Chapter 7

SEMI–BLIND METHODS FOR FIR MULTICHANNEL ESTIMATION

Elisabeth de Carvalho
Stanford University
David Packard Electrical Engineering
Star Lab
350 Serra Mall
Stanford, CA 94305-9515, USA
Tel: +1 650-724-3633 Fax: +1 650-723-9251
email: carvalho@dsl.stanford.edu

Dirk Slock
Institut EURECOM
2229 route des Crêtes, B.P. 193,
06904 Sophia Antipolis Cedex, FRANCE
Tel: +33 4 93 00 26 26 Fax: +33 4 93 00 26 27
email: slock@eurecom.fr

Corresponding Author: Elisabeth de Carvalho

Abstract

The purpose of semi-blind equalization is to exploit the information used by blind methods as well as the information coming from known symbols. Semi–blind techniques robustify the blind and training problems, and offer better performance than these two methods. We first present identifiability conditions and performance bounds for semi–blind 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 (CRBs) and a comparison of semi–blind techniques with blind and training sequence based techniques is presented. Three categories of semi–blind methods are considered. Optimal semi–blind methods take into account all blind information and all the information coming from the known symbols, even if not grouped. Mainly Maximum–Likelihood (ML) methods will be considered. Suboptimal semi–blind solutions are also investigated when the known symbols are grouped in a training sequence: the suboptimal semi–blind criteria appear as a linear combination of a training sequence criterion and a blind ML criterion. Thirdly, we present methods that combine a given blind criterion with a training sequence based criterion.

7.1 Introduction

7.1.1 Training Sequence based Methods and Blind Methods

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 actual mobile communication standards include a training sequence to estimate the channel, like in GSM [1]. In most cases, training methods appear as robust methods but present some disadvantages. Firstly, bandwidth efficiency decreases as a non–negligible part of the data burst may 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, such as when explicit synchronization between the receiver and the transmitter is not possible.

Blind equalization techniques allow the estimation of the channel or the equalizer based only on the received signal without any training symbols. The introduction of multichannels, or SIMO models where a single input symbol stream is transmitted through multiple symbol rate linear channels, has given rise to a plethora of new blind estimation techniques that do not require higher order–statistics. The most popular second–order statistics (SOS) based estimation techniques suffer from a lack of robustness: channels must satisfy diversity conditions and many blind SOS methods fail when the channel length is overestimated. Furthermore, the blind techniques leave an indeterminacy in the channel or the symbols, a scale or constant phase or a discrete phase factor. This suggests that SOS blind techniques should not be used alone but with some form of additional information. However, the same

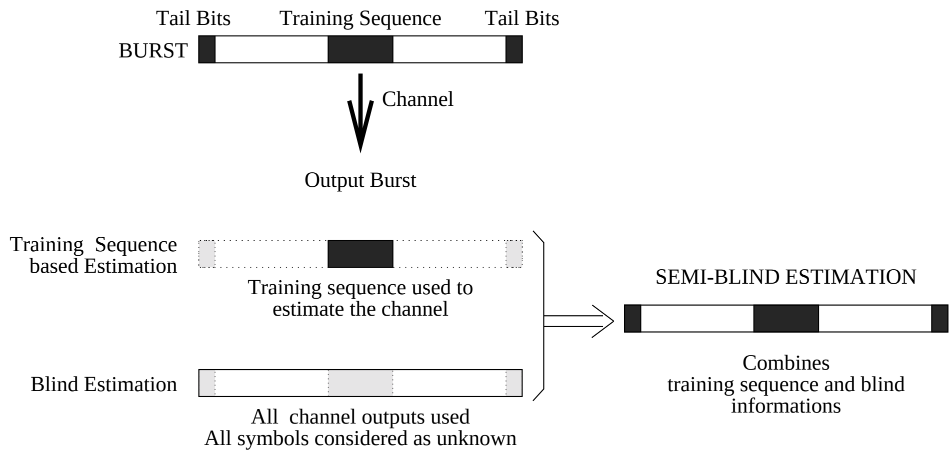

Figure 7.1. Semi–Blind Principle: example of a GSM burst.

may also be true for TS based methods, especially when the sequence is too short for a certain channel length. Semi–blind techniques provide a solution to overcome these problems.

7.1.2 Semi–Blind Principle

In this chapter, we assume a transmission by burst, *i.e.* the data is organized and transmitted by burst, and we furthermore assume that known symbols are present in each burst in the form of a training sequence aimed at estimating the channel or simply in the form of some known symbols used for synchronization or as guard intervals, like in the GSM or the DECT bursts. In this typical 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 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 7.1) and exploit the positive aspects of both techniques as stated in section 7.1.1.

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 the estimation of longer channel impulse responses than is possible with a certain training sequence length, a feature that is of interest for the application of mobile communications in mountainous areas. For methods based on the second–order moments of the data (which we will call Gaussian methods), one known symbol is sufficient to make any channel identifiable. In addition, it allows the use of 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 interesting 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.

In section 7.2, we describe the multichannel model. In section 7.4, we give identifiability conditions for FIR multichannel estimation for the deterministic and Gaussian classes of methods. In section 7.5, the CRBs are used as a performance measure for semi–blind estimation, which is compared to TS and blind channel estimation. Semi-blind estimation is shown to be superior to either TS or blind methods. In section 7.7, we describe optimal semi–blind methods which are able to take into account all blind information and all the information coming from the known symbols, even if not grouped. These methods are mainly based on ML. In section 7.9, we present suboptimal deterministic semi–blind methods which can be constructed when the known symbols are grouped in a training sequence. The suboptimal semi–blind criteria appear as a linear combination of a TS based criterion and a blind ML criterion. The proposed ML based methods are solved in an iterative quadratic fashion. In section 7.10, we linearly combine a blind criterion with a TS criterion: particularly the Subchannel Response Matching (SRM) and Subspace Fitting (SF) blind criteria are considered. We provide a performance study of the resulting semi–blind deterministic quadratic criteria in section 7.11. At last, in section 7.12, we give an example of the Gaussian method: semi–blind covariance matching.

Throughout this chapter, we shall use the following notation:

$(.)^*$, $(.)^T$, $(.)^H$	conjugate, transpose, conjugate transpose
$(.)^+$	Moore–Penrose Pseudo–Inverse
$\mathrm{tr}(A)$, $\det(A)$	trace and determinant of matrix A
$\mathrm{vec}(A)$	$[A_{i,1}^T \;\; A_{i,2}^T \cdots A_{i,n}^T]^T$
$\otimes$	Kronecker product
$\hat{\theta}$, θ^o	estimate of parameter θ, true value of parameter θ
E_X	mathematical Expectation w.r.t. the random quantity X
$\mathrm{Re}(.)$, $\mathrm{Im}(.)$	real and imaginary part
I	Identity matrix with adequate dimension
w.r.t.	with respect to

7.2 Problem Formulation

We consider here linear modulation (nonlinear modulations such as GMSK can be linearized with good approximation [2], [3]) over a linear channel with additive noise. The received signal after a linear receiver filter is then the convolution of

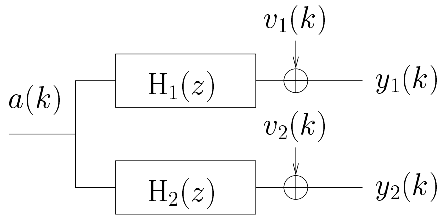

Figure 7.2. Multichannel model: example with 2 subchannels.

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, plus additive noise. 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, the single-user case will be considered.

We describe here the FIR multichannel model used throughout the chapter. This multichannel model applies to different cases: oversampling w.r.t. the symbol rate of a single received signal [4], [5], [6] or the separation into the real (in–phase) and imaginary (quadrature) component of the demodulated received signal if the symbol constellation is real [7], [8]. In the context of mobile digital communications, a third possibility appears in the form of multiple received signals from an array of sensors. These three sources for multiple channels can also be combined.

In the multichannel model, a sequence of symbols $a(k)$ is received through m channels of length N and with coefficients $\boldsymbol{h}(i)$ (see figure 7.2):

$$\boldsymbol{y}(k) = \sum_{i=0}^{N-1} \boldsymbol{h}(i) a(k-i) + \boldsymbol{v}(k), \tag{7.2.1}$$

$\boldsymbol{v}(k)$ is an additive independent white Gaussian noise, $r_{\boldsymbol{vv}}(k-i) = \mathrm{E}\, \boldsymbol{v}(k)\boldsymbol{v}(i)^H = \sigma_v^2 I_m\, \delta_{ki}$. Assume we receive M samples, concatenated in the vector $\boldsymbol{Y}_M(k)$:

$$\boldsymbol{Y}_M(k) = \mathcal{T}_M(h)\, A_{M+N-1}(k) + \boldsymbol{V}_M(k) \tag{7.2.2}$$

$\boldsymbol{Y}_M(k) = [\boldsymbol{y}^T(k) \cdots \boldsymbol{y}^T(k-M+1)]^T$, similarly for $\boldsymbol{V}_M(k)$, and the input burst is $A_{M+N-1}(k) = [a(k) \cdots a(k-M-N+2)]^T$. $\mathcal{T}_M(h)$ is a block Toeplitz matrix with M block rows and $\left[\boldsymbol{H} \;\; 0_{m\times(M-1)}\right]$ as first block row:

$$\boldsymbol{H} = [\boldsymbol{h}(0) \;\cdots\; \boldsymbol{h}(N-1)] \text{ and } h = \left[\boldsymbol{h}^T(0) \cdots \boldsymbol{h}^T(N-1)\right]^T. \tag{7.2.3}$$

The channel transfer function is $\mathbf{H}(z) = \sum_{i=0}^{N-1} \boldsymbol{h}(i) z^{-i} = [\mathrm{H}_1(z) \cdots \mathrm{H}_m(z)]^T$, where $\mathrm{H}_i(z)$ is the transfer function of the i^{th} subchannel.

