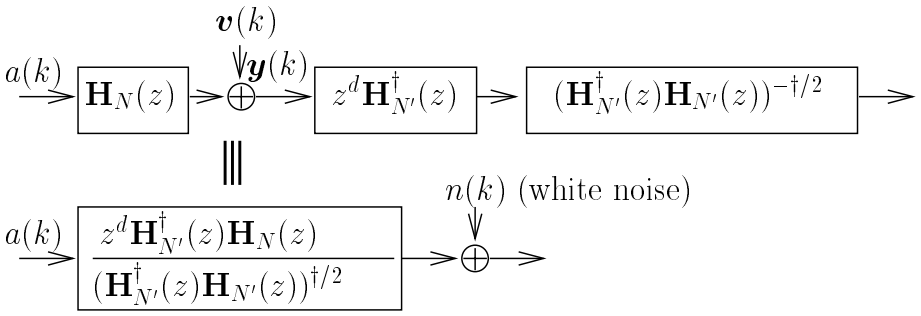



Figure 13.3: WMFB: reduced-order multichannel MF followed by a noise whitener.

where $\mathrm{P}_{\mathbf{H}(z)} = \mathbf{H}(z)(\mathbf{H}^\dagger(z)\mathbf{H}(z))^{-1}\mathbf{H}^\dagger(z)$. Note that it is insensitive to the delay in the reduced order channel model.

It is interesting to analyze the variation of WMFB($N'$) as a function of the reduced order $N'$. For $N = N'$ we get WMFB($N$)= $\frac{\sigma_a^2}{\sigma_v^2}\|\mathbf{H}_N\|^2$. It is not difficult to show that, in the limiting case $N' = 1$ (purely spatial channel model), we get WMFB(1)= $\frac{\sigma_a^2}{\sigma_v^2}\lambda_{\max}\left(\boldsymbol{H}_N^H\boldsymbol{H}_N\right)$. We then can derive the following bounds

$$1 \leq \frac{\mathrm{WMFB}(N)}{\mathrm{WMFB}(1)} = \frac{\mathrm{tr}\left(\boldsymbol{H}_N^H\boldsymbol{H}_N\right)}{\lambda_{\max}\left(\boldsymbol{H}_N^H\boldsymbol{H}_N\right)} \leq \min(m,N)\,. \tag{13.7}$$

We see that a reduced-order model does not degrade WMFB a lot: in the case of 2 subchannels the maximal degradation will be a factor of 2, which could seem surprising when considering a purely spatial model only. The lower bound is attained when $\boldsymbol{h}(i) \sim \boldsymbol{h}(0)$, $i = 1,\ldots,N-1$. In that case, $\mathbf{H}_N(z) = \boldsymbol{h}(0)\mathrm{H}_1(z)/h_1(0)$. The spatio-temporal channel factors into a spatial filter and a temporal one, and the optimal processing factors correspondingly: the full spatio-temporal treatment gets replaced by the cascade of a purely spatial combiner followed by a purely temporal treatment. The upper bound is attained when either $\boldsymbol{H}_N\boldsymbol{H}_N^H \sim I_m$ or $\boldsymbol{H}_N^H\boldsymbol{H}_N \sim I_N$, whichever is of full rank. In that case, the individual channel impulse responses are orthonormal. In a statistical set-up, if the $m$ channel impulse responses are i.i.d., then the upper bound is approached as the delay spread grows.

Consider now the case of burst processing. Let $\mathcal{T}_N$ and $\mathcal{T}_{N'}$ denote $\mathcal{T}_M(h_N)$ and $\mathcal{T}_M(h_{N'})$ respectively and consider the Cholesky factorization $\mathcal{T}_{N'}^H\mathcal{T}_{N'} = LL^H$. Then the $M+N'-1$ reduced-order WMF outputs are

$$L^{-1}\mathcal{T}_{N'}^H\boldsymbol{Y} = L^{-1}\mathcal{T}_{N'}^H\mathcal{T}_N A + L^{-1}\mathcal{T}_{N'}^H\boldsymbol{V}\,. \tag{13.8}$$

The covariance matrix of the noise component is $\sigma_v^2 I_{M+N'-1}$ while the covariance matrix of

the signal part is $\sigma_a^2 L^{-1}\mathcal{T}_{N'}^H\mathcal{T}_N\mathcal{T}_N^H\mathcal{T}_{N'}L^{-H}$. The sum of the SNRs of all WMF outputs is then

$$\sum_{i=1}^{M+N'-1}\frac{\sigma_a^2}{\sigma_v^2}\left(L^{-1}\mathcal{T}_{N'}^H\mathcal{T}_N\mathcal{T}_N^H\mathcal{T}_{N'}L^{-H}\right)_{ii}=\frac{\sigma_a^2}{\sigma_v^2}\,\mathrm{tr}\left(\mathrm{P}_{\mathcal{T}_{N'}}\mathcal{T}_N\mathcal{T}_N^H\right)\ . \tag{13.9}$$

This point of view corresponds to (a) in figure 13.2. To find the equivalent of (c) in figure 13.2, consider passing the previous WMF output $L^{-1}\mathcal{T}_{N'}^H\boldsymbol{Y}$ through the scalar MF $\mathcal{T}_N^H\mathcal{T}_{N'}L^{-H}$. This gives the $M+N-1$ outputs

$$\mathcal{T}_N^H\mathrm{P}_{\mathcal{T}_{N'}}\boldsymbol{Y}=\mathcal{T}_N^H\mathrm{P}_{\mathcal{T}_{N'}}\mathcal{T}_N A+\mathcal{T}_N^H\mathrm{P}_{\mathcal{T}_{N'}}\boldsymbol{V}\ . \tag{13.10}$$

As seen in section 13.1.2, the sum of the output SNRs in figure 13.2(c) is equal to the expression in equation (13.9). It is also possible to find the equivalent of (b) in figure 13.2, the sum of output SNRs giving again (13.9). What we call burst WMFB is again the average WMFB over the burst:

$$\mathrm{WMFB}=\frac{1}{M+N-1}\mathrm{tr}\left(\mathrm{P}_{\mathcal{T}_{N'}}\mathcal{T}_N\mathcal{T}_N^H\right)\ . \tag{13.11}$$

### 13.2.2 ISI Canceler Matched Filter Bound (ICMFB)

We now go all the way in suboptimality. We will not only assume that the multichannel MF is based on the reduced channel model but in fact that the whole receiver is. To find the optimal performance in this case, consider MLSE. The received burst through the channel $\mathbf{H}_N(z)$ is $\boldsymbol{Y}(k)=\mathcal{T}(h_N)A_{M+N-1}(k)+\boldsymbol{V}(k)$. The channel estimation procedure has given a reduced-order model $z^{-d}\mathbf{H}_{N'}(z)$ in which $\mathbf{H}_{N'}(z)$ is known but the delay d may be unknown. Based on the reduced-order model $z^{-d}\mathbf{H}_{N'}(z)$, the MLSE problem is

$$\begin{aligned}&\min_{\substack{a(i)\,\in\,\mathcal{A}_p\\ d\,\in\,\{0,1,\ldots,N-N'\}}}\quad\|\boldsymbol{Y}(k)-\mathcal{T}(h_{N'})A_{M+N'-1}(k-d)\|^2\end{aligned} \tag{13.12}$$

where $\mathcal{A}_p$ is the symbol alphabet. We obtain the ISI Canceler Matched Filter Bound (ICMFB) by considering the detection of a single symbol $a(i)$ assuming that the other symbols are known. It is easy to see that the continuous processing version of this leads to the structure in figure 13.4(a) (except for a scale factor). As the true channel is not known, the terms containing coefficients of the channel estimation error contribute to noise in the SNR computation. Hence, the equivalent structures in (b) and (c).

The output SNR in figure 13.4(c) is:

$$\mathrm{ICMFB}=\max_{d\in\{0,\ldots,N-N'\}}\frac{\|\mathbf{H}_{N'}\|^2\sigma_a^2}{\sigma_v^2+\sigma_a^2\|\mathbf{H}_{N'}^{\dagger}(z)(z^d\mathbf{H}_N(z)-\mathbf{H}_{N'}(z))\|^2/\|\mathbf{H}_{N'}\|^2}\ . \tag{13.13}$$

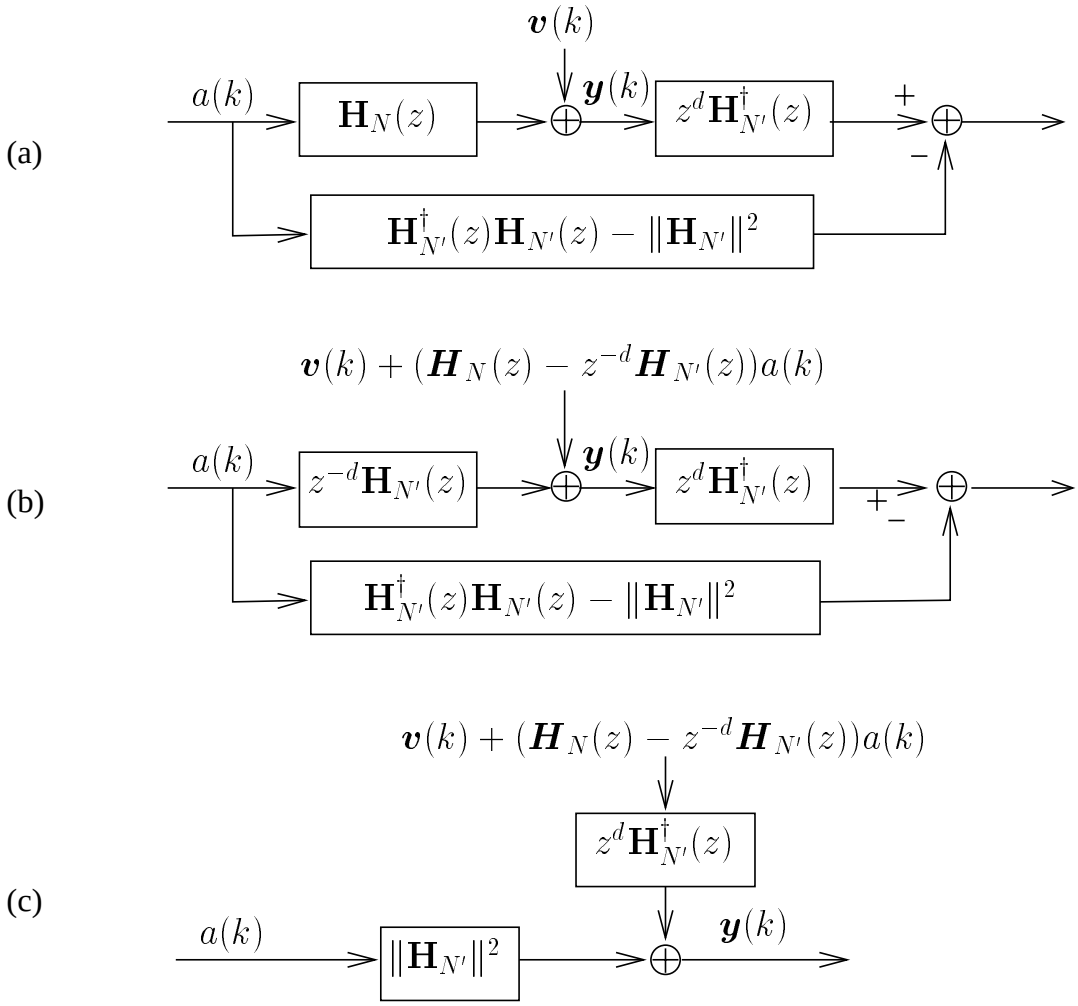



Figure 13.4: ICMFB: MFB for MLSE with the reduced-order channel model.

In contrast to WMFB, the delay $d$ in the reduced-order channel model plays a role in ICMFB. Note that the presence of an adjustable delay creates local minima for MLSE. Remark also that for $N' = N$, ICMFB=WMFB=MFB.

## 13.3 Two Applications

### 13.3.1 Deterministic Maximum Likelihood Channel Estimation

As a first example, we shall investigate the effect of model reduction in blind DML channel estimation. In [30], it is proved that, asymptotically in the number of data, the DML criterion approximates the channel with a lower order model such that the output SNR of the Whitened Matched Filter corresponding to this lower order model gets maximized. Asymptotically, the reduced-order channel estimate obtained with the DML is the one that maximizes WMFB.

Some simulations were performed for $m = 2$ channels and average SNR per subchannel of 10dB. In order to concentrate on the model reduction effects and not on the estimation errors, the averaged likelihood function was maximized, solved by IQML. DML only allow the estimation of the channel up to a multiplicative constant. WMFB on the other hand is quite sensitive to the choice of this scale factor: we have determined the magnitude of this scale factor using the norm constraint $\|\mathbf{H}_{N'}\| = \|\mathbf{H}_N\|$.

We considered continuous processing WMFB and ICMFB measures, but since the IQML method will normally be applied to a burst of data $\boldsymbol{Y}_M(k)$, we also considered the burst processing WMFB measure.