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 (7.2.2) with $k = M-1$ to

$$\boldsymbol{Y} = \mathcal{T}(h)A + \boldsymbol{V}. \tag{7.2.4}$$

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 can 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 the corresponding parts in the channel output as $\mathcal{T}(h)A = \mathcal{T}_K(h)A_K + \mathcal{T}_U(h)A_U$.

Irreducible, Reducible, and 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:

$$\mathbf{H}(z) = \mathbf{H}_I(z)\mathrm{H}_c(z), \tag{7.2.5}$$

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.

Commutativity of Convolution We shall exploit the commutativity property of convolution:

$$\mathcal{T}(h)A = \mathcal{A}h \tag{7.2.6}$$

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

$$\mathcal{A}_1 = \begin{bmatrix} a(M-1) & a(M-2) & \cdots & a(M-N) \\ a(M-2) & \iddots & \iddots & \vdots \\ \vdots & \iddots & \iddots & \vdots \\ a(0) & \cdots & \cdots & a(-N+1) \end{bmatrix}. \tag{7.2.7}$$

Minimum Zero-Forcing (ZF) Equalizer Length, Effective Number of Channels The Bezout identity states that for an FIR irreducible channel, FIR ZF equalizers exist [9]. The minimum length for such a FIR ZF equalizer is

$$\underline{M} = \min\{M : \mathcal{T}_M(h) \text{ has full column rank}\}. \tag{7.2.8}$$

One may note that $\mathcal{T}_M(h)$ has full column rank for $M \geq \underline{M}$. In [10], it is shown that if the mN elements of $\boldsymbol{H}$ are considered random, more precisely independently distributed with a continuous distribution, then

$$\underline{M} = \left\lceil \frac{N-1}{m-1} \right\rceil \quad \text{with probability 1}, \tag{7.2.9}$$

and $\underline{M} = 1$ for $N = 1$. In this case, the channel is irreducible w.p. 1. One could consider other (perhaps more realistic) channel models. Consider, for example, 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 $m_e = \text{rank}(\boldsymbol{H})$ which in this case would equal (w.p. 1)

$$m_e \; = \; \min\{m, N, K\} \; . \tag{7.2.10}$$

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 (7.2.9), 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 $\underline{M}$ [9] could then be applied.

In general, for an irreducible channel, $\underline{M} \leq N-1$ [11] in which the upper bound would correspond to $m_e = 2$. Note that $m_e = 1$ corresponds to a reducible channel (in which case $\underline{M} = \infty$).

7.3 Classification of Semi–Blind Methods

Semi–blind methods can be classified (roughly) according to the amount of a priori knowledge on the unknown input symbols that gets exploited, see figure 7.3:

1. No information exploited: deterministic methods.
   These methods are directly based on the structure of the received signal and particularly on the structure of the convolution matrix $\mathcal{T}(h)$. Among the blind deterministic methods, one can find the Subspace Fitting (SF) method [12], the Subchannel Response Matching [13] method, the deterministic Maximum Likelihood method [14], [5] and also the least squares smoothing method or two–sided linear prediction approach [15], [16]. Semi–blind extensions of these blind methods already exist as stated later in this chapter.

2. Second–order statistics: Gaussian methods.
   These methods exploit the second–order moments of the data. The (blind) prediction method [5], [17] or the (blind) covariance matching method [18] belong to this category, but also the semi–blind Gaussian Maximum Likelihood (GML) approach [19] which treats the unknown input symbols as Gaussian random variables. A semi–blind covariance matching approach will be presented later in this chapter.

3. Higher–order statistics.
   These methods exploits second– and higher–order statistics of the data [20].

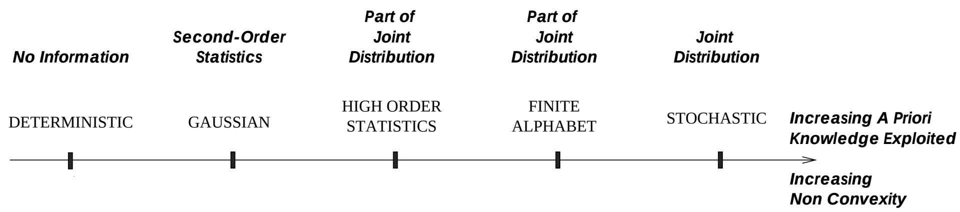

Figure 7.3. Classification of (semi)blind channel identification methods according to the assumed A Priori knowledge on the unknown input symbols.

4. Finite Alphabet of the input symbols.
   These methods exploit the finite alphabet nature of the input symbols. Among these approaches, we find the ML methods [21].

5. Complete distribution of the input symbols.
   In these methods, the true distribution of the input symbols is exploited; for example, for BPSK, we exploit the true discrete distribution the input symbols (*i.e.* their finite alphabet plus their probabilities). The stochastic ML (SML), which was presented in its semi–blind version in [22], belongs to this category.

The more information on the unknown input symbols gets exploited, the better the channel estimation. However, at the same time, the associated methods are typically more costly, with cost functions presenting local minima. In this chapter, we mainly focus on deterministic and Gaussian methods which can be solved in a simple way with sometimes quadratic cost functions/closed-form solutions.

7.4 Identifiability Conditions for Semi–Blind Channel Estimation

In this section, we define identifiability conditions for semi–blind deterministic and Gaussian channel estimation. We first recall results on blind estimation and then extend them to semi–blind estimation. One remarkable property is that, in the deterministic case, semi–blind techniques can estimate the channel even when the known symbols are arbitrarily dispersed. The following results are detailed and proved in [23].

7.4.1 Identifiability Definition

Let θ be the parameter to be estimated and $\boldsymbol{Y}$ the observations. We denote by $f(\boldsymbol{Y}|\theta)$ the probability density function of $\boldsymbol{Y}$. In a regular case (*i.e.* in a non blind case), θ is called identifiable if [24]:

$$\forall\, \boldsymbol{Y}, \quad f(\boldsymbol{Y}|\theta) = f(\boldsymbol{Y}|\theta') \quad \Rightarrow \quad \theta = \theta'. \tag{7.4.1}$$

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 α: $\alpha \in \mathbb{C}$

in the deterministic model and $|\alpha| = 1$ in the Gaussian model (α is real in the case of real symbols). The identifiability condition (7.4.1) will be for θ to equal θ' 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 $\boldsymbol{Y}$.

7.4.2 TS Based Channel Identifiability

We recall here the identifiability conditions for TS based channel estimation. From (7.2.6), $\mathcal{T}(h)A = \mathcal{A}h$. 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 $\mathbf{H}(z)$ is identifiable by TS estimation if and only if*

(i) Burst Length $M \geq N$ or number of known symbols $M_K \geq 2N-1$.

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

The burst length M is the length of $\boldsymbol{Y}$, expressed in symbol periods.

7.4.3 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 θ is based on the mean only; the covariance matrix only contains information about σ_v^2, the estimation of which hence gets decoupled from the estimation of θ. A_U and h are identifiable if:

$$\mathcal{T}(h)A = \mathcal{T}(h')A' \Rightarrow$$
$$\begin{cases} A_U = A'_U \text{ and } h = h' & \text{for semi–blind and TS based estimation} \\ A = \dfrac{1}{\alpha}A' \text{ and } h = \alpha h' & \text{for blind estimation} \end{cases} \tag{7.4.2}$$

with α complex, for a complex input constellation, and real, for a real input constellation. Identifiability is hence defined from the noise–free data $\mathcal{T}(h)A$ (the mean).

Blind Channel Identifiability

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 $M \geq N + 2\underline{M}$.

[1]For a definition of the notion of modes, see for example [14], [25]

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

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 (*i.e.* the column space of $\mathcal{T}(h)$ or its orthogonal complement). These conditions appear to be sufficient for all the deterministic methods listed in section 7.3 except for SRM [14].

Semi–Blind Channel Identifiability

We distinguish here between the case of grouped known symbols (training sequence case) and the case of arbitrarily dispersed known symbols. In both cases, the channel can be identified for the same number of (non-zero) known symbols. When the known symbols are all equal to zero, the channel cannot be identified when they are grouped. However, when they are dispersed in the burst, the channel can be identified (up to a scale factor). We will consider the general case of a reducible channel $\mathbf{H}(z) = \mathbf{H}_I(z)\mathrm{H}_c(z)$: $\underline{M}_I$ is defined as $\underline{M}_I = \min\{M : \mathcal{T}_M(h_I) \text{ has full column rank}\}$.

Grouped known Symbols

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* $\mathbf{H}(z)$ *(or hence of* $\mathrm{H}_c(z)$*).*

(iii) Grouped known symbols: number $M_K \geq 2N_c{-}1$ *(which is also a necessary condition), with number of excitation modes* $\geq N_c$.

For an irreducible channel, 1 known symbol is sufficient. For a monochannel, $2N{-}1$ grouped known symbols are sufficient. If $2N{-}1$ grouped known symbols containing N independent modes are available, condition (*ii*) becomes superfluous. For a reducible channel $\mathbf{H}(z) = \mathbf{H}_I(z)\mathrm{H}_c(z)$, $\mathbf{H}_I(z)$ can be identified blindly while $\mathrm{H}_c(z)$ can be identified by TS. In general, at least as many known symbols are needed as the number of (continuous) parameters that cannot be determined by blind estimation.

Arbitrarily Dispersed Known Symbols In [26], we prove that the regularity of the Fisher Information Matrix (FIM) is equivalent to local identifiability for the deterministic model (and also the Gaussian model). By studying the regularity of the FIM, the treatment of the arbitrarily dispersed known symbols case (also treated in [27] to some extent) becomes tractable. Strictly speaking, we can prove only local identifiability. In general, FIM regularity implies a finite number of solutions in A and h [28], however, for a sufficient burst length [27] (larger than the ones given below), the solution becomes unique and we have global identifiability.

- Non-Zero Known Symbols

  Theorem 1: *The channel* $\mathbf{H}(z)$ *is locally identifiable with probability 1*[2] *if*

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

  Note that no further conditions on the excitation modes of the known symbols are required. Furthermore, a monochannel can be identified if $2N - 1$ arbitrarily dispersed known symbols are available.

- Known Symbols equal to Zero

  When the known symbols are all equal to zero, 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 $A_K = A'_K$. We have shown in [26] that the channel is semi–blindly locally identifiable up to a scale factor if and only if the FIM is 1-singular. The position of the known symbols cannot be totally arbitrary though. In fact, we have the following theorem:

  Theorem 2: *The channel* $\mathbf{H}(z)$ *is locally identifiable with probability 1 up to a scale factor if*

  (i) *Burst length* $\geq N_I + 2\underline{M}_I$.
  (ii) *Number of excitation modes* $\geq N + \underline{M}_I$.
  (iii) *Number of known symbols (zeros)* $\geq 2N_c - 2$, *which is also a* **necessary** *condition.*
  (iv) *The known symbols (zeros) are "sufficiently" dispersed: there are at least* $N_c - 1$ *known symbols that do not belong to a group of* N_c *or more known symbols.*

  If only some known symbols are equal to zero, Theorem 1 can be applied provided condition (iv) of Theorem 2 is added.

Semi–Blind Robustness to Channel Length Overestimation

A major disadvantage of the deterministic blind methods is their non robustness to channel length overestimation. Semi–blind estimation overcomes this problem. 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*

[2] refers to the distribution of the channel coefficients

(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 $\geq N'{-}N_I{+}1$.

These results are also valid with probability one for arbitrarily dispersed known symbols.

7.4.4 Identifiability in the Gaussian Model

In the Gaussian model, $\boldsymbol{Y} \sim \mathcal{N}(\mathcal{T}_K(h)A_K, \sigma_a^2 \mathcal{T}_U(h)\mathcal{T}_U^H(h) + \sigma_v^2 I)$ and $\theta = [h^T\ \sigma_v^2]^T$. σ_a^2 is the variance of the input symbols. 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.

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 to be identifiable if:

$$C_{YY}(h, \sigma_v^2) = C_{YY}(h', \sigma_v^{2\prime}) \Rightarrow h' = e^{j\varphi}h \text{ and } \sigma_v^{2\prime} = \sigma_v^2. \tag{7.4.3}$$

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

Unlike in the deterministic case, zeros can be identified: it is only not possible to determine if the zeros are minimum or maximum–phase (discrete valued ambiguity). So if it is known that the channel is minimum-phase, the channel can be identified.

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 $M \geq \underline{M} + 1$

These conditions are sufficient conditions for the prediction, covariance matching and GML methods. Note that not all the non–zero correlations (lag 0 to $N-1$) are necessary for identification but only the first $\underline{M}+1$.

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* $\mathbf{H}(z)$ *is identifiable blindly up to a phase factor if*

(i) $H_c(z)$ *is minimum–phase.*

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

In the monochannel case, the noise variance σ_v^2 cannot be estimated and hence neither h. However, if we consider σ_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.

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 has the advantage of allowing 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_U 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), $\mathcal{A}_K = a(k)I_{Nm}$. The Gaussian model appears thus more robust than the deterministic model as it allows identification of any channel, reducible or not, multi or monochannel, under the following conditions:

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.

These results on identifiability indicate the superiority of the semi–blind methods over TS and blind methods, as well as the superiority of Gaussian methods over the deterministic methods. This last fact is especially true in the multiuser case [29].

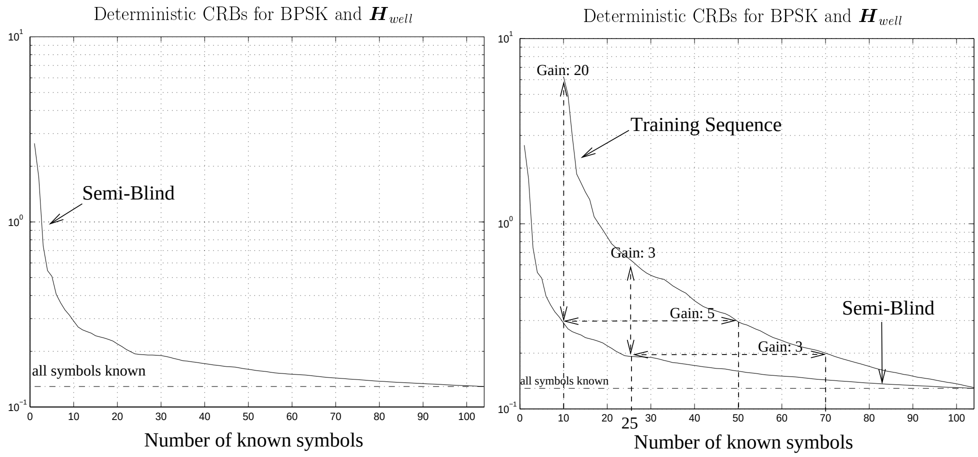

Figure 7.4. CRBs for deterministic semi–blind channel estimation (left); comparison between deterministic semi–blind and TS channel estimation (right).

7.5 Performance Measure: Cramér–Rao Bounds

We compare here the performance of semi–blind channel estimation to blind and training based estimation through the Cramér–Rao bounds for the channel with the input symbols being considered as nuisance parameters. 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. 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; the burst length is $M = 100$.

Three different types of channels are tested: an irreducible channel $\boldsymbol{H}_{well}$, an ill–conditioned channel with nearly a (common) zero $\boldsymbol{H}_{ill}$, a reducible channel with irreducible part $\boldsymbol{H}_I$ and reducible part H_c.

The results are mainly shown in the deterministic case; the Gaussian curves present similar shapes [26]. In figure 7.4 (left), we show the CRBs of semi–blind channel estimation for a fixed number of input symbols and variable number of known symbols. We see a dramatic improvement of the performance with very few known symbols.

In figure 7.4 (right), we compare the performance of semi–blind and training modes. For 10 known symbols, we have a gain of performance of 20 brought by semi–blind estimation. For the same estimation quality, 10 known symbols for semi–blind estimation require 50 known symbols for TS estimation. For 25 known symbols, we have a performance gain of a factor 3; one requires 70 known symbols in training mode to get the performance of semi–blind estimation.

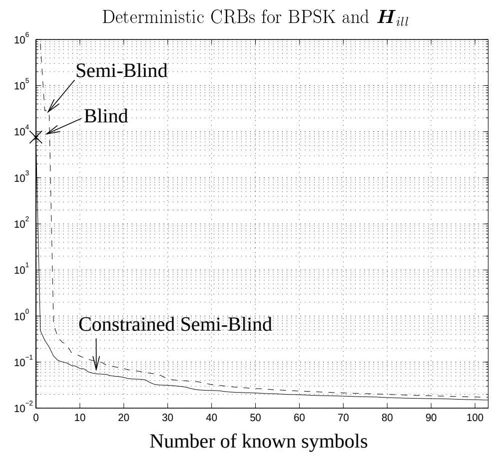

Figure 7.5. Comparison between deterministic blind and semi–blind channel estimation for $\boldsymbol{H}_{ill}$.

Deterministic blind channel estimation leaves a scale factor indeterminacy in the channel. One common way to adjust this scale factor is to impose constraints on the parameters; different constraints can be considered. Here, we consider the blind CRB computed under a norm constraint $||h||^2 = ||h^o||^2$ and a phase constraint $\text{Im}(h^H h^o) = 0$. This choice of constraints is first motivated by the common use of the norm constraint. Furthermore, the resulting constrained CRB is equal to the pseudo–inverse of the Fisher information matrix for the channel (with the input symbols considered as nuisance parameters) [30]. This is a particular constrained CRB as it yields the lowest value for the trace of the CRB among all sets of a minimal number of independent constraints. The previous norm and phase constraints give the same constrained CRB as the linear constraint $h^H h^o = h^{o\,H} h^o$: in general, a set of constraint is equivalent to another set of linear constraints, in the sense that they give the same constrained CRB, see [30] for further details.

In figure 7.5, we compare semi–blind and blind modes evaluated under the norm and phase constraints for an ill–conditioned channel. The superiority of the semi–blind mode can be noticed especially for a small number of known symbols.

In figure 7.6, the case of a reducible channel is shown. The CRB is drawn w.r.t. the number of unknown symbols in the burst, for a fixed number, 10, of known symbols. In the deterministic case, the blind part brings asymptotically no information to the estimation of the zeros of the channel. In the Gaussian case, however, the blind part brings information to the estimation of H_c.

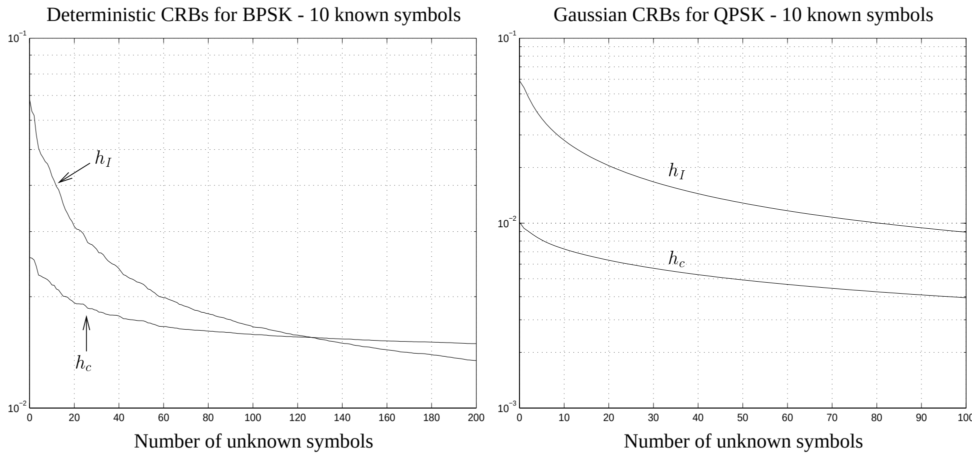

Figure 7.6. CRBs for deterministic (left) and Gaussian (right) semi–blind estimation of a reducible channel w.r.t. the number of unknown symbols. 10 symbols are known.