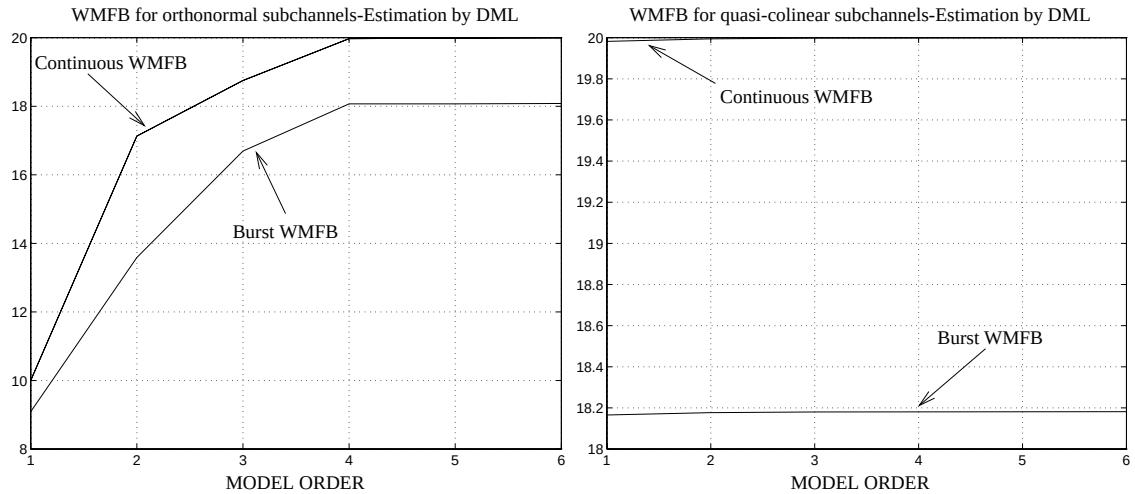



Figure 13.5: WMFB as a function of $N' = 1, \ldots, N$ for $m = 2$, $M = 50$, $N = 6$ for orthonormal (left) and almost colinear (right) impulse responses.

First, we illustrate equation (13.7), where the minimal and maximal degradations when a model of reduced order 1 is considered, are shown. Figure 13.5 shows the evolution of the continuous and burst processing WMFB as a function of $N'$ for a case in which the two impulse responses are orthonormal and a case in which they are almost colinear. In the first case, we see a degradation of approximatively 1/2 from $N' = N$ to $N' = 1$ as predicted in (13.7). In the second case, quasi no degradations are visible. We note here that the burst WMFB is lower than its continuous processing version. This is due to the degradations occurring at the edges of the burst w.r.t. continuous mode performances.

Some other simulations were done for less particular channels. The evolution of WMFB and ICMFB as a function of $N'$ is shown in figure 13.6 for the following two decaying channels:

$$\begin{aligned} \boldsymbol{H}_1 &= \begin{bmatrix} 1.0000 & 0.8000 & 0.5000 & 0.6000 & 0.1000 & 0.0050 \\ -1.5000 & 1.4000 & -0.9000 & 1.1000 & -0.0300 & 0.0050 \end{bmatrix} \\ \boldsymbol{H}_2 &= \begin{bmatrix} 1.0000 & 0.5000 & -0.1500 & 0.0550 & 0.0145 & -0.0014 \\ 1.5000 & -0.9500 & 0.3050 & 0.0695 & 0.0431 & -0.0043 \end{bmatrix} \end{aligned} \tag{13.14}$$

Both continuous and burst mode WMFB are ploted, as well as ICMFB.

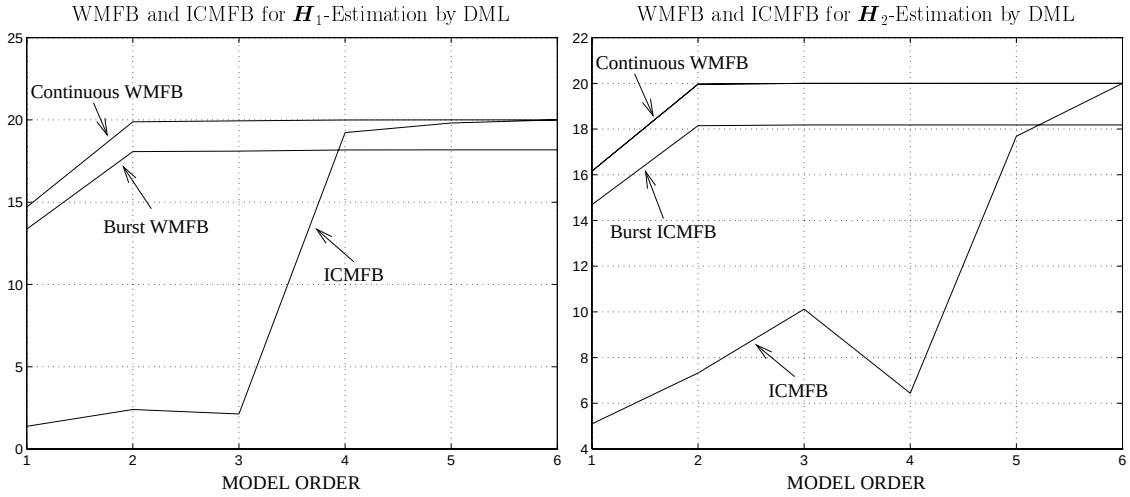



Figure 13.6: Comparison of WMFB and ICMFB as a function of $N'$ for channels $H_1$ and $H_2$, $m = 2$, $N = 6$, $M = 50$ in the case of DML estimation.

We notice that WMFB is greater than ICMFB for all reduced orders $N'$. The degradations due to reduced-order modeling are less severe for WMFB than for ICMFB, especially for low orders. This verifies equation (13.7), where we saw that maximal degradation for WMFB due to a model of reduced-order 1 is limited. Furthermore, degradations for WMFB occur mostly for $N' = 1$. As DML reduced-order models maximizes WMFB, WMFB is decreasing as $N'$ decreases.

For channel $\boldsymbol{H}_1$, ICMFB decreases considerably when the reduced model is of order 3. This is probably due to the fact that the channel contains most of its energy in its first 4 coefficients, which shows that ICMFB is sensitive to the energy contained in the reduced-order channel.

### 13.3.2 Training Sequence based Channel Estimation

In this second example, the channel is estimated by a white training sequence. The channel estimate of reduced-order $N'$ becomes the part of $N'$ consecutive coefficients of $\boldsymbol{H}_N$ which contains the most energy. This estimation procedure produces a value for the delay d. However, this value for d may not be the best one for MLSE. Hence the problem formulation in (13.12) and the ensuing bound in (13.13) with optimization over d are still meaningful. Nevertheless, the optimal d thus obtained will usually equal the d obtained with channel estimation by training sequence.

We see in figure 13.7 that WMFB is not decreasing anymore, but remains high and always greater than ICMFB. Although training sequence based channel estimation does not maximize ICMFB, it tends to. Indeed, in equation (13.13), the numerator $\|\mathbf{H}_{N'}\|^2$ gets maximized, and the coefficient in $z^0$ of $\mathbf{H}^{\dagger}_{N'}(z)(z^d\mathbf{H}_N(z) - \mathbf{H}_{N'}(z))$ becomes equal to 0. In particular, we see how ICMFB improves, for channel $\boldsymbol{H}_2$ when the reduced-order channel is estimated by training sequence compared to DML estimation.

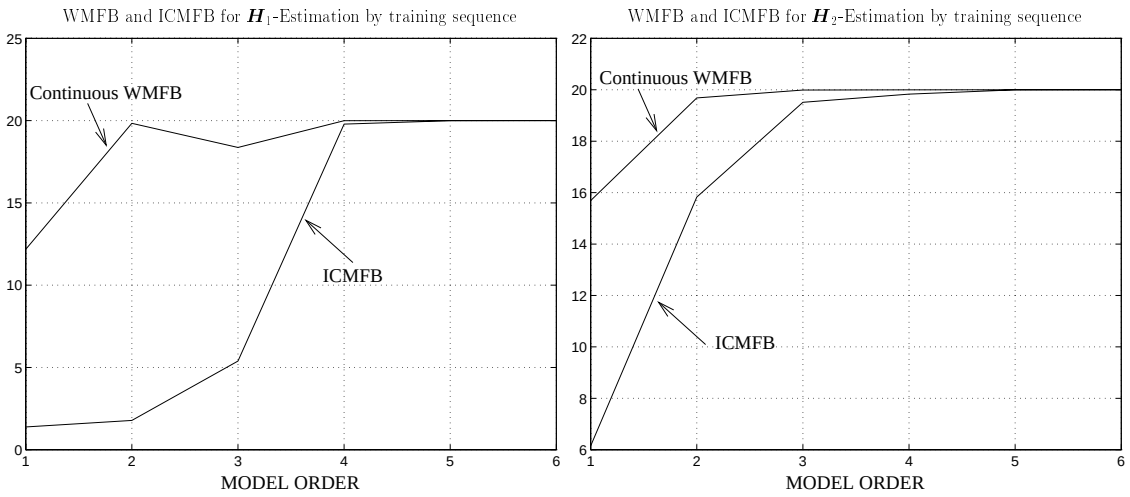



Figure 13.7: Comparison of WMFB and ICMFB as a function of $N'$ for channels $H_1$ and $H_2$, $m = 2$, $N = 6$, $M = 50$ in the case of training sequence based channel estimation.

## 13.4 Conclusion

In this chapter, we have proposed two performance bounds for MLSE when the channel order is underestimated. These measures are interesting when complexity reduction, through channel order reduction, have to be done, for the implemention of the Viterbi equalizer for example. In simulations, we use these bounds to see the effect of the channel order reduction for certain channel estimation methods. In [110], MFBs were developed in the case of colored noise (representing interferers) for an erroneous noise covariance model. MFBs taking into account channel errors are also under study.

# GENERAL CONCLUSION

In this thesis, we have presented an extensive study on semi–blind channel estimation. The first theoretical part has concentrated on performance bounds for semi–blind, blind, and training sequence based channel estimation: the superiority of semi–blind techniques over blind and training sequence based techniques have been shown. The comparison between semi–blind and blind methods have motivated an analysis of performance under constraints. FIM regularity and local identifiability are equivalent, and the number of independent constraints needed to regularize the estimation problem is equal to the number of singularities of the FIM. We have proposed a bound for blind estimation, which is the pseudo–inverse of the FIM, giving, for a minimal number of independent constraints, the minimal CRB and that we have interpreted as a CRB with some specific constraints. We think that this study is important because it gives a way to systematically characterize most constrained estimation problems, and furthermore, it provides a clear characterization of blind channel performance, which is sometimes done in a non satisfactory way.

Different ways of building semi–blind criteria have been presented. The optimal semi–blind criteria should be based on methods that naturally incorporate the knowledge of symbols: this is the case of ML methods, that we studied in this thesis, and also of methods that estimate directly the symbols like [76]. These optimal methods provide semi–blind solutions when the symbols are arbitrarily dispersed in the burst. This symbol configuration is undesirable in general as the associated semi–blind criteria will require computationally demanding algorithms.

For grouped known symbol, *i.e.* a training sequence, low complex solutions can be built because the structure of the blind problem is kept. By neglecting some information about the known or unknown symbols, ML easily allows one to construct semi–blind criteria that are a linear combination of a blind and a training sequence based criteria. Especially when the training sequence is short, it appears important to be able to take into account the overlap zone where known and unknown symbols appear at the same time. One of the solutions we have proposed for that is to combine blind DML to the optimally weighted least squares criterion; this combination corresponds to a mixed deterministic and Gaussian point of view and was shown to give the best results in our simulations. A semi–blind SRM solution was also proposed that proved to perform very well with performance closed to the ML performance.

The last step of this study was to find how to construct a semi–blind criterion based on a

given blind criterion. We propose a weighted combination of the blind and training sequence criteria which is typically possible when the blind criteria are based on least–squares. The semi–blind SRM example is built that way and by approximating the weighting matrix of blind SRM by a diagonal matrix, the criterion becomes a linear combination of blind SRM and a training sequence based criterion. Some words are said about the subspace fitting and covariance matching based semi–blind methods.

Constructing a semi–blind criterion as a linear combination of a blind and a training sequence based criterion may be a difficult task. Through the different examples studied, we could see that the resulting semi–blind criteria are very sensitive to the value of the coefficients of the linear weighting, and, for certain values of the weighting, the semi–blind criterion may give worse performance than the pure training sequence criteria. We could notice that when the linearly combined semi—blind criterion is correctly constructed, as with our semi–blind SRM for example, the performance is quite stable around the value of the chosen weight.

Blind algorithms have also been studied. We have provided new solutions to solve DML that are not complex, and need few iterations to converge. PQML is particularly powerful as it attains asymptotically the ML performance. Gaussian ML, well-known in the DOA context, is often misunderstood in the channel identification context. We brought more understanding to it by interpreting it as a form of covariance matching criterion and comparing it to the classical optimally weighted least squares covariance matching criterion. Through simulations, GML appears as the most powerful method among all the methods exploiting the second–order moments of the data; covariance matching attains GML performance for an infinite size of covariance matrix. Fast implementations of GML have been proposed and some others are still under investigation; fast implementations for covariance matching are also the subject of ongoing studies.

At last, we would like to insist on the interest of the non–causal DFE which is not extensively used whereas it possesses the potential of providing lower probability errors than the classical DFE. The piece–wise linear soft decision scheme we proposed offers interesting perspectives especially in a multi–user context applied to techniques like parallel interference cancellation.

# RÉSUMÉ EN FRANÇAIS

# 1 Introduction

La plupart des standards de communications mobiles actuels incluent une séquence de symboles connus, la séquence d'apprentissage, pour estimer le canal. Les méthodes d'apprentissage utilisent les échantillons du signal reçu contenant des symboles connus uniquement et tous les autres échantillons, contenant des symboles inconnus sont ignorés. Les méthodes aveugles sont basées sur tous les échantillons, mais la connaissance de symboles à l'entrée n'est pas mise à profit. Le but de l'estimation de canal semi–aveugle est d'exploiter à la fois l'information aveugle et l'information provenant des symboles connus et ainsi de combiner les aspects positifs des deux techniques.

Les techniques semi-aveugles permettent d'estimer les réponses impulsionnelles de canaux plus longues que possible avec une certaine longueur de séquence d'apprentissage; pour une longueur de canal et une qualité d'estimation données, elles permettent l'utilisation de séquences d'apprentissage plus courtes comparées à une technique d'apprentissage; les méthodes semi–aveugles sont également plus robustes que les méthodes aveugles qui requièrent des conditions de régularité sur le canal. Enfin, elles offrent de meilleures performances d'estimation que les méthodes aveugles et les méthodes d'apprentissage.