7.6 Performance Optimization Issues

Values of the known symbols (Deterministically) white input symbol sequences, in the sense that $\mathcal{A}^H\mathcal{A} = M\sigma_a^2 I$, optimize the performance of training sequence based estimation. Optimization of the semi–blind CRB w.r.t. the known symbols leads to channel dependent results; 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 tests 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) or equal to 0 (bottom), 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, performance is better when the known symbols are grouped than uniformly separated in the burst. For the random channel, both choices seem e-quivalent. 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 (such as *e.g.* COST 207 models [31]). Our simulation experience tends to indicate that the training sequence position should be such that $\mathcal{T}_K(h)$ has maximal energy. In other words, putting the training sequence in the middle of the burst is never a bad choice, regardless of the channel.

7.7 Optimal Semi–Blind Methods

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 in the burst.

- Semi–blind identifiability conditions should also be respected: for example, the methods should work for only one known symbol for irreducible channels, something that is not systematically satisfied by the suboptimal methods.

- With a sufficient number of known symbols, the optimal semi–blind methods should be able to identify any channel, particularly monochannels.

Optimal semi–blind methods are methods that naturally incorporate the knowledge of symbols. Maximum–Likelihood methods fulfill this condition. Methods estimating directly the input symbols like [32] are also good candidates for optimal semi–blind methods. We describe now the semi–blind ML methods.

Deterministic Maximum Likelihood (DML) In DML, both the channel h and the unknown symbols A_U are to be estimated. The DML criterion can be written as:

$$\min_{h,A_U} \|\boldsymbol{Y} - \mathcal{T}(h)A\|^2 = \min_{h,A_U} \|\boldsymbol{Y} - \mathcal{T}_K(h)A_K - \mathcal{T}_U(h)A_U\|^2 \tag{7.7.1}$$

This criterion can be optimized using alternating minimizations between h and A_U (see [33], [34] for the blind version of the algorithm and [35] for the semi–blind version). This algorithm offers the advantage of decreasing the cost function at each iteration. The blind version converges to the ML minimum asymptotically (in SNR and/or number of data) [34]. However, this algorithm requires a large number of iterations which renders it less practical.

When solving criterion (7.7.1) w.r.t. A_U and replacing the expression of A_U found in the criterion, we get the DML criterion for h:

$$\min_{h} \left(\boldsymbol{Y} - \mathcal{T}_K(h)A_K\right)^H P^{\perp}_{\mathcal{T}_U(h)} \left(\boldsymbol{Y} - \mathcal{T}_K(h)A_K\right) \tag{7.7.2}$$

P_X is the orthogonal projection onto the column space of X and $P^{\perp}_X = I - P_X$ is the projection onto its orthogonal complement. This criterion can be optimized by the method of scoring [36], which represents a computationally heavy solution. In section 7.9, the criterion (7.7.2) will be simplified in the case of grouped known symbols and low complexity solutions will be proposed.

Gaussian Maximum Likelihood (GML) In the Gaussian approach, the parameters to be estimated are the channel and the noise variance. The GML criterion has the form:

$$\min_{h,\sigma_v^2} \ \ln\det C_{YY} + \left(\boldsymbol{Y} - \mathcal{T}_K(h)A_K\right)^H C^{-1}_{YY} \left(\boldsymbol{Y} - \mathcal{T}_K(h)A_K\right) \tag{7.7.3}$$

Again, this criterion can be optimized by the method of scoring [36].

Maximum Likelihood with finite alphabet constraints on the input symbols (FA–ML) The FA–ML criterion is similar to the DML criterion except that the finite alphabet (denoted $\mathcal{A}_p$) constraint on the input symbols is imposed.

$$\min_{h,A_U \in \mathcal{A}_p} \|\boldsymbol{Y} - \mathcal{T}_K(h)A_K - \mathcal{T}_U(h)A_U\|^2 \tag{7.7.4}$$

FA–ML can be solved by alternating minimizations between h and A_U, with A_U constrained to the finite alphabet. The most problematic estimation is that of the symbols because of the FA constraint. In [21] where the blind version of (7.7.4) is optimized, the FA constraint is first ignored and then the estimates are projected onto the nearest discrete value of the finite alphabet. In [37], this technique is extended to the semi–blind case. In [38], the blind version of (7.7.4) is optimized w.r.t. A via the Viterbi algorithm.

Stochastic Maximum Likelihood (SML) 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. $f(\boldsymbol{Y}|h) = \sum_{A \in \mathcal{A}_p} f(\boldsymbol{Y}|A,h) f(A) \sim \sum_{A \in \mathcal{A}_p} f(\boldsymbol{Y}|A,h)$, so the SML criterion is:

$$\min_{h,\sigma_v^2} \frac{1}{\sigma_v^2} \sum_{A_U \in \mathcal{A}_p} \exp\left[-\frac{1}{\sigma_v^2} \|\boldsymbol{Y} - \mathcal{T}_K(h) A_K - \mathcal{T}_U(h) A_U\|^2\right] \qquad (7.7.5)$$

Direct optimization of the SML criterion represents a costly solution. The Expectation–Maximization (EM) [39] algorithm can be used to solve SML using the Hidden Markov Model (HMM) framework: see [40], for a description of different methods. The EM algorithm will converge to the SML solution given a good initialization. A semi–blind SML is formulated in [22].

When the known symbols are dispersed, the "blind problem part" looses its structure as $\mathcal{T}_U(h)$ has no particular structural properties. As a consequence, fast algorithms cannot be built. So for complexity reasons but also for performance reasons (see section 7.6), when the choice is possible, it is preferable to have grouped known symbols.

Suboptimal semi–blind criteria can be constructed when the known symbols are grouped. The first group of proposed semi–blind methods is again based on ML. In that case, a ML based semi–blind criterion can be written as:

Semi–blind Criterion = α_1 Training sequence criterion + α_2 Blind criterion

The weights α_1 and α_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 [19] and [41].

The most interesting semi–blind criteria are the ones based on a blind criterion that is quadratic; the semi-blind criterion of the form above is then also quadratic.

We now mainly focus on deterministic ML methods. We describe a method to optimize blind DML in an iterative quadratic fashion and then combine blind DML to three different TS based criteria.

7.8 Blind DML

A low cost solution to solve blind DML is based on a linear parameterization of the noise subspace [9], [14]. For a given $\boldsymbol{H}(z)$, there exists an $\mathbf{H}^{\perp}(z)$ such that $\mathbf{H}^{\perp}(z)\mathbf{H}(z) = 0$ and $\mathcal{T}(h^{\perp})\mathcal{T}(h) = 0$ where $\mathcal{T}(h^{\perp})$ is the convolution matrix built from $\mathbf{H}^{\perp}(z)$. For example, for $m = 2$ (2 subchannels), $\mathbf{H}^{\perp}(z) = [-\mathrm{H}_2(z) \;\; \mathrm{H}_1(z)]$;

for $m > 2$ different choices are possible, we will consider [42]:

$$\mathbf{H}^{\perp}(z) = \begin{bmatrix} -\mathrm{H}_2(z) & \mathrm{H}_1(z) & 0 & \cdots & 0 \\ 0 & -\mathrm{H}_3(z) & \mathrm{H}_2(z) & \cdots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \mathrm{H}_m(z) & 0 & \cdots & 0 & -\mathrm{H}_1(z) \end{bmatrix}. \tag{7.8.1}$$

The DML criterion for h can then be written as:

$$\min_{\|h\|=1} \boldsymbol{Y}^H \mathcal{T}^H(h^{\perp}) \underbrace{\left[\mathcal{T}(h^{\perp})\mathcal{T}^H(h^{\perp})\right]}_{\mathcal{R}(h)}{}^{+} \mathcal{T}(h^{\perp})\boldsymbol{Y}\,. \tag{7.8.2}$$

We only specify here the norm constraint on the channel; a phase constraint is necessary in the complex channel case [30]. The channel is assumed irreducible.

Iterative Quadratic ML (IQML) is an iterative algorithm which, at each iteration considers the denominator $\mathcal{R}(h) = \mathcal{R}$ as constant, evaluated at the channel estimate from the previous iteration. The criterion becomes quadratic in this way. Using the commutativity of convolution, we can write $\mathcal{T}(h^{\perp})\boldsymbol{Y} = \mathcal{Y}h$; the IQML criterion can then be written as:

$$\min_{\|h\|=1} h^H \mathcal{Y}^H \mathcal{R}^+ \mathcal{Y} h\,. \tag{7.8.3}$$

In the noiseless case, $\mathcal{Y}h^o = 0$: the true channel h^o nulls the (quadratic) criterion (regardless of initialization) and the solution is h^o. At high SNR, a first iteration gives a consistent estimate of h and a second iteration gives the ML solution. At low SNR however, IQML results in a biased estimation of the channel and its performance is poor. Indeed, asymptotically ($M \to \infty$) the IQML cost function becomes equivalent to its expected value by the law of large numbers:

$$\begin{aligned} &\mathrm{tr}\left\{\mathcal{T}^H(h^{\perp})\mathcal{R}^+\mathcal{T}(h^{\perp})\mathrm{E}\boldsymbol{Y}\boldsymbol{Y}^H\right\} = \\ &\mathrm{tr}\left\{\mathcal{T}^H(h^{\perp})\mathcal{R}^+\mathcal{T}(h^{\perp})\boldsymbol{X}\boldsymbol{X}^H\right\} + \sigma_v^2 \mathrm{tr}\left\{\mathcal{T}^H(h^{\perp})\mathcal{R}^+\mathcal{T}(h^{\perp})\right\}\,. \end{aligned} \tag{7.8.4}$$

$h = h^o$ nulls the first term but is not in general the minimal eigenvector of the second term and hence of the sum.

7.8.1 Denoised IQML (DIQML)

The first approach [41] we propose removes the noise from the IQML criterion. We subtract from the IQML criterion an estimate of the noise contribution ($\widehat{\sigma_v^2}$ denotes a consistent estimate of the noise variance):

$$\begin{aligned} &\min_{\|h\|=1} \mathrm{tr}\left\{P_{\mathcal{T}^H(h^{\perp})}\left(\boldsymbol{Y}\boldsymbol{Y}^H - \widehat{\sigma_v^2} I\right)\right\} \Leftrightarrow \min_{\|h\|=1}\left(\boldsymbol{Y}^H P_{\mathcal{T}^H(h^{\perp})}\boldsymbol{Y} - \widehat{\sigma_v^2}\underbrace{\mathrm{tr}\left\{P_{\mathcal{T}^H(h^{\perp})}\right\}}_{\text{constant}}\right) \\ &\Leftrightarrow \min_{\|h\|=1}\left\{h^H\mathcal{Y}^H\mathcal{R}^+(h)\mathcal{Y}h - \widehat{\sigma_v^2}\mathrm{tr}\{\mathcal{T}(h^{\perp})\mathcal{R}^+(h)\mathcal{T}^H(h^{\perp})\}\right\}\,. \end{aligned} \tag{7.8.5}$$

(7.8.5) is solved in the IQML way: we consider $\mathcal{R}(h) = \mathcal{R}$ as constant at each iteration and the problem becomes quadratic:

$$\min_{\|h\|=1} h^H \left\{ \mathcal{Y}^H \mathcal{R}^+ \mathcal{Y} - \widehat{\sigma_v^2} \mathcal{D} \right\} h \tag{7.8.6}$$

where $h^H \mathcal{D} h = \text{tr}\{\mathcal{T}^H(h^\perp)\mathcal{R}^+\mathcal{T}(h^\perp)\}$.

The choice for $\widehat{\sigma_v^2}$ turns out to be crucial. For a finite amount of data, the central matrix $\mathcal{Q} = \mathcal{Y}^H\mathcal{R}^+\mathcal{Y} - \widehat{\sigma_v^2}\mathcal{D}$ is indefinite if $\widehat{\sigma_v^2}$ is not properly chosen: in that case, DIQML will not improve IQML. In order to have a well–defined minimization problem at each iteration, we choose the $\widehat{\sigma_v^2}$ which renders $\mathcal{Q}(h)$ positive with 1 singularity. The problem becomes:

$$\min_{\|h\|=1,\lambda} h^H \left\{ \mathcal{Y}^H \mathcal{R}^+ \mathcal{Y} - \lambda \mathcal{D} \right\} h\,. \tag{7.8.7}$$

with the non–negativity constraint for the central matrix. λ is the minimal generalized eigenvalue of $\mathcal{Y}^H\mathcal{R}^+\mathcal{Y}$ and $\mathcal{D}$ and h is the associated generalized eigenvector. Asymptotically in the number of data, a first iteration gives a consistent estimate of h and a second iteration gives the global minimizer whatever the initialization [43]. Also, $\lambda \to \sigma_v^2$. The performance of IQML is inferior to that of DML. A similar approach was developped independently [44] with a less judicious choice for the noise variance estimate; some related work can also be found in [45]. The next proposed algorithm, PQML, will give the same performance as DML.

7.8.2 Pseudo Quadratic ML (PQML)

PQML [41] is an iterative algorithm which at each iteration tries to null the true gradient of DML. This gradient can be written as $\mathcal{P}(h)h$ where $\mathcal{P}(h)$ is ideally a positive definite matrix. At each iteration $\mathcal{P}(h) = \mathcal{P}$ is considered as constant (evaluated from the previous iteration). The true (unconstrained) DML gradient can then also be interpreted as the (unconstrained) gradient of the following pseudo-quadratic criterion:

$$\min_{\|h\|=1} h^H \mathcal{P}\, h\,. \tag{7.8.8}$$

For the DML problem, the matrix $\mathcal{P}(h)$ can be written as $\mathcal{P}(h) = \mathcal{Y}^H\mathcal{R}^+\mathcal{Y} - \mathcal{B}^H(h)\mathcal{B}(h)$. When $M \to \infty$, $\mathcal{B}^H(h)\mathcal{B}(h) \to \sigma_v^2\mathcal{D}$ + signal term. Evaluated at a consistent h, the signal term becomes negligible, and the effect of $\mathcal{B}^H(h)\mathcal{B}(h)$ is to remove the noise contribution from the IQML Hessian but in a (statistically) more efficient way than DIQML does.

Again, for a finite amount of data, the matrix $\mathcal{P}(h)$ will be indefinite. So by analogy with DIQML, we introduce a scalar λ that will render the matrix $\mathcal{P}(h) = \mathcal{Y}^H\mathcal{R}^+\mathcal{Y} - \lambda\mathcal{B}^H(h)\mathcal{B}(h)$ positive with one singularity. The criterion becomes:

$$\min_{\|h\|=1,\lambda} h^H \left\{ \mathcal{Y}^H \mathcal{R}^+ \mathcal{Y} - \lambda\, \mathcal{B}^H \mathcal{B} \right\} h\,. \tag{7.8.9}$$

with non–negativity constraint on the central matrix. Asymptotically, with a consistent initialization, PQML converges to its minimum at the first iteration and provides the same asymptotic performance as DML. Also, $\lambda \to 1$ in this case. Both DIQML and PQML require a complexity that is linear in the burst length M, and so will the semi–blind criteria presented below.

7.9 Three Suboptimal DML based Semi–Blind Criteria

7.9.1 Split of the Data

The output burst can be decomposed into 3 parts, see figure 7.7 (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 to the blind part of the semi–blind criteria.

7.9.2 Least Squares–DML

In the first semi–blind approach, the overlap zone is incorporated into the blind part of the semi–blind criterion. The data is split as $\boldsymbol{Y} = [\boldsymbol{Y}_{TS}^T \ \boldsymbol{Y}_B^T]^T$, see figure 7.7:

- $\boldsymbol{Y}_{TS} = \mathcal{T}_{TS}(h)A_{TS} + \boldsymbol{V}_{TS}$ 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 unknown symbols and especially the overlap observations, where we do not exploit the knowledge of the known symbols, which will hence be treated as unknown. So, some information is lost. This loss of information can be critical, especially when the training sequence is very short, of less than N symbols.

We apply the DML principle to:

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_{TS} \\ \boldsymbol{Y}_B \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathcal{T}_{TS}(h)A_{TS} \\ \mathcal{T}_B(h)A_B \end{bmatrix}, \sigma_v^2 I \right) \tag{7.9.1}$$

As $\boldsymbol{Y}_{TS}$ and $\boldsymbol{Y}_B$ are decoupled in terms of noise components, the DML criterion for $\boldsymbol{Y}$ is the sum of the DML criteria for $\boldsymbol{Y}_{TS}$ and $\boldsymbol{Y}_B$:

$$\min_{h, A_B} \left\{ \|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h)A_{TS}\|^2 + \|\boldsymbol{Y}_B - \mathcal{T}_B(h)A_B\|^2 \right\} . \tag{7.9.2}$$

This criterion can be optimized by alternating minimizations w.r.t. h and A_B. We can also solve w.r.t. A_B and substitute the solution to get the following semi–blind

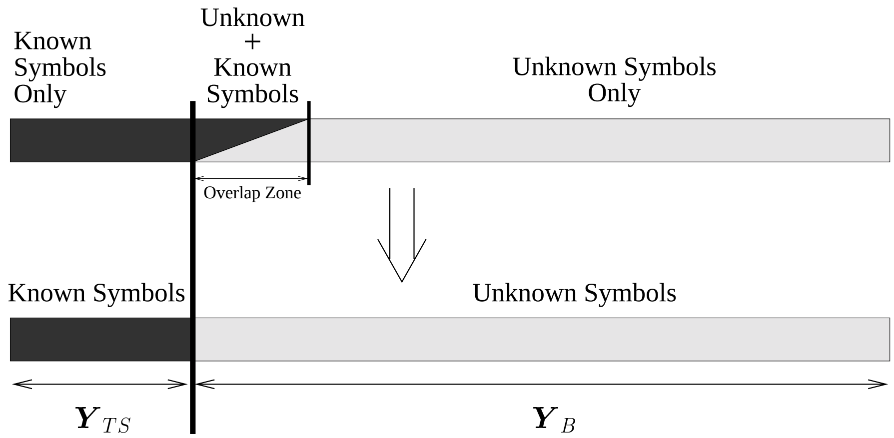

Figure 7.7. Output Burst: split of the data for LS–DML.

DML criterion for h:

$$\min_h \left\{ \|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_{TS}\|^2 + \boldsymbol{Y}_B^H P_{\mathcal{T}_B^H(h^\perp)} \boldsymbol{Y}_B \right\} . \tag{7.9.3}$$

This criterion can also be optimized in the semi–blind PQML way:

$$\min_h \left\{ \|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_{TS}\|^2 + h^H \left(\mathcal{Y}_B^H \mathcal{R}_B^+ \mathcal{Y}_B - \lambda \mathcal{B}_B^H \mathcal{B}_B \right) h \right\} \tag{7.9.4}$$

where λ is chosen as the minimal generalized eigenvalue of $\mathcal{Y}_B^H \mathcal{R}_B^+ \mathcal{Y}_B$ and $\mathcal{B}_B^H \mathcal{B}_B$. We will call (7.9.4) Least–Squares PQML (LS–PQML). LS–PQML represents a suboptimal way of solving the semi–blind problem and the semi–blind identifiability condition for the number of known symbols required no longer holds exactly. For irreducible channels, the criterion requires at least N known symbols to be well–defined: $\mathcal{P}_B(h) = \mathcal{Y}_B^H \mathcal{R}_B^+ \mathcal{Y}_B - \hat{\lambda} \mathcal{B}_B^H \mathcal{B}_B$ is indeed positive semi–definite with 1 singularity, and with N known symbols, $\mathcal{A}_{TS}^H \mathcal{A}_{TS}$ has rank 1, which is sufficient to allow $\mathcal{P}_B(h)$ to be positive definite. For a reducible channel with $N_c - 1$ zeros, asymptotically $\mathcal{P}_B(h) \to \mathcal{X}_B^H \mathcal{R}^+ \mathcal{X}_B$ has N_c singularities, and $N + N_c - 1$ known symbols are necessary to have a well–conditioned problem. Furthermore, for monochannels the semi–blind criterion reduces to its TS part since the blind part does not add any information (in fact, the blind criterion is not defined for monochannels). LS–PQML as well as the following algorithms need an initialization and can be initialized by TS or by one of the algorithms of section 7.10.