Le principal objet de cette thèse est l'étude de l'identification semi–aveugle de canaux FIR multiples, avec une transmission des données d'entrée par paquet. Nous présentons tout d'abord les conditions d'identifiabilité et des bornes de performance pour l'estimation semi–aveugle. On montre que les méthodes semi–aveugles peuvent identifier n'importe quel canal, avec peu de symboles connus et même lorsque ceux-ci sont dispersés arbitrairement dans le paquet. Des bornes de performance, les bornes de Cramér–Rao, permettent une comparaison des techniques semi-aveugles avec les techniques d'apprentissage et les techniques aveugles. De plus, une étude de performance sous contraintes est donnée pour caractériser les performances de l'estimation aveugle.

Les méthodes semi–aveugles proposées sont basées sur le principe de Maximum de Vraisemblance (MV) qui offre la possibilité d'incorporer la connaissance de symboles d'entrée. Lorsque les symboles connus sont groupés dans une séquence d'apprentissage, des méthodes sous–optimales sont proposées: les critères correspondant sont sous la forme d'une combinaison linéaire d'un critère basé sur la séquence d'apprentissage et du critère MV aveugle. Afin de construire des critères semi–aveugles performants, nous nous concentrons également sur l'étude des méthodes MV aveugles. Enfin, nous proposons des solutions pour construire des critères semi–aveugles combinant un certain critère aveugle (qui n'est pas un critère MV) et un critère d'apprentissage.

Dans cette thèse, des structures d'égaliseurs sont également proposés. La structure des égaliseurs en mode paquet et en particulier la structure d'un annulateur d'interférences entre symboles que nous appelons Non–Causal Decision–Feedback Equalizer (NCDFE) et qui utilise les décisions passées mais également futures. Une implémentation du NCDFE basée sur des

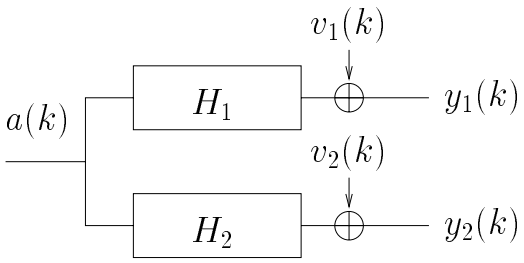



Figure 1: Modèle multicanal; exemple de 2 sous–canaux.

décisions douces est présentée. Enfin, des bornes de performance sur "Maximum Likelihood Sequence Estimation" (MLSE) pour des modèles de canaux d'ordre réduit sont données.

## 2 Formulation du Problème

Nous considérons un modèle multicanal FIR où une séquence de symboles $a(k)$ est envoyée à travers $m$ canaux linéaires de longueur $N$ et de coefficients $\boldsymbol{h}(i)$ (voir la figure 1):

$$\boldsymbol{y}(k) = \sum_{i=0}^{N-1} \boldsymbol{h}(i)a(k-i) + \boldsymbol{v}(k), \tag{1}$$

$\boldsymbol{v}(k)$ est un bruit additive Gaussien blanc, $r_{\boldsymbol{vv}}(k-i) = \mathrm{E}\,\boldsymbol{v}(k)\boldsymbol{v}(i)^H = \sigma_v^2 I_m\,\delta_{ki}$. Supposons que nous recevons $M$ échantillons, concaténés dans le vecteur $\boldsymbol{Y}_M(k)$:

$$\boldsymbol{Y}_M(k) = \mathcal{T}_M(h)\, A_{M+N-1}(k) + \boldsymbol{V}_M(k)\,. \tag{2}$$

$\boldsymbol{Y}_M(k) = [\boldsymbol{y}^H(k-M+1)\cdots\boldsymbol{y}^H(k)]^H$, et similairement pour $\boldsymbol{V}_M(k)$ et $A_M(k)$. La fonction de transfert du canal est $\mathbf{H}(z) = \sum_{i=0}^{N-1}\boldsymbol{h}(i)z^{-i} = [\mathrm{H}_1(z)\cdots\mathrm{H}_m(z)]^T$. $\mathcal{T}_M(h)$ est une matrice bloc Toeplitz avec $M$ ligne bloc et $\left[\boldsymbol{H}\ 0_{m\times(M-1)}\right]$ comme première ligne bloc ($\boldsymbol{H} = [\boldsymbol{h}(0)\cdots\boldsymbol{h}(N-1)]$). On notera de plus: $h = [\boldsymbol{h}^T(0)\ \cdots\ \boldsymbol{h}^T(N-1)]^T$. La longueur du canal est $N$ ce qui implique que $\boldsymbol{h}(0) \neq 0$ et $\boldsymbol{h}(N-1) \neq 0$; la réponse impulsionnelle est nulle en dehors de l'intervalle indiqué. On simplifiera la notation (2) pour $k = M-1$ en:

$$\boldsymbol{Y} = \mathcal{T}(h)A + \boldsymbol{V}\,. \tag{3}$$

**Commutativité de la Convolution** Nous aurons besoin de la propriété de commutativité de la convolution:

$$\mathcal{T}(h)A = \mathcal{A}_m h \tag{4}$$

où: $\mathcal{A}_m = \mathcal{A}_1 \otimes I_m$,

$$\mathcal{A}_1 = \begin{bmatrix} a(M-1) & a(M-2) & \cdots & a(M-N) \\ a(M-2) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ a(0) & \cdots & \cdots & a(-N+1) \end{bmatrix}. \tag{5}$$

Pour simplifier, on notera $\mathcal{A}$ la matrice $\mathcal{A}$.

**Modèle Semi–Aveugle** Le vecteur des symboles d'entrée s'écrit comme $A = \mathcal{P} \begin{bmatrix} A_K \\ A_U \end{bmatrix}$ où le vecteur $A_K$ regroupe les $M_K$ symboles connus et le vecteur $A_U$ regroupe les $M_U = M+N-1-M_K$ symboles inconnus. Les symboles connus peuvent être arbitrairement dispersés dans le paquet et $\mathcal{P}$ désigne la matrice de permutation appropriée. Dans le cas de l'estimation aveugle, $A = A_U$, alors que $A = A_K = A_{TS}$ dans le cas de l'estimation par apprentissage. La sortie du canal peut être décomposée en la contribution des symboles connus et la contribution des symboles inconnus: $\mathcal{T}(h)A = \mathcal{T}_K(h)A_K + \mathcal{T}_U(h)A_U$.

**Canaux irréductibles, réductibles et à minimum de phase** Un canal est dit irréductible si tous ces sous–canaux $\mathrm{H}_i(z)$ n'ont pas de zéros en commun, et réductible sinon. Un canal réductible peut être décomposé comme:

$$\mathbf{H}(z) = \mathbf{H}_I(z)\mathrm{H}_c(z), \tag{6}$$

où $\mathbf{H}_I(z)$, de longueur $N_I$, est irréductible et $\mathrm{H}_c(z)$ de longueur $N_c = N - N_I + 1$ est un monocanal pour lequel nous supposons $\mathrm{H}_c(\infty) = h_c(0) = 1$ (monique). Un canal est dit à minimum de phase si tous ces zéros sont à l'intérieur du cercle unité. Ainsi $\mathbf{H}(z)$ est à minimum de phase si et seulement si $\mathrm{H}_c(z)$ est à minimum de phase.

**Modèles pour les symboles d'entrée** Les méthodes aveugles peuvent être classifiées (approximativement) selon le niveau d'information exploité sur les symboles d'entrée: voir figure 2.

1. Pas d'information exploitée: les méthodes déterministes.

   Ces méthodes sont basées directement sur la structure du signal reçu, et plus particulièrement sur la structure de la matrice de convolution $\mathcal{T}(h)$. Parmi les techniques déterministes, on trouve la méthode de "subspace fitting" [56], la méthode "Subchannel Response Matching" (SRM) [17, 18], la méthode de maximum de vraisemblance déterministe (DML) [9] ou encore les techniques de "least–squares smoothing" [25, 26].

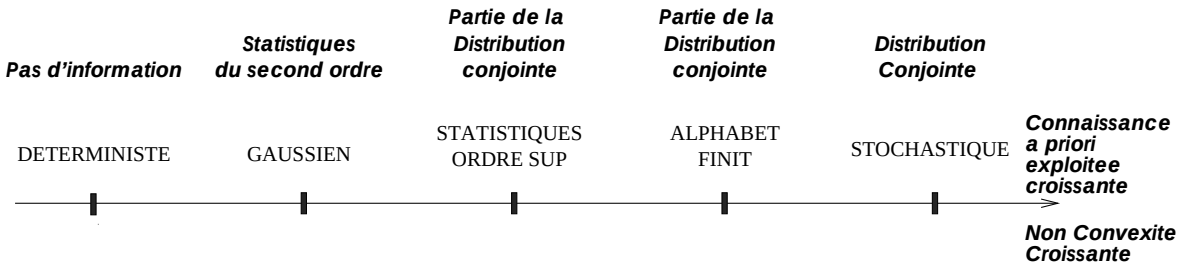



Figure 2: Classification des méthodes d'identification de canal selon la connaissance a priori exploitée sur les symboles d'entrée.

2. Statistiques du second ordre: les méthodes Gaussiennes.

   Ce sont des méthodes qui utilisent les moments du second–ordre des données, comme la méthode de prédiction [9] ou de covariance matching [37]. On trouve également la méthode de maximum de vraisemblance Gaussienne (GML) [40].

3. Statistiques d'ordre supérieur.

4. Alphabet fini des symboles d'entrée.

   Ces méthodes sont basées sur le signal reçu et exploitent en plus l'alphabet fini des symboles d'entrée. Parmi ces méthodes, on trouve les méthodes de maximum de vraisemblance telles que [45].

5. La distribution complète des symboles d'entrée.

   Dans ces méthodes, la vraie distribution des symboles d'entrée est exploitée; par exemple pour une BPSK on exploite le fait que les symboles prennent la valeur $+1$ ou $-1$ avec probabilité $\frac{1}{2}$. La méthode de maximum de vraisemblance stochastique (SML) [49] appartient à cette catégorie.

Cette classification peut être adaptée au cas semi–aveugle: les différentes catégories correspondent aux différentes modélisations des symboles inconnus.

Plus on exploite d'information sur les symboles d'entrée, meilleures vont être les performances d'estimation, mais, en même temps, plus coûteuses vont être les méthodes associées avec des fonctions de coût présentant des minima locaux. Dans cette thèse, nous nous sommes surtout concentrés sur les méthodes déterministes et Gaussiennes: on verra que certaines de ces méthodes peuvent être résolues de façon simple avec des fonctions de coût quadratiques, sans minima locaux donc.

# 3 Conditions d'Identifiabilité Semi–Aveugle

Un paramètre $\theta$ sera dit identifiable s'il peut être déterminé de façon unique à partir de sa fonction de densité de probabilité:

$$\forall \boldsymbol{Y}, f(\boldsymbol{Y}|\theta) = f(\boldsymbol{Y}|\theta') \Rightarrow \theta = \theta' . \tag{7}$$

Dans le cas de l'estimation aveugle déterministe et Gaussienne, l'identification du canal se fait au mieux à un facteur d'échelle ou de phase près; l'identifiabilité sera l'identifiabilité à l'ambiguïté aveugle près.

Nous résumons ici les conditions d'identifiabilité pour l'estimation de canal par apprentissage et pour l'estimation aveugle et semi–aveugle. On ne précise que les conditions sur le canal et le nombre de symboles connus; dans la thèse, des conditions sur la longueur du paquet du signal de réception, sur les modes d'excitation des symboles d'entrée sont également fournies.

## 3.1 Modèle Déterministe

- Estimation par apprentissage: n'importe quel canal peut être estimé pourvu qu'on ait $2N - 1$ symboles connus.
- Estimation aveugle: un canal irréductible sera identifiable à un facteur d'échelle près.
- Estimation semi–aveugle: n'importe quel canal peut être estimé pourvu qu'on ait $2N_c - 1$ symboles connus (non tous nuls). Ainsi:
  - pour un canal irréductible, il faut 1 symbole connu.
  - pour un monocanal, il faut $2N - 1$ symboles connus.

  Un résultat important démontré dans cette thèse est que ces dernières conditions sont valables pour des positions arbitraires des symboles connus dans le paquet. Le cas de symboles connus tous égaux à 0 est également traité: dans ce cas–là, la canal sera identifiable à un facteur d'échelle près avec $2N_c - 2$ symboles connus suffisamment dispersés dans le paquet, c'est–à–dire qu'on doit avoir au moins $N_c - 1$ symboles connus qui n'appartiennent pas à un groupe de $N_c$ symboles connus ou plus.

## 3.2 Modèle Gaussien

- Estimation aveugle: un canal irréductible sera identifiable à un facteur de phase près. Les zéros d'un canal sont identifiables mais on ne peut déterminer s'ils sont à minimum ou maximum de phase; ainsi, si on sait que le canal est à minimum de phase, il sera identifiable.

- Estimation semi–aveugle: n'importe quel canal sera identifiable pourvu qu'on ait un seul symbole connu (non nul) dans le paquet (qui n'est pas situé près des bords du paquet, *i.e.* dans les $N-1$ premiers ou derniers symboles).

Au point de vue identifiabilité, on voit donc que le modèle Gaussien est plus robuste que le modèle déterministe. De plus, l'estimation semi–aveugle a besoin de moins de symboles connus pour permettre l'identifiabilité que l'estimation par apprentissage.

# 4 Bornes de Performance

Pour caractériser les performances de l'estimation semi–aveugle, nous utilisons les bornes de Cramér–Rao (CRB). La plupart des définitions seront données pour le cas de paramètres réels; dans la thèse, nous traitons également le cas complexe.