7.9.3 Alternating Quadratic DML (AQ–DML)

Here, the overlap zone will be incorporated into the training part of the criterion. The data is split as $\boldsymbol{Y} = [\boldsymbol{Y}_{AQ}^T \; \boldsymbol{Y}_B^T]^T$, see figure 7.8:

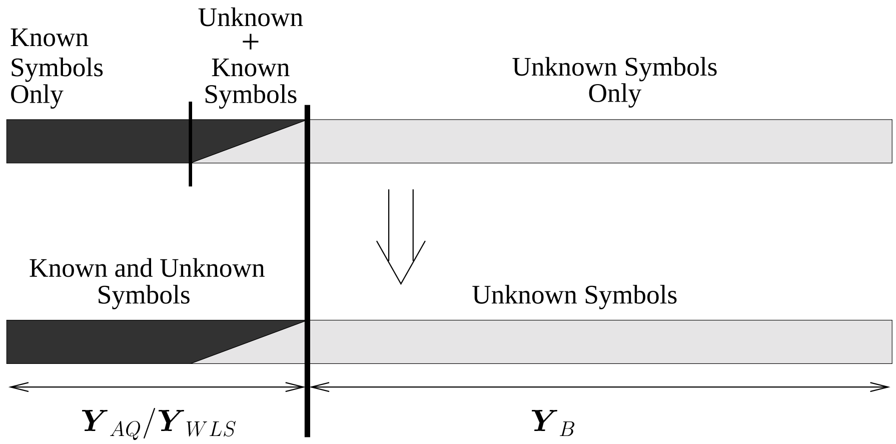

Figure 7.8. Output Burst: split of the data for AQ–DML and WLS-DML.

- $\boldsymbol{Y}_{AQ} = \mathcal{T}_{AQ}(h)A_{AQ} + \boldsymbol{V}_{AQ} = \mathcal{T}'_K(h)A_{TS} + \mathcal{T}'_U(h)A'_U + \boldsymbol{V}_{AQ}$ groups all the observations containing the known symbols A_{TS} plus the overlap observations. The unknown symbols A'_U in $\boldsymbol{Y}_{AQ}$ are considered as deterministic.
- $\boldsymbol{Y}_B = \mathcal{T}_B(h)A_U + \boldsymbol{V}_B$ groups all the observations containing only unknown symbols which are considered as deterministic unknown quantities.

DML is applied to $[\boldsymbol{Y}_{AQ}^T \; \boldsymbol{Y}_B^T]^T$, which gives:

$$\min_{h,A_B,A'_U} \left\{\|\boldsymbol{Y}_{AQ} - \mathcal{T}'_K(h)A_{TS} - \mathcal{T}'_U(h)A'_U\|^2 + \|\boldsymbol{Y}_B - \mathcal{T}_B(h)A_B\|^2\right\} . \qquad (7.9.5)$$

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 $\min\limits_{h,A'_U} \|\boldsymbol{Y}_{AQ} - \mathcal{T}'_K(h)A_{TS} - \mathcal{T}'_U(h)A'_U\|^2$ is solved by alternating minimizations w.r.t. A'_U and h. We keep only the estimate of A'_U, to form the new estimate of $\widehat{A}_{AQ}$: $\widehat{A}_{AQ}^{(i+1)} = [A_K^T \;\; {A'}_U^{(i+1)^T}]^T$.
   - Solve the semi–blind criterion to get $\hat{h}^{(i+1)}$ (with A'_U frozen).
     We can perform alternating minimizations w.r.t. A_B and h, starting from $\hat{h}^{(i)}$, based on the criterion :

$$\min_{h,A_B} \left\{\|\boldsymbol{Y}_{AQ} - \mathcal{T}_{AQ}(h)A_{AQ}^{(i+1)}\|^2 + \|\boldsymbol{Y}_B - \mathcal{T}(h)A_B\|^2\right\} . \qquad (7.9.6)$$

We can alternatively solve the PQML based criterion:

$$\min_h \left\{ \|\boldsymbol{Y}_{AQ} - \mathcal{T}_{AQ}(h) A_{AQ}^{(i+1)}\|^2 + h^H \left(\mathcal{Y}_B^H \mathcal{R}_B^+(\hat{h}^{(i)}) \mathcal{Y}_B - \lambda\, \mathcal{B}_B^H(\hat{h}^{(i)}) \mathcal{B}_B(\hat{h}^{(i)}) \right) h \right\}. \tag{7.9.7}$$

7.9.4 Weighted–Least–Squares–PQML (WLS–PQML)

WLS–PQML is based on the same decomposition as for 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 σ_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 $\boldsymbol{Y}_B$:

$$\left\{ \begin{array}{l} \boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_{WLS} \\ \boldsymbol{Y}_B \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathcal{T}'_K(h) A_{TS} \\ \mathcal{T}_B(h) A_B \end{bmatrix}, \begin{bmatrix} C_{Y_{WLS} Y_{WLS}} & 0 \\ 0 & \sigma_v^2 I \end{bmatrix} \right), \\ C_{Y_{WLS} Y_{WLS}} = \sigma_a^2 \mathcal{T}'_U(h) \mathcal{T}'^H_U(h) + \sigma_v^2 I \end{array} \right. \tag{7.9.8}$$

The mixed ML criterion is:

$$\min_{h, A_B, \sigma_v^2} \left\{ \ln\det C_{Y_{WLS} Y_{WLS}} + \left(\boldsymbol{Y}_{WLS} - \mathcal{T}'_K(h) A_{TS}\right)^H C^{-1}_{Y_{WLS} Y_{WLS}} \left(\boldsymbol{Y}_{WLS} - \mathcal{T}'_K(h) A_{TS}\right) + \ln\det \sigma_v^2 I + \frac{1}{\sigma_v^2} \|\boldsymbol{Y}_B - \mathcal{T}_B(h) A_B\|^2 \right\}. \tag{7.9.9}$$

We consider σ_v^2 as known (in practice, it will be estimated separately). $C_{Y_{WLS} Y_{WLS}}$ is considered as constant (computed using the channel estimate from the previous iteration). The criterion then becomes:

$$\min_{h, A_B} \left\{ \|\boldsymbol{Y}_{WLS} - \mathcal{T}_{WLS}(h) A_{WLS}\|^2_{C^{-1}_{Y_{WLS} Y_{WLS}}} + \frac{1}{\sigma_v^2} \|\boldsymbol{Y}_B - \mathcal{T}_B(h) A_B\|^2 \right\}. \tag{7.9.10}$$

Solved in the PQML fashion, the criterion becomes:

$$\min_h \left\{ \|\boldsymbol{Y}_{WLS} - \mathcal{T}_{WLS}(h) A_{WLS}\|^2_{C^{-1}_{Y_{WLS} Y_{WLS}}} + \frac{1}{\sigma_v^2} h^H \left(\mathcal{Y}_B^H \mathcal{R}^+ \mathcal{Y}_B - \lambda \mathcal{B}_B^H \mathcal{B}_B \right) h \right\} \tag{7.9.11}$$

The approximation of considering $C_{Y_{WLS} Y_{WLS}}$ as constant is justified in [41] by using a semi–blind PQML strategy.

AQ–PQML and WLS–PQML outperform LS–PQML because the information coming from the known symbols in the overlap zone is used.

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–defined problem.

A performance study of the criteria is provided in section 7.11.

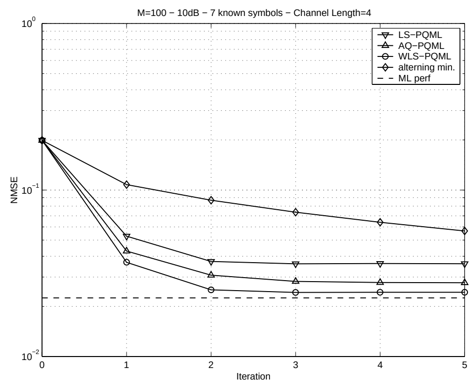

Figure 7.9. Semi–blind algorithms.

7.9.5 Simulations

In figure 7.9, we show the NMSE given by all the PQML based semi–blind algorithms as well as the optimal semi–blind DML in (7.7.1) solved by AQML 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 lower limit for TS identifiability). The PQML based semi–blind criterion and AQML are initialized by TS. The semi–blind algorithms improve dramatically TS performance. WLS–PQML is the best of the algorithms with performance close to the theoretical ML performance (this was confirmed by other simulations). We also notice the slow convergence of AQML (which in fact requires many iterations for A'_U per indicated iteration for h).

To illustrate the lack of robustness of blind methods, we show in figure 7.10 simulations with 5000 Monte–Carlo runs for channel, noise and input symbols: the particularly poor performance of blind estimation can be noticed. This does not mean that blind PQML is a weak algorithm. Among the 5000 realizations of the channels, some gave ill–conditioned channels resulting in poor performance and yielding poor averaged performance. The semi–blind algorithm does not suffer from this problem.

7.10 Semi–Blind Criteria as a Combination of a Blind and a TS Based Criteria

Some semi–blind algorithms have been proposed that linearly combine a TS and a blind criterion: in [46] a semi–blind criterion is proposed based on the (unweighted) Subspace Fitting (SF) criterion; in [47], [48], another one is based on the blind CMA criterion.

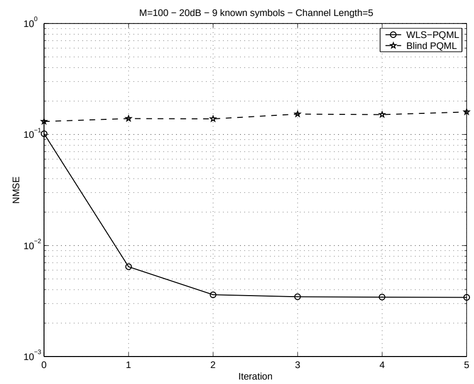

Figure 7.10. Comparison between semi–blind PQML and blind PQML.

7.10.1 Semi–Blind SRM Example

Semi–Blind Subchannel Response Matching (SRM) [49], [13], [14], [25] illustrates the fact that one has to be careful when building a semi–blind criterion that way. Blind SRM can be seen as a non–weighted version of IQML: $\min_h h^H \mathcal{Y}^H \mathcal{Y} h$. Consider the following semi–blind cost function:

$$\alpha\, h^H \mathcal{Y}_B^H \mathcal{Y}_B h + \|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_{TS}\|^2 \,. \tag{7.10.1}$$

This criterion is based on the decomposition of figure 7.7. An intuitive way to weigh both TS and blind parts is to associate them with the number of data they are built from, as suggested in [46] for semi–blind subspace fitting. In the SRM case, the optimal α would then be equal to 1.

In figure 7.11, we show the NMSE for the channel averaged over 100 Monte–Carlo realizations of channel (with i.i.d. coefficients), noise and input symbols. The NMSE for (7.10.1) is plotted w.r.t. the value of α in dotted lines. For $\alpha = 1$, semi–blind SRM gives worse performance than TS estimation.

The blind SRM criterion gives unbiased estimates only under a 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. In [50], the criterion as is 7.10.1 was proposed without being denoised or properly scaled.

For the channel estimates to be unbiased, the term $\mathcal{Y}_B^H \mathcal{Y}_B$ needs to be denoised. We remove $\lambda_{min}(\mathcal{Y}_B^H \mathcal{Y}_B) I$ from $\mathcal{Y}_B^H \mathcal{Y}_B$ (where $\lambda_{min}(\mathcal{Y}_B^H \mathcal{Y}_B)$ denotes 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 weight α remains unsolved. A way to determine this factor would be to optimize the asymptotic performance of the semi–blind SRM channel estimate w.r.t. α. However, since an analytical optimization appears impossible, one would have to resort to search techniques which would represent an increase in complexity.

In the next section, we construct semi–blind SRM as an approximation of semi–blind DIQML: in this way, the blind SRM part will be automatically denoised and scaled.

Semi–Blind SRM as an Approximation of DIQML

We know that the semi–blind ML criterion gives the optimal weighting between the blind and training sequence parts:

$$h^H \mathcal{Y}_B^H \mathcal{R}^+(h) \mathcal{Y}_B h + \|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_{TS}\|^2 . \tag{7.10.2}$$

We now neglect the off-diagonal terms in $\mathcal{R}(h)$: $\mathcal{R}(h) \approx \mathcal{D}(h)$. For $m = 2$ channels, the diagonal elements are constant and equal to $\|h\|^2$. For $m > 2$, the diagonal contains the squared norm of pairs of subchannels. For example, for $m = 3$, the first three diagonal elements are: $\|\mathrm{H}_1\|^2 + \|\mathrm{H}_2]\|^2$, $\|\mathrm{H}_2\|^2 + \|\mathrm{H}_3\|^2$, $\|\mathrm{H}_1\|^2 + \|\mathrm{H}_3\|^2$ and these three values are repeated along the diagonal M times.

With this approximation, (7.10.2) becomes:

$$\|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_{TS}\|^2 + h^H \mathcal{Y}_B^H \mathcal{D}^+(h) \mathcal{Y}_B h \tag{7.10.3}$$

and we optimize this criterion in the DIQML fashion in order to denoise it; $\mathcal{D}(h) = \mathcal{D}$ is considered as constant. The semi–blind criterion becomes:

$$\min_h \left\{ \|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_{TS}\|^2 + h^H \left(\mathcal{Y}_B^H \mathcal{D}^{-1} \mathcal{Y}_B - \widehat{\sigma_v^2} \mathcal{D}_v \right) h \right\} \tag{7.10.4}$$

where $h^H \mathcal{D}_v h = \mathrm{tr}\left\{ \mathcal{T}_B^H(h^\perp) \mathcal{D}^{-1} \mathcal{T}_B(h^\perp) \right\}$, and $\widehat{\sigma_v^2}$ is the minimal 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 estimated from the denoised sample second–order moment $r_{yy}(0) = \sigma_a^2 \mathcal{T}_1(h) \mathcal{T}_1^H(h)$: $\hat{r}_{yy}(0) - \widehat{\sigma_v^2} I = \frac{1}{M} \sum_{k=0}^{M-1} \boldsymbol{y}(k) \boldsymbol{y}^H(k) - \widehat{\sigma_v^2} I$. As will be seen in the simulations, at low SNR, the weight on the blind part should in fact be smaller than the true value $\mathcal{D}^o$ of $\mathcal{D}$ indicates. So instead of estimating the energy of each channel from the denoised $r_{yy}(0)$, we use the noisy version. The resulting $\mathcal{D}$ will be larger than $\mathcal{D}^o$.

In general, the different channels tend to have the same energy, so that $\mathcal{D}$ can in turn be approximated by a multiple of an identity matrix, the multiple being $\widehat{\|h\|^2} \frac{m}{2}$. When $m = 2$, this approximation is exact. With $\mathcal{D}$ being a multiple of identity, $\widehat{\sigma_v^2}$ is the minimal eigenvalue of $\mathcal{Y}_B^H \mathcal{Y}_B$, and the semi–blind criterion becomes:

$$\min_h \left\{ \frac{2}{m} \frac{1}{\widehat{\|h\|^2}} \; h^H \left(\mathcal{Y}_B^H \mathcal{Y}_B - \lambda_{min}(\mathcal{Y}_B^H \mathcal{Y}_B) I \right) h + \|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_{TS}\|^2 \right\} . \tag{7.10.5}$$

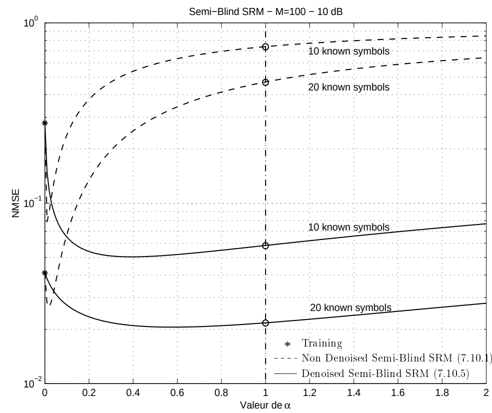

Figure 7.11. Semi–blind subspace fitting built as a linear combination of blind SRM and training sequence based criteria.

An alternative to this semi–blind SRM criterion is to use the decomposition of figure 7.8 and to use WLS or AQML to handle the training sequence part.

In figure 7.11 in solid lines, we show the performance of the corrected semi–blind criterion. The scalar α scales the blind part in (7.10.5). The value $\alpha = 1$ gives approximately the optimal performance and we notice that in fact the performance is roughly constant in the neighborhood of $\alpha = 1$.

In figure 7.12, semi–blind SRM is used to initialize the different DML based semi–blind algorithms in the case where the training sequence is too short to estimate the channel.

7.10.2 Subspace Fitting Example

Blind Subspace Fitting

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):

$$R_{Y_L Y_L} = \mathcal{T}_L(h) \left[\sum_{k=0}^{M-L-1} A_L(k) A_L^H(k) \right] \mathcal{T}_L^H(h) + \sigma_v^2 I \quad . \tag{7.10.6}$$

Provided the blind deterministic identifiability conditions [DetB] are fulfilled, the matrix $\sum_{k=0}^{M-L-1} A_L(k) A_L^H(k)$ is non-singular, so the space spanned by the matrix $\mathcal{T}_L(h) \left[\sum_{k=0}^{M-L-1} A_L(k) A_L^H(k)\right] \mathcal{T}_L^H(h)$ is the signal subspace. $R_{Y_L Y_L}$ admits $L + N - 1$ (the dimension of the signal subspace) eigenvectors belonging to the signal subspace, and $Lm - (L + N - 1)$ eigenvectors belonging to the noise subspace and all associated to the eigenvalue σ_v^2. The eigendecomposition of $R_{Y_L Y_L}$ is:

$$R_{Y_L Y_L} = V_{\mathcal{S}} \Lambda_{\mathcal{S}} V_{\mathcal{S}}^H + V_{\mathcal{N}} \Lambda_{\mathcal{N}} V_{\mathcal{N}}^H \tag{7.10.7}$$

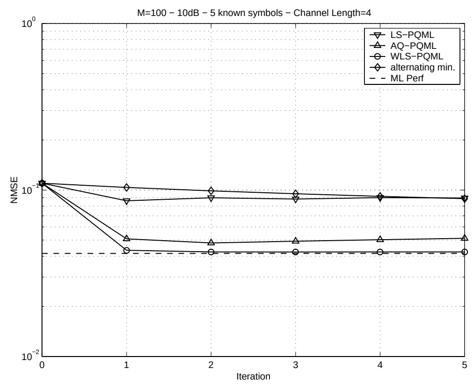

Figure 7.12. DML based semi–blind algorithms initialized by semi–blind SRM. Case of a too short training sequence to estimate the channel via TS alone.

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 $\widehat{V}_{\mathcal{S}}$ and $\widehat{V}_{\mathcal{N}}$ be estimates of the signal and noise eigenvectors obtained from the sample covariance matrix. Signal Subspace Fitting (SSF) tries to fit the column space of $\mathcal{T}(h)$ to that of $\widehat{V}_{\mathcal{S}}$ through the quadratic criterion:

$$\min_{\|h\|^2=1} \|P_{\widehat{V}_{\mathcal{N}}} \mathcal{T}(h)\|^2 \quad \Leftrightarrow \quad \min_{\|h\|^2=1} h^H \mathcal{S}^H \mathcal{S}\, h \ . \tag{7.10.8}$$

(see citeAbedMeraim:Subsp97, for a description of the structure of $\mathcal{S}$).