La CRB donne une borne inférieure sur la matrice de corrélation $C_{\tilde{\theta}\tilde{\theta}}$ des erreurs d'un estimateur non biaisé:

$$C_{\tilde{\theta}\tilde{\theta}} \geq CRB = \mathcal{J}_{\theta\theta}^{-1} \tag{8}$$

$$\mathcal{J}_{\theta\theta} = E\left(\frac{\partial f(\boldsymbol{Y}|\theta)}{\partial \theta}\right)\left(\frac{\partial f(\boldsymbol{Y}|\theta)}{\partial \theta}\right)^T \tag{9}$$

où $\mathcal{J}_{\theta\theta}$ est la matrice d'information de Fisher (FIM).

## 4.1 CRB pour l'Estimation sous Contraintes

Pour les modèles déterministe et Gaussien, on démontre le théorème suivant:

**Theorem 25** *La FIM est non singulière en $\theta$ si et seulement si $\theta$ est identifiable localement (i.e. il existe un voisinage ouvert de $\theta$ sur lequel on a identifiabilité).*

L'estimation aveugle ne peut estimer la totalité des paramètres et il s'ensuit qu'on n'aura pas identifiabilité locale. La FIM aveugle est donc singulière et le résultat classique de la CRB (8) ne peut être directement appliqué.

L'estimation aveugle est en général faite sous contraintes: nous proposons donc des CRBs pour l'estimation sous contraintes. On considérera les contraintes sous la forme de la fonction implicite suivante:

$$\mathcal{K}_\theta = 0 \tag{10}$$

où $\mathcal{K}_\theta : \mathbb{R}^n \to \mathbb{R}^k$ est continûment différentiable. On note $n$ le nombre d'éléments de $\theta$. On supposera que ces contraintes résultent en une estimation non biaisée de $\theta$. On définit la matrice $\mathcal{V}_\theta$ telle que:

$$\text{espace}\left\{\mathcal{V}_\theta\right\} = \left(\text{espace}\left\{\frac{\partial \mathcal{K}_\theta^T}{\partial \theta}\right\}\right)^\perp \tag{11}$$

où espace $\{.\}$ désigne l'espace engendré par les colonnes de son argument et $(.)^{\perp}$ désigne l'espace orthogonal à son argument; on suppose que les colonnes de $\mathcal{V}_\theta$ sont indépendantes.

**Theorem 26** *La CRB sous les contraintes $\mathcal{K}_\theta = 0$ est*

$$CRB_C = \mathcal{V}_\theta \left(\mathcal{V}_\theta^T \mathcal{J}_{\theta\theta} \mathcal{V}_\theta\right)^{-1} \mathcal{V}_\theta^T \,. \tag{12}$$

*Cette CRB est définie si et seulement si $\mathcal{V}_\theta^T \mathcal{J}_{\theta\theta} \mathcal{V}_\theta$ est régulière, ce qui équivaut à dire que $\theta$ est localement identifiable sous les contraintes $\mathcal{K}_\theta$.*

## 4.2 Contraintes Particulières

Dans la thèse, nous avons mis en évidence des contraintes particulières importantes. Quand espace $\left\{\dfrac{\partial \mathcal{K}_\theta^T}{\partial \theta}\right\} = \text{null}(\mathcal{J}_{\theta\theta})$,

$$CRB_C = \mathcal{J}_{\theta\theta}^{+} \,. \tag{13}$$

Ces contraintes donnent, pour un nombre minimal de contraintes indépendantes, la plus petite valeur de $\text{tr}\{CRB_C\}$.

Ce résultat peut également être appliqué au cas de l'estimation conjointe des vecteurs de paramètres $\theta_1$ et $\theta_2$ où les contraintes $\mathcal{K}_{\theta_1}$ ne sont appliquées que sur $\theta_1$; On suppose que $\mathcal{J}_{\theta\theta}$ est singulière tandis que $\mathcal{J}_{\theta_2\theta_2}$ est régulière. La CRB sous la contrainte $\mathcal{K}_{\theta_1}$ est:

$$CRB = \mathcal{V}_{\theta_1} \left(\mathcal{V}_{\theta_1}^T \mathcal{J}_{\theta_1\theta_1}(\theta) \mathcal{V}_{\theta_1}\right)^{-1} \mathcal{V}_{\theta_1}^T \tag{14}$$

où

$$\mathcal{J}_{\theta_1\theta_1}(\theta) = \mathcal{J}_{\theta_1\theta_1} - \mathcal{J}_{\theta_1\theta_2} \mathcal{J}_{\theta_2\theta_2}^{-1} \mathcal{J}_{\theta_2\theta_1} \,. \tag{15}$$

$\mathcal{J}_{\theta_1\theta_1}^{-1}(\theta)$ serait la CRB pour $\theta_1$ si $\mathcal{J}_{\theta_1\theta_1}(\theta)$ était inversible. $\mathcal{V}_{\theta_1}^T \mathcal{J}_{\theta_1\theta_1}(\theta) \mathcal{V}_{\theta_1}$ est supposé inversible. Maintenant, si espace $\left\{\dfrac{\mathcal{K}_{\theta_1}^T}{\partial \theta}\right\} = \text{null}(\mathcal{J}_{\theta_1\theta_1}(\theta))$, la CRB contrainte est:

$$CRB_C = \left(\mathcal{J}_{\theta_1\theta_1}(\theta)\right)^{+} \,. \tag{16}$$

## 4.3 Application à l'Estimation Aveugle

On suppose ici que les conditions d'identifiabilité aveugle sont vérifiées.

### Modèle Déterministe

Dans le cas d'un canal complexe, la FIM déterministe a 2 singularités; son espace nul est engendré par:

$$h_{S_1} = \begin{bmatrix} \text{Re}(h) \\ \text{Im}(h) \end{bmatrix} \quad \text{et} \quad h_{S_2} = \begin{bmatrix} -\text{Im}(h) \\ \text{Re}(h) \end{bmatrix} \,. \tag{17}$$

Le premier vecteur peut être interprété comme correspondant à l'ambiguïté sur la norme et le second vecteur comme correspondant à l'ambiguïté de phase.

Pour l'estimation aveugle déterministe, on considère les contraintes suivantes:

- Une contrainte quadratique:
$$h^H h = h^{o\,H} h \tag{18}$$
qui permet d'ajuster la norme du canal.

- Dans le cas d'un canal complexe, une contrainte supplémentaire pour ajuster le facteur de phase:
$$h_{S_2}^{o\,T} h_R = h_{S_2}^{o\,T} h_R^o = 0\,. \tag{19}$$

On dénote $(.)^o$ la vraie valeur de son argument et $h_R = \left[\mathrm{Re}(h^T)\ \ \mathrm{Im}(h^T)\right]^T$. Ces contraintes laissent une ambiguïté de signe mais qui ne conduit pas à une FIM singulière. Pour évaluer le MSE, cette ambiguïté peut être résolue en imposant $h^{o\,T} h > 0$.

### Modèle Gaussien

La FIM Gaussienne pour un canal complexe a une seule singularité; l'espace nul est engendré par le vecteur $h_{S_2}$. Une seule contrainte est donc nécessaire pour régulariser le problème d'estimation; cette contrainte est la contrainte de phase utilisée dans le cas déterministe:

$$h_{S_2}^{o\,T} h_R = h_{S_2}^{o\,T} h_R^o = 0\,. \tag{20}$$

Pour les deux modèles, la CRB sous contraintes est:

$$CRB_C = \left(\mathcal{J}_{hh}(\theta)\right)^+\,. \tag{21}$$

Dans le cas déterministe, la contrainte de norme est usuellement utilisée, mais sans autre justification que celle de donner des solutions simples pour les méthodes associées. Ici, nous choisissons également cette contrainte parce que, adjointe de la contrainte de phase (19), elle donne comme CRB la pseudo inverse de $\mathcal{J}_{hh}(\theta)$. C'est ainsi que l'on justifie également la contrainte de phase choisie dans le cas Gaussien.

## 4.4 Évaluations Numériques des CRBs

Dans la figure 3, nous montrons la mesure $\mathrm{tr}(CRB)$, pour un paquet de longueur fixe $M = 100$, le canal $\boldsymbol{H}_{well}$ (voir Chapitre 5, appendice D). Les symboles d'entrée appartiennent à une BPSK et sont choisis aléatoirement. Nous ne présentons ici que le cas déterministe.

Dans la figure 3 à gauche, les performances sont montrées en fonction du nombre de symboles connus dans le paquet. On voit une amélioration spectaculaire des performances

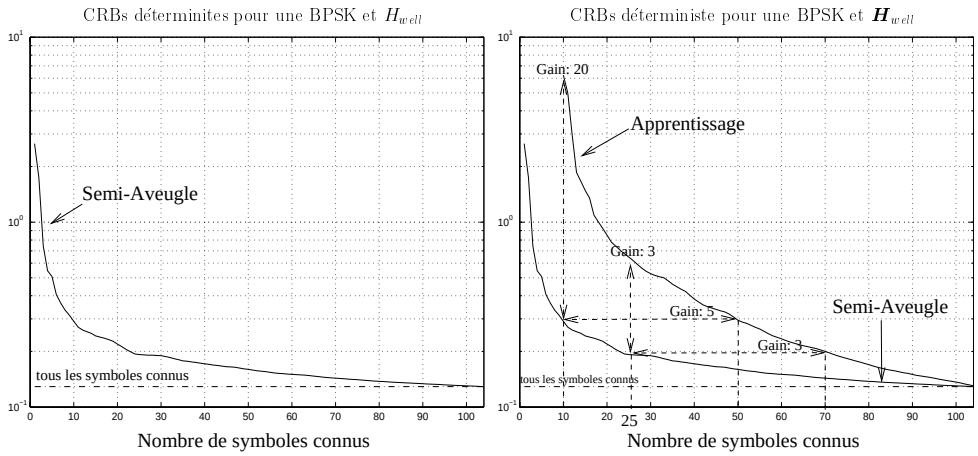



Figure 3: CRBs pour l'estimation semi–aveugle (gauche) et comparaison des modes semi–aveugle et d'apprentissage (droite).

semi–aveugles avec très peu de symboles connus (10 symboles connus); passé ce stade, il faut un nombre plus important de symboles connus pour avoir une amélioration significative.

Dans la figure 3 à droite, on compare les performances de l'estimation semi–aveugle avec l'estimation par apprentissage. Pour 10 symboles connus, on a un gain de performance de 20 apporté par l'estimation semi–aveugle. Pour un niveau de qualité d'estimation désiré, il faudrait 10 symboles connus pour l'estimation semi–aveugle alors qu'il en faudrait 50 pour l'estimation par apprentissage. Pour 25 symboles connus, on a un gain de performance de 3; il faudrait 70 symboles connus en apprentissage pour obtenir les performances de l'estimation semi–aveugle.

## 4.5 Optimisation des Symboles Connus

### Valeur des Symboles Connus

Pour l'estimation par apprentissage, la séquence optimale est une séquence blanche (au sens déterministe), c'est–à–dire telle que $\mathcal{A}^H\mathcal{A}$ est une diagonale constante. Pour l'estimation semi–aveugle, l'optimisation de la séquence va dépendre du canal: une séquence blanche ne va pas optimiser les performances mais reste sans doute parmi les meilleurs choix.

### Symboles Groupés ou Dispersés

Lorsque les symboles connus sont groupés, on aura de meilleures performances lorsque ces symboles sont choisis aléatoirement. Par contre, lorsqu'ils sont tous identiques, les performances seront meilleures lorsque les symboles connus sont dispersés dans le paquet.

# 5 Méthodes Semi–Aveugles Optimales

Dans une première approche, nous nous sommes intéressés à des méthodes semi–aveugles optimales: ce sont des méthodes qui permettent d'exploiter toute l'information venant des symboles connus et des symboles inconnus, et notamment des zones de transitions qui contiennent à la fois des symboles connus et des symboles inconnus. De telles méthodes sont des méthodes qui incorporent naturellement la connaissance de symboles à l'entrée. Ainsi, nous avons opté pour les méthodes de maximum de vraisemblance semi–aveugles.

### DML Semi–Aveugle

$\boldsymbol{Y} = \mathcal{T}(h)A + \boldsymbol{V}$: comme $\boldsymbol{V} \sim \mathcal{N}(0, \sigma_v^2 I)$, $\boldsymbol{Y} \sim \mathcal{N}(\mathcal{T}(h)A, \sigma_v^2 I)$. Le critère DML s'écrit comme:

$$\max_{A_U,h} f(\boldsymbol{Y}|h) \quad \Leftrightarrow \quad \min_{A_U,h} \|\boldsymbol{Y} - \mathcal{T}(h)A\|^2 \quad \Leftrightarrow \quad \min_{A_U,h} \|\boldsymbol{Y} - \mathcal{T}_K(h)A_K - \mathcal{T}_U(h)A_U\|^2 \tag{22}$$

où $f(\boldsymbol{Y}|h)$ est la densité de probabilité de $\boldsymbol{Y}$. Quand on minimise ce critère en fonction de $A_U$ et on remplace l'expression obtenue dans le critère, on obtient un critère DML pour $h$ uniquement:

$$\min_h \left(\boldsymbol{Y} - \mathcal{T}_K(h)A_K\right)^H P^{\perp}_{\mathcal{T}_U(h)} \left(\boldsymbol{Y} - \mathcal{T}_K(h)A_K\right) . \tag{23}$$

Le premier critère (22) peut être optimisé par des minimisations alternées entre $h$ et $A_U$. Cette méthode simple présente un certain nombre de propriétés avantageuses: la fonction de coût DML décroît à chaque itération et avec une bonne initialisation, on converge vers le minimum global. Cette méthode a pour désavantage de présenter une convergence lente qui rend son utilisation quelque peu prohibitive. Le critère (23) en $h$ peut être résolu par la méthode de scoring [40].

### GML Semi–Aveugle

$\boldsymbol{Y} = \mathcal{T}_K(h)A_K + \mathcal{T}_U(h)A_U + \boldsymbol{V}$: comme $A_U \sim \mathcal{N}(0, \sigma_a^2 I)$ et $\boldsymbol{V} \sim \mathcal{N}(0, \sigma_v^2 I)$, alors $\boldsymbol{Y} \sim \mathcal{N}(\mathcal{T}_K(h)A_K, C_{YY})$ avec $C_{YY} = \sigma_a^2 \mathcal{T}_U(h)\mathcal{T}_U^H(h) + \sigma_v^2 I$. Le critère GML s'écrit comme:

$$\min_{h,\sigma_v^2} \left\{ \ln\det C_{YY} + \left(\boldsymbol{Y} - \mathcal{T}_K(h)A_K\right)^H C_{YY}^{-1} \left(\boldsymbol{Y} - \mathcal{T}_K(h)A_K\right) \right\} . \tag{24}$$

Ce critère peut être optimisé par la méthode de scoring. Il est à noter que l'hypothèse Gaussienne sur les symboles inconnus, qui est une hypothèse non valide, n'est utilisée que pour construire le critère. Dans la thèse, une forte connexion entre GML aveugle et la méthode de covariance matching basée sur un critère aux moindres carrés optimalement pondéré. On montre que GML peut être vu comme une certaine forme de covariance matching et a des performances équivalentes à la méthode de covariance matching quand le nombre de données et la taille de la matrice de covariance tendent vers l'infini.

Une étude de performance des méthodes MV déterministes et Gaussiennes est présentée dans le chapitre 6.

Dans une seconde approche qui est sous–optimale, on considère que les symboles connus sont groupés sous forme d'une séquence d'apprentissage. Dans ce cas–là, on verra, qu'avec peu d'approximations, le critère MV semi–aveugle s'écrira sous la forme de la somme du critère MV et du critère d'apprentissage. Comme le problème semi–aveugle garde la structure du problème aveugle, il va être possible de construire des algorithmes rapides semi–aveugles. Ainsi, pour des raisons de complexité algorithmique, mais également pour des raisons de performance, il est préférable d'avoir des symboles connus groupés.

Pour avoir des critères semi–aveugles performants, il apparaît important de trouver des méthodes performantes pour résoudre le problème MV aveugle.

## 6 DML Aveugle

Une solution peu coûteuse pour optimiser le critère DML aveugle est basée sur une paramétrisation de l'espace bruit (*i.e.* complément orthogonal à l'espace colonne de $\mathcal{T}(h)$). Une telle paramétrisation $\mathbf{H}^{\perp}(z)$ vérifie $\mathbf{H}^{\perp}(z)\mathbf{H}(z) = 0$ et $\mathcal{T}^H(h^{\perp})\mathcal{T}(h) = 0$ où $\mathcal{T}(h^{\perp})$ est la matrice de convolution construite à partir de $\mathbf{H}^{\perp}(z)$. Par exemple, dans le cas $m = 2$ (2 sous–canaux), $\mathbf{H}^{\perp}(z) = [-\mathrm{H}_2(z) \;\; \mathrm{H}_1(z)]$. Le critère DML pour $h$ s'écrit alors comme:

$$\min_{\|h\|=1} \boldsymbol{Y}^H \mathcal{T}^H(h^{\perp}) \underbrace{\left[\mathcal{T}(h^{\perp})\mathcal{T}^H(h^{\perp})\right]^{+}}_{\mathcal{R}(h)} \mathcal{T}(h^{\perp})\boldsymbol{Y} \,. \tag{25}$$

On ne précise que la contrainte de norme sur le canal: une contrainte de phase est également nécessaire dans le cas d'un canal complexe.

Iterative Quadratic ML (IQML) est un algorithme itératif qui permet de résoudre (25): à chaque itération, IQML considère le dénominateur $\mathcal{R}(h) = \mathcal{R}$ comme constant, calculé à partir de l'itération précédente. Le critère devient alors quadratique. Utilisant la commutativité de la convolution, on peut écrire $\mathcal{T}(h^{\perp})\boldsymbol{Y} = \mathcal{Y}h$; le critère IQML s'écrit:

$$\min_{\|h\|=1} h^H \mathcal{Y}^H \mathcal{R}^{+} \mathcal{Y} h \,. \tag{26}$$

Dans le cas sans bruit, $\mathcal{Y}h^o = 0$: le vrai canal annule le critère (quadratique) et donc la solution est $h^o$. A haut SNR, une première itération donnera une estimée consistente de $h$ et une seconde itération donnera la solution DML. A bas SNR cependant, IQML donne une estimation biaisée du canal et ses performances sont mauvaises. En effet, asymptotiquement dans le nombre des donne'es, la fonction de coût IQML est équivalente à son espérance par

la loi des grands nombres:

$$\mathcal{T}^H(h^\perp)\mathcal{R}^+\mathcal{T}(h^\perp)\mathrm{E}\boldsymbol{Y}\boldsymbol{Y}^H = \mathrm{tr}\left\{\mathcal{T}^H(h^\perp)\mathcal{R}^+\mathcal{T}(h^\perp)\boldsymbol{X}\boldsymbol{X}^H\right\} + \sigma_v^2\mathrm{tr}\left\{\mathcal{T}^H(h^\perp)\mathcal{R}^+\mathcal{T}(h^\perp)\right\}. \tag{27}$$

$h^o$ annule le premier terme mais n'est pas en général un vecteur propre du second terme et donc de la somme.

## 6.1 Denoised IQML (DIQML)

La première méthode que nous proposons va débruiter le critère IQML. Nous soustrayons au critère IQML une approximation de la contribution du bruit ($\widehat{\sigma_v^2}$ est une estimée consistente de la variance du bruit):

$$\begin{aligned}&\min_{\|h\|=1}\mathrm{tr}\left\{P_{\mathcal{T}^H(h^\perp)}\left(YY^H - \widehat{\sigma_v^2}I\right)\right\} \quad \Leftrightarrow \\ &\min_{\|h\|=1}\left\{h^H\mathcal{Y}^H\mathcal{R}^+(h)\mathcal{Y}h - \widehat{\sigma_v^2}\mathrm{tr}\{\mathcal{T}(h^\perp)\mathcal{R}^+(h)\mathcal{T}^H(h^\perp)\}\right\}.\end{aligned} \tag{28}$$

(28) est résolu de la même manière que IQML: on considère $\mathcal{R}(h) = \mathcal{R}$ comme constant à chaque itération et le problème devient quadratique:

$$\min_{\|h\|=1} h^H\left\{\mathcal{Y}^H\mathcal{R}^+\mathcal{Y} - \widehat{\sigma_v^2}\mathcal{D}\right\}h \tag{29}$$

où $h^H\mathcal{D}h = \mathrm{tr}\{\mathcal{T}^H(h^\perp)\mathcal{R}^+\mathcal{T}(h^\perp)\}$.

Le choix de $\widehat{\sigma_v^2}$ s'avère crucial. Pour un nombre fini de données, la matrice centrale $\mathcal{Q} = \mathcal{Y}^H\mathcal{R}^+\mathcal{Y} - \widehat{\sigma_v^2}\mathcal{D}$ sera indéfinie si $\widehat{\sigma_v^2}$ n'est pas choisi proprement: dans ce cas là, DIQML n'aura pas de bonnes performances. Pour avoir un problème de minimisation bien défini à chaque itération, on va choisir le $\widehat{\sigma_v^2}$ qui rend $\mathcal{Q}(h)$ positive avec 1 singularité. Le problème devient alors:

$$\min_{\|h\|=1,\lambda} h^H\left\{\mathcal{Y}^H\mathcal{R}^+\mathcal{Y} - \lambda\mathcal{D}\right\}h\,. \tag{30}$$

$\lambda$ est la valeur propre généralisée minimale de $\mathcal{Q}$ et $h$ le vecteur propre généralisé associé.

## 6.2 Pseudo Quadratic ML (PQML)

PQML est un algorithme itératif qui à chaque itération va tenter d'annuler le vrai gradient de DML. Ce gradient peut être décomposé sous la forme $\mathcal{P}(h)h$ où $\mathcal{P}(h)$ est une matrice idéalement positive. A chaque itération $\mathcal{P}(h) = \mathcal{P}$ est considéré comme constant (évalué grâce à l'itération précédente). Le problème devient quadratique:

$$\min_{\|h\|=1} h^H\mathcal{P}\,h\,. \tag{31}$$

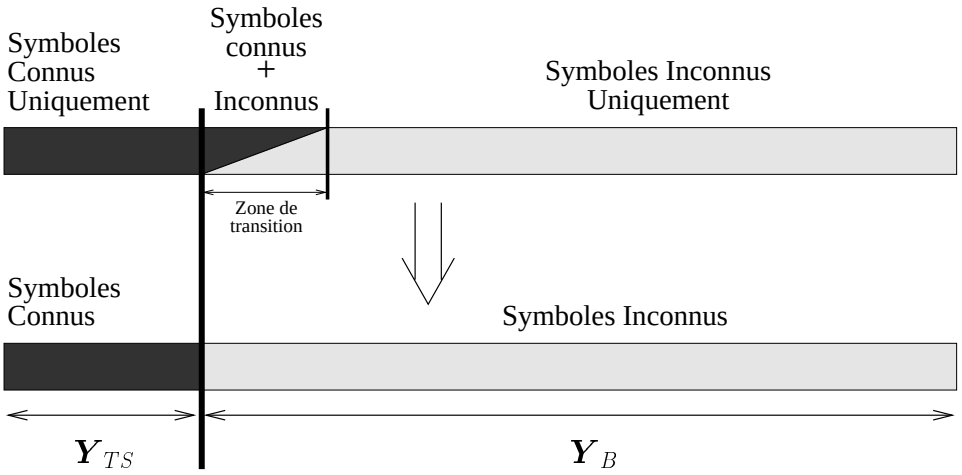



Figure 4: Paquet de sortie: décomposition des données pour LS–PQML

Pour le problème DML, la matrice $\mathcal{P}(h)$ peut être mise sous la forme $\mathcal{P}(h) = \mathcal{Y}^H \mathcal{R}^+ \mathcal{Y} - \mathcal{B}(h)$. Quand $M \to \infty$, $\mathcal{B}(h) \to \sigma_v^2 \mathcal{D}$ + terme signal. Évalué en un $h$ consistent, le terme signal devient négligeable, et l'effet de $\mathcal{B}(h)$ est de débruiter le critère IQML mais d'une façon plus efficace que DIQML.

Les performances de PQML et DIQML sont étudiées: PQML a de meilleures performances que DIQML. De plus PQML a les mêmes performances asymptotiques que DML.

# 7 Méthodes Semi–Aveugles Sous–Optimales

On suppose que les symboles connus sont groupés et pour simplifier, sont situés au début du paquet. Le paquet de sortie (figure 4) peut être décomposé en 3 zones: une zone ne contenant que des symboles connus, une zone contenant à la fois des symboles connus et inconnus et une zone ne contenant que des symboles inconnus.

Les critères semi–aveugles que nous proposons vont être basés sur une décomposition en 2 zones, le zone de transition étant assimilée à la partie aveugle ou à la partie apprentissage du critère.

## 7.1 Least Squares–PQML (LS–PQML)

La première méthode assimile la zone de transition à la partie aveugle et les symboles connus dans cette zone sont considérés comme inconnus, ce qui entraîne une perte d'information. On applique DML à $\left[\boldsymbol{Y}_{TS}^T \;\; \boldsymbol{Y}_B^T\right]^T$ avec $\boldsymbol{Y}_{TS} = \mathcal{T}_{TS}(h) A_{TS} + \boldsymbol{V}_{TS}$ et $\boldsymbol{Y}_B = \mathcal{T}_B(h) A_B + \boldsymbol{V}_B$. Le critère semi–aveugle s'écrit:

$$\min_{h, A_U} \left\{ \|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_{TS}\|^2 + \|\boldsymbol{Y}_B - \mathcal{T}_B(h) A_B\|^2 \right\} \tag{32}$$

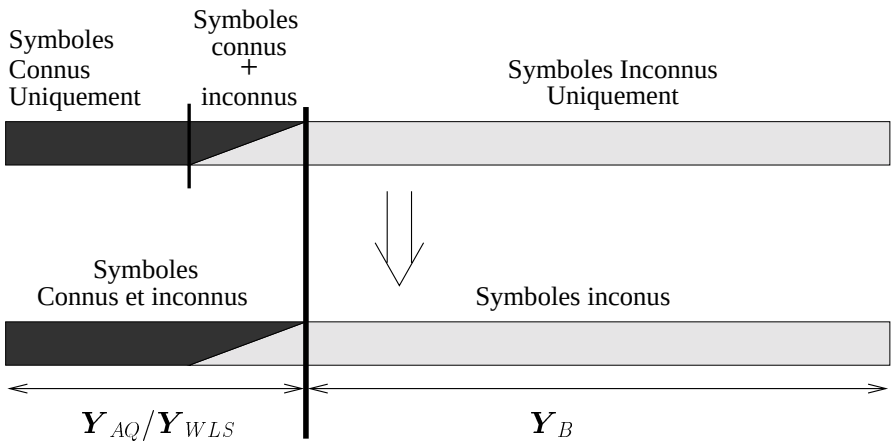



Figure 5: Paquet de sortie: décomposition des données pour AQ/WLS–PQML

c'est–à–dire la somme du critère aux moindres carrés pour $\boldsymbol{Y}_{TS}$ et du critère DML pour $\boldsymbol{Y}_B$. Ce critère semi–aveugle peut être résolu en utilisant des minimisations alternées entre $h$ et $A_B$. On peut également utiliser le critère DML en $h$ pour la partie aveugle en la résolvant par PQML:

$$\min_{h,\lambda} \left\{ \|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_{TS}\|^2 + h^H \left\{ \mathcal{Y}_B^H \mathcal{R}^+ \mathcal{Y}_B - \lambda \mathcal{B} \right\} h \right\} . \tag{33}$$

## 7.2 Alternating Quadratic–PQML (AQ–PQML)

Dans la deuxième méthode, la zone de transition est assimilée à la partie apprentissage, les symboles inconnus dans cette zone sont considérés comme déterministes. On applique DML à $\left[\boldsymbol{Y}_{AQ}^T \;\; \boldsymbol{Y}_B^T\right]^T$ avec $\boldsymbol{Y}_{AQ} = \mathcal{T}'_K(h) A_{TS} + \mathcal{T}'_U(h) A'_U + \boldsymbol{V}_{AQ}$. Le critère semi–aveugle s'écrit:

$$\min_{A_B, A'_U, h} \left\{ \|\boldsymbol{Y}_{AQ} - \mathcal{T}'_K(h) A_{TS} - \mathcal{T}'_U(h) A'_U\|^2 + \|\boldsymbol{Y}_B - \mathcal{T}_B(h) A_B\|^2 \right\} . \tag{34}$$

Ce critère peut être optimisé en utilisant des minimisations alternées. Si on résoud la partie aveugle par PQML:

$$\min_{A'_U, h, \lambda} \left\{ \|\boldsymbol{Y}_{AQ} - \mathcal{T}'_K(h) A_{TS} - \mathcal{T}'_U(h) A'_U\|^2 + h^H \left\{ \mathcal{Y}_B^H \mathcal{R}^+ \mathcal{Y}_B - \lambda \mathcal{B} \right\} h \right\} \tag{35}$$

et on résoud par minimisations alternées entre $h$ et $A'_U$.

## 7.3 Weighted Least Squares–PQML (WLS–PQML)

Dans la troisième méthode, la zone de transition est de nouveau assimilée à la partie apprentissage, mais les symboles inconnus dans cette zone sont cette fois–ci considérés comme des

variables aléatoires Gaussiennes. On applique GML à $\boldsymbol{Y}_{WLS} = \mathcal{T}'_K(h)A_{TS}+\mathcal{T}'_U(h)A'_U+\boldsymbol{V}_{WLS}$ et DML à $\boldsymbol{Y}_B$, and après quelques approximations, on obtient:

$$\min_{A_B,h} \left\{ \|\boldsymbol{Y}_{WLS} - \mathcal{T}'_K(h)A_{TS}\|^2_{C_{Y_{WLS}Y_{WLS}}} + \|\boldsymbol{Y}_B - \mathcal{T}_B(h)A_B\|^2 \right\} \tag{36}$$

avec $C_{Y_{WLS}Y_{WLS}} = \sigma_a^2\mathcal{T}_{WLS}(h)\mathcal{T}^H_{WLS}(h)+\sigma_v^2 I$. On peut de nouveau utiliser des minimisations alternées. WLS–PQML s'écrit:

$$\min_{h,\lambda} \left\{ \|\boldsymbol{Y}_{WLS} - \mathcal{T}'_K(h)A_{TS}\|^2_{C_{Y_{WLS}Y_{WLS}}} + h^H \left\{ \mathcal{Y}_B^H \mathcal{R}^+ \mathcal{Y}_B - \lambda \mathcal{B} \right\} h \right\} . \tag{37}$$

Dans la figure 6, on montre les performances simulées des différents algorithmes basés sur PQML pour un paquet de longueur $M = 100$, un SNR de 10dB, une longueur de canal de 4 (les coefficients du canal sont choisis aléatoirement). 1000 réalisations de Monte–Carlo sur les symboles d'entrée inconnus et sur le bruit sont effectuées. On a 7 symboles connus, c'est–à–dire le nombre minimal pour que l'estimation par apprentissage soit définie. On remarque que la méthode la plus performante est WLS–PQML avec des performances simulées proches des performances optimales de ML, et la moins performante est LS-PQML. On montre de plus les performances de l'algorithme de minimisations alternées basées sur le critère optimal (22). On remarque la convergence lente de cet algorithme.

Dans la figure 7, on effectue 5000 réalisations de Monte–Carlo sur le canal (aléatoire), le bruit et les symboles d'entrée. On y montre les performances du PQML aveugle (le facteur d'échelle du canal est estimé grâce à la séquence d'apprentissage). On remarque que le PQML aveugle a de mauvaises performances: ceci n'est pas dû au fait que PQML est un mauvais algorithme, mais au fait que le problème aveugle est très sensible au conditionnement du canal: sur les 5000 réalisations, des canaux mal conditionnés ont été tirés qui ont rendu les performances aveugles moyennes mauvaises. L'estimation semi–aveugle permet par contre de robustifier le problème.

# 8 Combinaison Critère Aveugle et Critère Apprentissage

## 8.1 Exemple de SRM

Certains critères semi–aveugles se présentent sous la forme d'une combinaison linéaire d'un critère aveugle (critère qui ne permet pas d'incorporer la connaissance de symboles connus) et d'un critère d'apprentissage.

Nous traitons ici l'exemple de SRM semi–aveugle. SRM aveugle peut être vu comme une version non pondérée de IQML: $\min_{\|h\|=1} h^H \mathcal{Y}^H \mathcal{Y}\ h$. Le critère SRM semi–aveugle s'écrit comme:

$$\min_h \left\{ \alpha h^H \mathcal{Y}_B^H \mathcal{Y}_B h + \|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h)A_{TS}\|^2 \right\} . \tag{38}$$

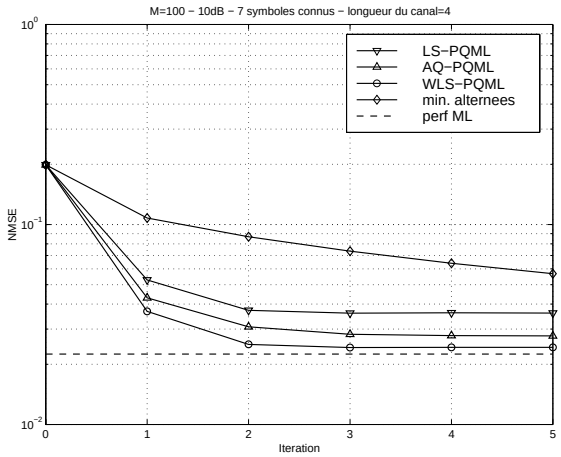



Figure 6: Algorithmes semi–aveugles.

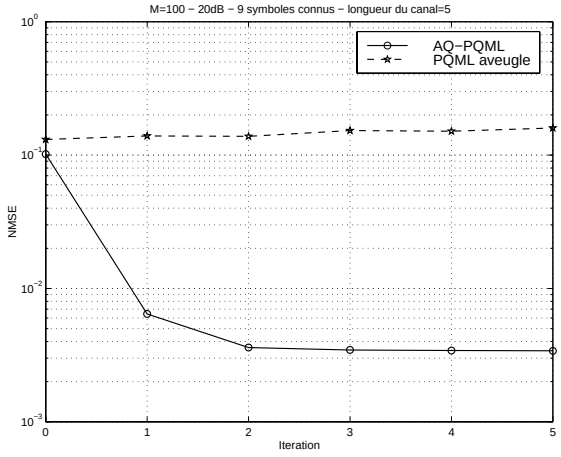



Figure 7: Comparaison entre un algorithme semi–aveugle et un algorithme aveugle.

Ce critère est basée sur la décomposition des données de la figure 4. Une façon intuitive de pondérer les deux parties aveugle et apprentissage est de leur associer le nombre de données sur lesquelles elles sont basées. Ainsi, dans notre cas, le $\alpha$ optimal serait égal à 1. Dans la figure 8, nous montrons le NMSE moyenné sur 100 réalisations du canal (dont les coefficients sont choisis aléatoirement), du bruit et des symboles d'entrée. On considère les cas de 10 et 20 symboles connus pour un burst de longueur 100. Pour $\alpha = 1$, on trouve des performances pires que celles données par apprentissage pur ($\alpha = 0$).

Cet exemple nous montre qu'un critère semi-aveugle bien construit ne consiste pas en une simple combinaison linéaire du critère aveugle et du critère d'apprentissage. Le critère SRM aveugle ne donne une estimée du canal non biaisé que sous la contrainte de norme constante. L'estimation semi–aveugle se faisant sans contraintes, les mauvaises performances de la partie aveugle rendent le critère semi–aveugle mauvais.

Il apparaît donc nécessaire de débruiter la partie SRM du critère semi-aveugle, il reste cependant à trouver le bon facteur $\alpha$. Pour cela, nous faisons un parallèle avec DIQML, et nous approximons la matrice $\mathcal{R}$ (30) par sa diagonale. Ainsi dans le cas $m = 2$, le critère SRM semi–aveugle est:

$$\min_h \left\{ \frac{1}{\widehat{\|h\|}^2} h^H \left( \mathcal{Y}_B^H \mathcal{Y}_B - \lambda_{min}(\mathcal{Y}_B^H \mathcal{Y}_B) \right) h + \| \boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_{TS} \|^2 \right\} \tag{39}$$

où $\lambda_{min}(\mathcal{Y}_B^H \mathcal{Y}_B)$ est la valeur propre minimale de $\mathcal{Y}_B^H \mathcal{Y}_B$, et $\widehat{\|h\|}$ est une estimée de la norme de $h$. Dans la figure 8, on montre les performances du SRM semi–aveugle corrigé. On voit que la valeur $\alpha = 1$ donne approximativement les performances optimales. On remarque en fait que les performances dépendent assez peu du facteur $\alpha$.

## 8.2 Autre Exemple

Considérons maintenant un critère semi–aveugle formé d'un critère aveugle quadratique de la forme $h^H \mathcal{U}_B^H \mathcal{U}_B h$, avec $\mathcal{U}_B h^o \overset{M_U \to \infty}{\to} 0$. Considérant la situation asymptotique où le nombre de symboles connus et inconnus tendent vers l'infini, on connaît la matrice de pondération optimale $W$ du critère aux moindres carrés pondérés:

$$\min_h \left\| \begin{array}{c} \mathcal{U}_B h \\ \boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_{TS} \end{array} \right\|_{W^+}^2 \tag{40}$$

(ce critère est basé sur la décomposition de la figure 4). En effet:

$$W = \mathrm{E} \left[ \begin{array}{c} \mathcal{U} h \\ \boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_{TS} \end{array} \right] \left[ \begin{array}{c} \mathcal{U} h \\ \boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_{TS} \end{array} \right]^H = \left[ \begin{array}{cc} W_B & 0 \\ 0 & \sigma_v^2 I \end{array} \right] \tag{41}$$

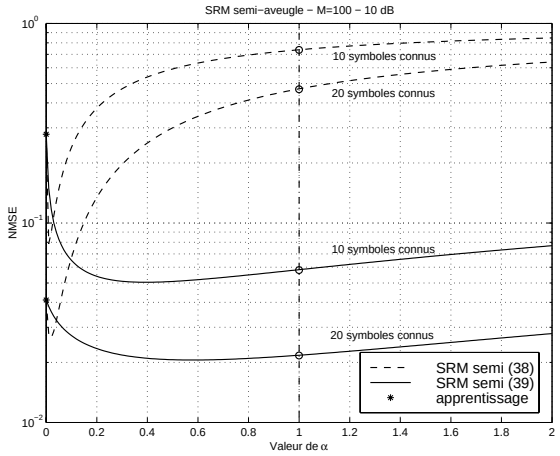



Figure 8: SRM semi–aveugle.

avec $W_B = \mathrm{E}\left(\mathcal{U}_B h h^H \mathcal{U}_B^H\right)$. Le critère semi–aveugle pondéré optimalement devient alors:

$$\min_h \left\{ h^H \mathcal{U}_B^H \, W_B^+ \, \mathcal{U}_B h + \frac{1}{\sigma_v^2} \| \boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_{TS} \|^2 \right\} \tag{42}$$

# 9 Gaussian ML

Des algorithmes aveugles et semi–aveugles pour optimiser GML basés sur la méthode de scoring sont également proposés. Dans des simulations, on met en évidence les performances supérieures de GML par rapport à DML. La méthode de scoring est cependant plus algorithmiquement complexe que l'algorithme DIQML ou PQML. Nous proposons une approximation de la méthode de scoring basée sur des expressions fréquentielles de la FIM gaussienne et du gradient de GML.

Dans le cas Gaussien, on peut également construire des critères semi–aveugles sous la forme de la somme d'un critère aveugle et d'un critère d'apprentissage comme cela a été fait dans le cas déterministe. Le découpage des données en deux parties devrait être différent cependant: en effet, non seulement le bruit au niveau de ces deux parties doit être découplé, mais la partie signal également. Cependant, nous avons vérifié par des simulations qu'on peut ignorer les corrélations au niveau signal et donc adopter les mêmes décompositions (figure 4 et 5) que dans le cas déterministe.

# 10 Décisions Douces Appliquées à l'Estimation Semi–Aveugle

Dans une dernière étape, nous appliquons une stratégie de décisions douces à l'estimation de canal semi–aveugle. On part d'une estimée du canal obtenue par une méthode semi–aveugle, à partir de laquelle un égaliseur est construit qui nous donne des estimés des symboles d'entrée inconnus. Des décisions sur les symboles les plus fiables sont prises et sont considérées comme des symboles connus; les autres estimés des symboles, non fiables, sont considérés comme inconnus. Un nouveau critère semi–aveugle est alors construit basé sur un nombre de symboles connus plus important. Cette idée prometteuse est appliquée à DML et à GML: il apparaît en fait difficile d'appliquer cette stratégie car elle change les statistiques conjointes de symboles et du bruit.

# 11 Structures de Récepteurs

## 11.1 Égaliseur en Mode Paquet

On étudie la structure des égaliseurs en mode paquet: on donne la structure des égaliseurs classiques linéaires et à retour de décision, ainsi que celle de l'annulateur d'interférences entre symboles, que l'on appellera NCDFE (Non Causal Decision Feedback Equalizer), qui utilise les décisions passées et futures: les décisions futures sont données par une autre égaliseur (linéaire ou DFE), les décisions passées sont données par cet autre égaliseur ou le NCDFE lui–même. Les différents égaliseurs sont donnés selon les critères zéro–forcing, MMSE et MMSE non biaisé. Contrairement au mode continu, les filtres optimaux sont variants dans le temps. Les performances des ces égaliseurs sont données en terme de SNR. Le SNR dépend de la position du symbole dans le paquet, ainsi que de la présence de symboles connus dans le paquet. Nous montrons qu'en choisissant correctement le nombre et la position des symboles connus, les filtres du traitement continu (qui sont invariants dans le temps) peuvent être organisés pour donner des performances satisfaisantes, de telle façon que le traitement optimal en mode paquet impliquant des filtres variants dans le temps peut être évité.

## 11.2 Égaliseur à Retour de Décision Non Causal

Quand il n'y a pas d'erreurs dans le retour de décision non causal du NCDFE, celui–ci donne en sortie un signal sans interférences entre symboles, et le "matched filter bound" est atteint. En pratique, il souffre de la propagation d'erreurs. Nous proposons une implémentation du NCDFE basée sur des décisions douces. A chaque étape, pour chaque symbole inconnu, des intervalles de fiabilité sont calculés sur la base du SNR du symbole, de la position du symbole dans le paquet et de la présence de symboles connus dans le paquet. Uniquement les symboles les plus fiables sont mis dans le retour de décisions du NCDFE. Les autres symboles sont classifiés comme inconnus et seront de nouveau estimés dans la prochaine itération du

NCDFE. Le nombre croissant de symboles assimilés comme connus, venant des itérations précédentes du NCDFE, améliore la qualité d'estimation des autres symboles inconnus. Ce schéma de décisions douces est comparé à d'autres schémas, et en particulier celui basé sur la tangente hyperbolique.

### 11.3 Matched Filter Bounds pour des Modèles de Canaux d'Ordre Réduit

Le Matched Filter Bound (MFB) correspond à une borne de performance pour les récepteurs quand le canal est connu. Nous proposons ici deux Matched Filter Bounds (MFB) destinés à caractériser les performances de récepteurs qui utilisent des modèles de canaux d'ordre réduit. La première borne utilise le modèle du canal pour effectuer un filtrage adapté spatio–temporel qui conduit à une réduction des données de multicanal à monocanal. Le reste du traitement reste optimal. La seconde mesure (ICMFB) borne les performances de l'algorithme de Viterbi basé sur le modèle de canal d'ordre réduit. Nous présentons deux méthodes fournissant des modèles d'ordre réduit pour illustrer ces deux mesures: l'estimation de canal aveugle par DML (qui maximise WMFB) et l'estimation de canal par apprentissage.

## 12 Conclusion

Dans cette thèse, nous avons présenté une étude approfondie sur l'estimation de canal semi–aveugle. Nous avons montré la supériorité des techniques semi–aveugles par rapport aux techniques aveugles et aux techniques d'apprentissage. Les méthodes semi–aveugles sont capables d'identifier n'importe quel canal avec très peu de symboles connus; ces derniers peuvent de plus être arbitrairement dispersés dans le paquet d'entrée. Les performances de l'estimation semi–aveugle sont également meilleures que celles de l'estimation aveugle et de l'estimation par apprentissage. Nous avons de plus fourni une étude sur les CRBs sous contraintes afin de caractériser les performances de l'estimation aveugle. Il existe différentes façons de construire des critères semi–aveugles. On a vu des critères optimaux qui sont basés sur des méthodes incorporant naturellement la connaissance de symboles d'entrée. Des critères sous–optimaux basés sur MV qui considèrent des symboles groupés et s'écrivent sous forme d'une combinaison linéaire du critère aveugle et du critère d'apprentissage: les poids de la combinaison linéaire sont donnés par MV. Enfin, nous avons examiné des critères combinant certains critères aveugles, comme subspace fitting, et un critère d'apprentissage. Dans une dernière partie, nous nous sommes concentrés sur des structures de récepteurs avec un intérêt particulier pour le NCDFE.

# BIBLIOGRAPHY


[1] Raymond Steele, *Mobile Radio Communications*, Pentech Press Limited, 1992.

[2] Y. Sato, "A Method of Self–Recovering Equalization for Multilevel Amplitude Modulation Systems," *IEEE Transactions on Communications*, vol. 23, pp. 679–682, June 1975.

[3] Editor Simon Haykin, *Blind Deconvolution*, Prentice Hall, 1994.

[4] W.A. Gardner, "A New Method of Channel Identification," *IEEE Transactions on Communications*, vol. 39, pp. 813–817, June 1991.

[5] L. Tong, G. Xu, and T. Kailath, "A New Approach to Blind Identification and Equalization of Multipath Channels," in *Proc. of the 25th Asilomar Conference on Signals, Systems & Computers*, Pacific Grove, CA, Nov. 1991, pp. 856–860.

[6] L. Tong, G. Xu, and T. Kailath, "Fast Blind Equalization Via Antenna Arrays," in *Proc. ICASSP*, Minneapolis, MN, April 27-30 1993, vol. IV, pp. 272–275.

[7] P.A. Laurent, "Exact and Approximate Construction of Digital Phase Modulations by Superposition of Amplitude Modulated Pulses (AMP)," *IEEE Trans. Communications*, vol. 34, February 1986.

[8] H. Trigui and D.T.M. Slock, "Cochannel Interference Cancellation within the Current GSM Standard," in *Proc. Workshop COST254*, July 7-9 1997, Toulouse, France.

[9] D.T.M. Slock, "Blind Fractionally-Spaced Equalization, Perfect-Reconstruction Filter Banks and Multichannel Linear Prediction," in *Proc. ICASSP 94 Conf.*, Adelaide, Australia, April 1994.

[10] D.T.M. Slock and C.B. Papadias, "Blind Fractionally-Spaced Equalization Based on Cyclostationarity," in *Proc. Vehicular Technology Conf.*, Stockholm, Sweden, June 1994.

[11] M. Kristensson, B.E. Ottersten, and D.T.M. Slock, "Blind Subspace Identification of a BPSK Communication Channel," in *Proc. of the 30th Asilomar Conference on Signals, Systems & Computers*, Pacific Grove, CA, Nov. 1996.

[12] A.J. van der Veen, "Analytical Method for Blind Binary Signal Separation," *IEEE Transactions on Signal Processing*, vol. 45, no. 4, pp. 1078–1082, April 1997.

[13] D.T.M. Slock, "Blind Joint Equalization of Multiple Synchronous Mobile Users Using Oversampling and/or Multiple Antennas," in *Proc. 28th Asilomar Conference on Signal, Systems & Computers*, Pacific Grove, CA, Nov. 1994.

[14] G. Harikumar and Y. Bresler, "FIR Perfect Reconstruction from Multiple Convolutions: Minimum Deconvolver Orders," *IEEE Transactions on Signal Processing*, vol. 46, no. 1, pp. 215–218, Jan. 1998.

[15] G.D. Forney, "Minimal Bases of Rational Vector Spaces with Applications to Multivariable Linear Systems," *SIAM Journal on Control*, vol. 13, no. 3, pp. 493–520, May 1975.

[16] Luc Deneire, Jaouhar Ayadi, and Dirk T.M. Slock, "Subspace Fitting Without Eigendecomposition," in $13^{th}$ *international Conference on Digital Signal Processing*, Santorini, Greece, July 1997.

[17] M. Gürelli and C.L. Nikias, "A New Eigenvector-Based Algorithm for Multichannel Blind Deconvolution of Input Colored Signals," in *Proc. ICASSP*, 1993, pp. 448–451.

[18] G. Xu, H. Liu, L. Tong, and T. Kailath, "A Least Squares Approach to Blind Channel Identification," *IEEE Transactions on Signal Processing*, vol. 43, no. 12, pp. 2982–2993, Dec. 1995.

[19] Y. Hua, "Fast Maximum Likelihood for Blind Identification of Multiple FIR Channels," *IEEE Transactions on Signal Processing*, vol. 44, no. 3, pp. 661–672, March 1996.

[20] J. Ayadi and D.T.M. Slock, "Cramér-Rao Bounds and Methods for Knowledge Based Estimation of Multiple FIR Channels," in *Proc. SPAWC 97 Conf.*, Paris, France, April 1997.

[21] H. Zeng and L. Tong, "Connections between the Least–Squares and Subspace Approaches to Blind Channel Estimation," *IEEE Transactions on Signal Processing*, vol. 44, no. 6, pp. 1593–1596, June 1996.

[22] D.T.M. Slock and C.B. Papadias, "Further Results on Blind Identification and Equalization of Multiple FIR Channels," in *Proc. ICASSP 95 Conf.*, Detroit, Michigan, May 1995.

[23] Elisabeth de Carvalho, Luc Deneire, and Dirk Slock, "Blind and Semi-Blind Maximum Likelihood Techniques for Multiuser Multichannel Identification," in *European*

*Association for Signal Processing EUSIPCO 98*, Island of Rhodes, Greece, September 1998.

[24] J. Ayadi and D.T.M. Slock, "Multichannel Estimation by Blind MMSE ZF Equalization," in *Signal Processing Advances in Wireless Communications (SPAWC)*, Annapolis, Maryland, USA, May 1999.

[25] M.K. Tsatsanis and Z. Xu, "Constrained Optimization Methods for Blind Equalization of Multiple FIR Channels," in *Proc. Asilomar Conference on Signals, Systems & Computers*, Pacific Grove, CA, Nov. 1998.

[26] L. Tong and Q. Zhao, "Blind Channel Estimation by Least Squares Smoothing," in *Proc. ICASSP 98 Conf.*, Seattle, USA, May 1998.

[27] J. Zhu, Z. Ding, and X.-R. Cao, "Column-Anchored Zeroforcing Blind Equalization for Multiuser Wireless FIR Channels," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 3, pp. 411, March 1999.

[28] David Gesbert, *Egalisation et Identification Multi–voies: Méthodes Auto–Adaptatives au Second–Ordre*, Ph.D. thesis, ENST, Paris, 1997.

[29] E. Pitié and P. Duhamel, "Bilinear Methods for Blind Channel Equalization: (No) Local Minimum Issue," in *Proc. ICASSP 98 Conf.*, Seattle, USA, May 1998.

[30] E. de Carvalho and D.T.M. Slock, "Maximum-Likelihood Blind Equalization of Multiple FIR Channels," in *Proc. ICASSP 96 Conf.*, Atlanta, USA, May 1996.

[31] P. Comon, "Analyse en Composantes Indépendantes et Identification Aveugle," *Traitement du Signal*, vol. 7, no. 3, pp. 435–450, Dec. 1990, Numéro spécial non linéaire et non gaussien.

[32] K. Abed Meraim, E. Moulines, and P. Loubaton, "Prediction Error Method for Second-Order Blind Identification," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 694–705, March 1997.

[33] Alexei Gorokhov, *Séparation autodidacte des mélanges convolutifs: méthodes du second ordre*, Ph.D. thesis, Ecole Nationale Supérieure des Télécommunications, 1997.

[34] Luc Deneire, *Estimation Aveugle de Canal et Accès Multiple par Répartition Spatiale*, Ph.D. thesis, ENST, Paris, 1998.

[35] Luc Deneire and Dirk Slock, "A Schur Method for Multiuser Multichannel Blind Identification," in *Proc. ICASSP 99 Conf.*, Phoenix, USA, March 1999.

[36] L. Deneire and D.T.M. Slock, "A Deterministic Schur Method for Multichannel Blind Identification," in *Signal Processing Advances in Wireless Communications (SPAWC)*, Annapolis, Maryland, USA, May 1999.

[37] G.B. Giannakis and S.D. Halford, "Asymptotically Optimal Blind Fractionally-Spaced Channel Estimation and Performance Analysis," *IEEE Transactions on Signal Processing*, vol. 45, no. 7, pp. 1815–1830, July 1997.

[38] H.H. Zeng and L. Tong, "Blind Channel Estimation Using the Second–Order Statistics: Asymptotic Performance and Limitations," *IEEE Transactions on Signal Processing*, vol. 45, no. 8, pp. 2060–2071, Aug. 1997.

[39] M. Kristensson, *On Parameter Estimation in Wireless Communications, Sensor Array Processing and Spectral Analysis*, Ph.D. thesis, Royal Institute of Technology (KTH), Sweden, 1998.

[40] E. de Carvalho and D.T.M. Slock, "Semi-Blind Maximum-Likelihood Estimation with Gaussian Prior for the Symbols using Soft Decisions," in *48th Annual Vehicular Technology Conference*, Ottawa, Canada, May 1998.

[41] E. de Carvalho and D.T.M. Slock, "Asymptotic Performance of ML Methods for Semi-Blind Channel Estimation," in *Proc. Asilomar Conference on Signals, Systems & Computers*, Pacific Grove, CA, Nov. 1997.

[42] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, "Exact and Large Sample ML Techniques for Parameter Estimation and Detection in Array Processing," in *Radar Array Processing*, S. Haykin, J. Litva, and T.J. Shepherd, Eds. Springer-Verlag, 1993.

[43] N. Seshadri, "Joint Data and Channel Estimation Using Blind Trellis Search Techniques," *IEEE Transactions on Communications*, vol. 42, pp. 1000–1011, Feb.-Apr. 1994.

[44] M.V. Eyuboglu and S.U. Qureshi, "Reduced–State Sequence Estimation with Set Partioning and Decision–Feedback," *IEEE Transactions on Communications*, vol. 36, pp. 13–20, Jan. 1988.

[45] S. Talwar, M. Viberg, and A. Paulraj, "Blind Separation of Synchronous Co-Channel Digital Signals Using an Antenna Array," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1184–1197, May 1996.

[46] P. Comon, O. Grellier, and B. Mourrain, "Blind Channel Identification with MSK Inputs," in *Asilomar Conference*, Pacific Grove, California, November 1-4 1998, Invited session.

[47] Dempster, N.M. Laird, and D.B. Rubin, "Maximum Likelihood from Incomplete Data via EM algorithm," *J. Royal Statisti. Soci.*, vol. 39, 1977.

[48] L. Tong and S. Perreau, "Multichannel Blind Identification: From Subspace to Maximum Likelihood Methods," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 1951–1968, Oct. 1998.

[49] H.A. Cirpan and M.K. Tsatsanis, "Stochatic Maximum–Likelihood Methods for Semi–Blind Channel Estimation," *IEEE Signal Processing Letters*, vol. 5, no. 1, pp. 21–24, Jan. 1998.

[50] B. Hochwald and A. Nehorai, "On Identifiability and Information-Regularity in Parameterized Normal Distributions," *Circuits, Systems ans Signal Processing*, vol. 16, no. 1, 1997.

[51] Y. Hua and M. Wax, "Strict Identifiability of Multiple FIR Channels Driven by an Unknown Arbitrary Sequence," *IEEE Transactions on Signal Processing*, vol. 44, no. 3, pp. 756–759, March 1996.

[52] E. de Carvalho and D.T.M. Slock, "Identifiability Conditions for Blind and Semi-Blind Multichannel Estimation," in *European Association for Signal Processing EUSIPCO 98*, Island of Rhodes, Greece, September 1998.

[53] L. Tong, G. Xu, and T. Kailath, "Blind Identification and Equalization Based on Second-Order Statistics: A Time Domain Approach," *IEEE Transactions on Information Theory*, vol. 40, no. 2, pp. 340–349, March 1994.

[54] Z. Ding and Y. Li, "On Channel Identification Based on Second-Order Cyclic Spectra," *IEEE Transactions on Signal Processing*, vol. 42, no. 5, pp. 1260–1264, May 1994.

[55] Luc Deneire, Elisabeth de Carvalho, and Dirk Slock, "Identifiability Conditions for Blind and Semi-Blind Multiuser Multichannel Identification," in *9th IEEE Signal Processing Workshop On Statistical Signal And Array Processing*, Portland, Oregon, USA, September 1998.

[56] E. Moulines, P. Duhamel, J.F. Cardoso, and S. Mayrargue, "Subspace Methods for the Blind Identification of Multichannel FIR filters," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 516–526, Feb. 1995.

[57] J.D. Gorman and A.O. Hero, "Lower Bounds for Parametric Estimation with Constraints," *IEEE Transactions on Information Theory*, vol. 26, no. 6, pp. 1285–1301, Nov. 1990.

[58] S.M. Kay, *Fundamentals of Statistical Signal Processing Estimation Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1993.

[59] Peter E. Caines, *Linear Stochastic Systems*, John Wiley & Sons, 1988.

[60] E. de Carvalho and D.T.M. Slock, "Blind and Semi–Blind FIR Multichannel Estimation: Cramér–Rao Bounds," Submitted to IEEE Transactions on Signal Processing.

[61] T. Marzetta, "On Simple Derivation of the Constrained Multiple Parameter Cramér–Rao Bound," *IEEE Transactions on Signal Processing*, vol. 41, no. 6, pp. 2247–2249, June 1993.

[62] P. Stoica and B.C. Ng, "On the Cramér–Rao Bound Under Parametric Constraints," *IEEE Signal Processing Letters*, vol. 5, no. 7, pp. 177–179, July 1998.

[63] E. de Carvalho and D.T.M. Slock, "Cramér-Rao Bounds for Semi-blind, Blind and Training Sequence based Channel Estimation," in *Proc. SPAWC 97 Conf.*, Paris, France, April 1997.

[64] J.L. Bapat, *Partially Blind Identification of FIR Channels for QAM Signals*, Ph.D. thesis, Pennsylvania State University, Aug. 1996.

[65] D.R. Morgan, J. Benesty, and M.M. Sondhi, "On the Evaluation of Estimated Impulse Responses," *IEEE Signal Processing Letters*, vol. 5, no. 7, pp. 174–176, July 1998.

[66] L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, Englewood Cliffs, NJ, 1987.

[67] E. de Carvalho and D.T.M. Slock, "Blind and Semi–Blind FIR Multichannel Estimation: Identifiability Conditions," Submitted to IEEE Transactions on Signal Processing.

[68] M. Kristensson, M. Jansson, and B. Ottersten, "Modified IQML and a Statistically Efficient Method for Direction Estimation without Eigendecomposition," in *Proc. ICASSP*, Seattle, USA, May 1998.

[69] J. Ayadi, E. de Carvalho, and D.T.M. Slock, "Blind and Semi-Blind Maximum Likelihood Methods for FIR Multichannel Identification," in *Proc. ICASSP 98 Conf.*, Seattle, USA, May 1998.

[70] N. Chotikakamthorn and J.A. Chambers, "On the IQML Algorithm for Multiple Signal Parameter Estimation," *IEE Proceedings-Radar, Sonar and Navigation*, vol. 5, no. 144, pp. 237–244, Oct. 1997.

[71] M.R. Osborne and G.K. Smyth, "A Modified Prony Algorithm for Fitting Functions Defined by Difference Equations," *SIAM J. Sci. Stat. Comput.*, vol. 12, no. 2, pp. 362–382, 1991.

[72] G. Harikumar and Y. Bresler, "Analysis and Comparative Evaluation of Techniques for Multichannel Blind Deconvolution," in *Proc. 8th IEEE Sig. Proc. Workshop Statistical Signal and Array Proc.*, Corfu, Greece, June 1996, pp. 332–335.

[73] A. Gorokhov and P. Stoica, "Generalized Quadratic Minimization and Signal Processing Applications," in *Proc. EUSIPCO 98 Conf.*, Island of Rhodes, Greece, Sept. 1998.

[74] S. Talwar, *Blind Space-Time Algorithms for Wireless Communication System*, Ph.D. thesis, Stanford University, Jan. 1996.

[75] Elisabeth de Carvalho and Dirk Slock, "Asymptotic Performance of ML Methods for Blind and Semi-Blind Channel Estimation," In preparation, for IEEE Transactions on Signal Processing.

[76] A.-J. Van der Veen, "Algebraic Methods for Deterministic Blind Beamforming," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 1951–1968, Oct. 1998.

[77] E. de Carvalho and D.T.M. Slock, "Maximum-Likelihood FIR Multi-Channel Estimation with Gaussian Prior for the Symbols," in *Proc. ICASSP 97 Conf.*, Munich, Germany, April 1997.

[78] Elisabeth de Carvalho and Dirk Slock, "Semi–Blind Maximum Likelihood Methods for FIR Multichannel Identification," In preparation, for IEEE Transactions on Signal Processing.

[79] European Telecommunications Standards Institute, "European Digital Cellular Telecommunications system (phase 2): Radio Transmission and Reception (GSM 05.05)," Tech. Rep., ETSI, Dec. 1995, Sophia Antipolis, France.

[80] A. Gorokhov and P. Loubaton, "Semi–Blind Second–Order Identification of Convolutive Channels," in *Proc. ICASSP 97 Conf.*, Munich, Germany, April 1997.

[81] A.M. Kuzminskiy and D. Hatzinakos, "Semiblind Estimation of Spatio–Temporal Filter Coefficients Based on a Training-Like Approach," *IEEE Signal Processing Letters*, vol. 5, no. 9, pp. 231–233, Sept. 1998.

[82] A.M. Kuzminskiy, L. Féty, P. Forster, and S. Mayrargue, "Regularized Semi–Blind Estimation of Spatio–Temporal Filter Coefficients for Mobile Radio Communications," in *Proc. GRETSI 97 Conf.*, Grenoble, France, 1997.

[83] K. Abed Meraim, J.-F. Cardoso, A.Y. Gorokhov, P. Loubaton, and E. Moulines, "On Subspace Methods for Blind Identification of Single–Input Multiple–Output FIR Systems," *IEEE Transactions on Signal Processing*, vol. 45, no. 1, pp. 42–55, Jan. 1997.

[84] J.L. Bapat, "Partially blind estimation: ML-based approaches and Cramér–Rao bound," *Signal Processing*, vol. 71, no. 3, pp. 265–277, Dec. 1998.

[85] E. de Carvalho and D.T.M. Slock, "A Fast Gaussian Maximum-Likelihood Method for Blind Multichannel Estimation," in *Signal Processing Advances in Wireless Communications (SPAWC)*, Annapolis, Maryland, USA, May 1999.

[86] P. Whittle, "The Analysis of Multiple Stationary Time Series," *J. Royal Statistic. Soc.*, vol. 45, pp. 125–139, 1953.

[87] D.T.M. Slock and E. de Carvalho, "Unbiased MMSE Decision-Feedback Equalization for Packet Transmission," in *Proc. EUSIPCO 96 Conf.*, Trieste, Italy, September 1996.

[88] E. de Carvalho and D.T.M. Slock, "Burst Mode Equalization: Optimal Approach and Suboptimal Continuous–Processing Approximation," Submitted, Signal Processing, Special Issue on Signal Processing Technologies for Short–Burst Wireless Communications.

[89] J.G. Proakis, "Adaptive Non Linear Filtering Techniques for Data Transmission," *IEEE Symp. on Adaptive Processes*, pp. XV.2.1–5, 1970.

[90] A. Gersho and T.L. Lim, "Adaptive Cancellation of Intersymbol Interference for Data Transmission," *Bell Syst. Tech. J.*, vol. 60, no. 11, pp. 1997–2021, Nov. 1981.

[91] D.T.M. Slock and E. de Carvalho, "Burst Mode Non-Causal Decision-Feedback Equalization and Blind MLSE," in *Proc. GLOBECOM 96 Conf.*, London, Great Britain, November 1996.

[92] E. de Carvalho and D.T.M. Slock, "Burst Mode Non-Causal Decision-Feedback Equalizer based on Soft Decisions," in *48th Annual Vehicular Technology Conference*, Ottawa, Canada, May 1998.

[93] J.M. Cioffi, G.P. Dudevoir, M.V. Eyuboglu, and G.D. Forney Jr., "MMSE Decision-Feedback Equalizers and Coding - Part I: Equalization Results," *IEEE Trans. Com.*, vol. 43, no. 10, pp. 2582–2594, Oct. 1995.

[94] G. Kawas Kaleh, "Channel Equalization for Block Transmission Systems," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 1, pp. 110–121, Jan. 1995.

[95] J. M. Cioffi and G. D. Forney Jr., "Generalized Decision-Feedback Equalization for Packet Transmission with ISI and Gaussian noise," in *Communications, Computation, Control, and Signal Processing, A Tribute to Thomas Kailath*, pp. 79–127. Norwell, MA: Kluwer, 1997.

[96] J. M. Cioffi and G. D. Forney Jr., "Canonical Packet Transmission on the ISI Channel with Gaussian Noise," in *Proc. GLOBECOM 96 Conf.*, London, Great Britain, November 1996.

[97] N. Al-Dhahir, "Time–Varying Versus Time–Invariant finite–Length MMSE–DFE on Stationary Dispersive Channels," *IEEE Transactions on Communications*, vol. 46, no. 1, pp. 11–15, Jan. 1998.

[98] H. Sugimoto, L.K. Rasmussen, T.J. Lim, and T. Oyama, "Mapping Functions for Successive Interference Cancellation in CDMA," in *48th Annual Vehicular Technology Conference*, Ottawa, Canada, May 1998.

[99] T. Oyama, S. Sun, H. Sugimoto, T.J. Lim, L.K. Rasmussen, and Y. Matsumoto, "Performance Comparison of Multi–Stage SIC and Limited Tree–Search Detection in CDMA," in *48th Annual Vehicular Technology Conference*, Ottawa, Canada, May 1998.

[100] M. Chiani, "Introducing Erasures in Decision-Feedback Equalization to Reduce Error Propagation," *IEEE Transactions on Communications*, vol. 45, no. 7, pp. 757–760, Jul. 1997.

[101] A. Fertner, "Improvement of Bit–Error–Rate in Decision Feedback Equalizer by Preventing Decision–Error Propagation," *IEEE Transactions on Communications*, vol. 46, no. 7, pp. 1872–1877, Jul. 1998.

[102] E. Dahlman and B. Gudmundson, "Performance Improvement in Decision Feedback Equalizers by using Soft Decisions," *Electonic Letters*, vol. 24, pp. 1084–1085, Aug. 1988.

[103] J.W.M. Bergmans, J.O. Voorman, and H.W. Wong-Lam, "Dual Decision Feedback Equalizer," *IEEE Transactions on Communications*, vol. 45, no. 5, pp. 514–518, May 1997.

[104] A. Duel-Hallen, J. Holtzman, and Z. Zvonar, "Multiuser Detection for CDMA Systems," *IEEE Personal Communications*, pp. 46–58, April 1995.

[105] S. Moshavi, "Multi-User Detection for DS-CDMA Communications," *IEEE Communications Magazine*, pp. 124–136, Oct. 1996.

[106] D. Divsalar, M.K. Simon, and D. Raphaeli, "Improved Parallel Interference Cancellation for CDMA," *IEEE Transactions on Communications*, vol. 46, no. 2, pp. 258–268, Feb. 1998.

[107] A.L.C. Hui and K. Ben Letaief, "Successive Interference Cancellation for Multiuser Asynchronous DS/CDMA Detectors in Multipath Fading Links," *IEEE Transactions on Communications*, vol. 46, no. 3, pp. 384–391, March 1998.

[108] R. Nilsson, F. Söberg, O. Edfors, P. Ödling, H. Eriksson, S.K. Wilson, and P.O. Börjesson, "A Low Complexity Threshold Detector Making Decisions in a Multiuser Environment," in *48th Annual Vehicular Technology Conference*, Ottawa, Canada, May 1998.

[109] D.T.M. Slock and E. de Carvalho, "Matched Filter Bounds for Reduced-Order Multichannel Models ," in *Proc. GLOBECOM 96 Conf.*, London, Great Britain, November 1996.

[110] H. Trigui and D.T.M. Slock, "Cochannel Interference Cancellation within the Current GSM Standard," in *Proc. of the IEEE 1998 International Conference on Universal Personal Communications*, Florence, Italy, Oct. 5-9 1998.