Semi–Blind Subspace Fitting

Consider now the following semi–blind cost function:

$$\alpha M_U \ h^H \mathcal{S}_B^H \mathcal{S}_B \ h + \|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_{TS}\|^2 \, . \tag{7.10.9}$$

We adopt here the decomposition of figure 7.7. In [46], αM_U was chosen equal to the number of data on which the blind criterion is based, *i.e.* $\alpha = 1$.

In figure 7.13 (left), we plot the NMSE of channel estimation w.r.t. α for different sizes L, for SNR=10dB and $M_K = 10$ known symbols. For $L = N$, the semi–blind criterion is relatively insensitive to the value of α. For L larger than N however, the criterion is visibly very sensitive to the value of α. The choice $\alpha = 1$ gives performance worse than that for training sequence based estimation for $L > N$. These simulations suggest that the linearly combined semi–blind algorithm is sensitive to the dimension of the noise subspace which varies when L varies.

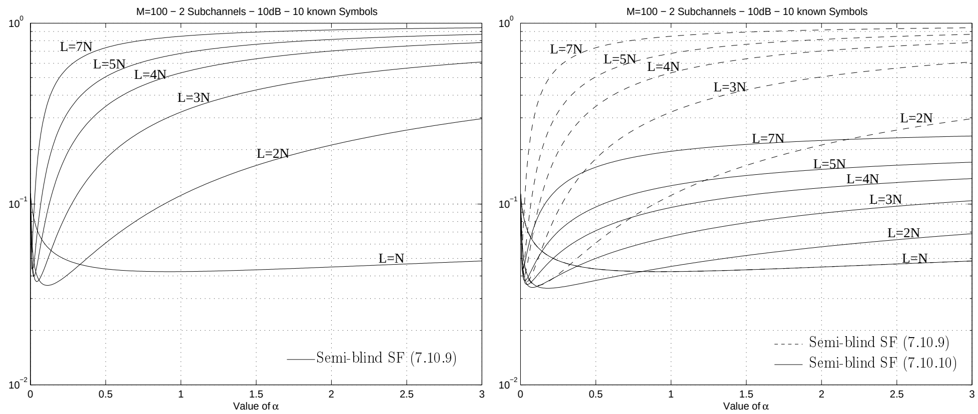

Figure 7.13. Semi–blind subspace fitting built as a linear combination of blind SF and training sequence based criteria.

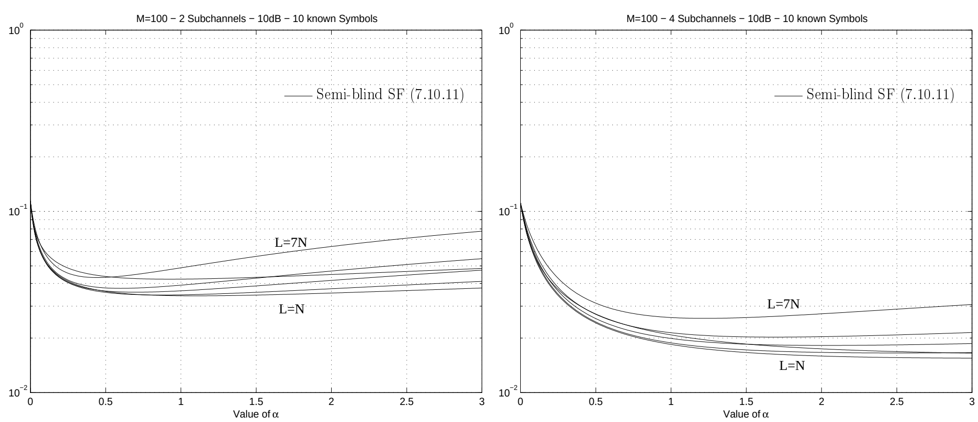

Figure 7.14. Semi–blind subspace fitting built as a linear combination of weighted and denoised blind SF and training sequence based criteria.

"Denoised" semi–blind SF

As will be seen in section 7.11, the behavior of the semi–blind criterion of the form (7.10.9) depends heavily on the "smallest" eigenvalues of $\mathcal{S}_B^H\mathcal{S}_B$. In particular, we will see that if the smallest eigenvalues are forced to 0, the performance of the semi–blind criterion is constant w.r.t. α. When the channel has N_c-1 zeros, $\mathcal{S}_B^H\mathcal{S}_B$ has theoretically N_c singularities. It is found that forcing the N_c (or more as discussed in section 7.11) smallest eigenvalues to zero will render the performance less sensitive to the value of α.

For complexity reasons, we force only one eigenvalue to zero, the smallest one, via $\mathcal{S}_B^H\mathcal{S}_B-\lambda_{min}(\mathcal{S}_B^H\mathcal{S}_B)I$. Note that this does not remove a noise contribution as in the case of semi–blind PQML or SRM; we just force the structure of estimated quantities to be closer to that of the theoretical quantity in the blind part of the criterion. The new semi–blind SF criterion is then:

$$\min_h \left\{\alpha M_U\, h^H\left(\mathcal{S}_B^H\mathcal{S}_B-\lambda_{min}(\mathcal{S}_B^H\mathcal{S}_B)I\right)h+\|\boldsymbol{Y}_{TS}-\mathcal{T}_{TS}(h)A_{TS}\|^2\right\}\ . \quad (7.10.10)$$

In figure 7.13 (right), we show the performance of this algorithm w.r.t. α for 500 Monte–Carlo realizations of the channel, noise and input symbols. We can notice the significant effect of the denoising on the semi–blind algorithm performance. The performance still depends on the value of α though (in fact this denoising is not sufficient here to render the performance optimal and constant around $\alpha=1$).

Denoised and scaled semi–blind SF

As a heuristic solution, we propose to scale the blind part of the semi–blind SF criterion by the dimension $\mathcal{N}$ of the noise subspace. The resulting criterion is:

$$\min_h \left\{\alpha\frac{M_U}{\mathcal{N}}h^H\left(\mathcal{S}_B^H\mathcal{S}_B-\lambda_{min}(\mathcal{S}_B^H\mathcal{S}_B)I\right)h+\|\boldsymbol{Y}_{TS}-\mathcal{T}_{TS}(h)A_{TS}\|^2\right\}\ . \quad (7.10.11)$$

In figure 7.14, we show the performance of this algorithm in the case of random channels with 2 subchannels (left) and 4 subchannels (right). We see that this semi–blind SF criterion gives satisfactory results for $\alpha=1$.

7.11 Performance of Semi–Blind Quadratic Criteria

We consider here the following general semi–blind quadratic criterion of the form:

$$\min_h \left\{\alpha M_U h^H\widehat{Q}_B h+\|\boldsymbol{Y}_{TS}-\mathcal{T}_{TS}(h)A_{TS}\|^2\right\}\ . \quad (7.11.1)$$

For instance, $\widehat{Q}_B=\frac{1}{M_U}\left[\mathcal{Y}_B^H\mathcal{R}^+\mathcal{Y}_B-\hat{\lambda}\mathcal{B}_B^H\mathcal{B}_B\right]$ for semi–blind PQML, $\widehat{Q}_B=\frac{2}{m\widehat{\|h\|^2}M_U}\left[\mathcal{Y}_B^H\mathcal{Y}_B-\lambda_{min}(\mathcal{Y}_B^H\mathcal{Y}_B)I\right]$ for semi–blind SRM and $\widehat{Q}_B=\frac{1}{\mathcal{N}}\left[\mathcal{S}_B^H\mathcal{S}_B-\lambda_{min}(\mathcal{S}_B^H\mathcal{S}_B)I\right]$ for semi–blind SF. $\widehat{Q}_B$ tends asymptotically to Q_B which has N_c eigenvalues equal to zero if the channel has N_c-1 zeros (e.g. due to channel length overestimation).

Let $\widehat{Q}_B = \widehat{\mathcal{W}}_1 \widehat{\Lambda}_1 \widehat{\mathcal{W}}_1^H + \widehat{\mathcal{W}}_2 \widehat{\Lambda}_2 \widehat{\mathcal{W}}_2^H$ be the eigencomposition of $\widehat{Q}_B$, $\widehat{\Lambda}_2 \to 0$ asymptotically. Note that the smallest eigenvalue of $\widehat{Q}_B$ is 0 for the semi–blind criteria proposed. One can show that the elements of $\widehat{\Lambda}_2$ (not exactly equal to 0) are of order $1/M_U$ for PQML, SRM and SF.

The solution of (7.11.1) is:

$$\hat{h} = \left(\alpha M_U \widehat{Q}_B + \mathcal{A}_{TS}^H \mathcal{A}_{TS}\right)^{-1} \mathcal{A}_{TS}^H \boldsymbol{Y}_{TS} \ . \tag{7.11.2}$$

we get for the channel estimation error $\Delta h = \hat{h} - h^o$:

$$\Delta h = -\alpha M_U \left(\alpha M_U \widehat{Q}_B + \mathcal{A}_{TS}^H \mathcal{A}_{TS}\right)^{-1} \widehat{Q}_B h^o + \left(\alpha M_U \widehat{Q}_B + \mathcal{A}_{TS}^H \mathcal{A}_{TS}\right)^{-1} \mathcal{A}_{TS}^H \boldsymbol{V}_{TS} \ . \tag{7.11.3}$$

We study the performance of the semi–blind criterion (7.11.1) for the asymptotic cases:

- M_U and M_K infinite, with condition $\frac{\sqrt{M_U}}{M_K} \to 0$, which accounts for the fact that the TS part of the criterion should not be negligible w.r.t. the blind part [51].
- M_U infinite, M_K finite, with $M_K \geq 2N-1$.

A related asymptotic analysis exists in [52] for semi–blind subspace; the second asymptotic condition is $M_K \ll M_U$, with M_K infinite however. For each asymptotic condition, the channel estimate is unbiased.

7.11.1 M_U and M_K infinite

In that case, it can be shown as in [53], [43] that:

$$C_{\Delta h \Delta h} = \left(\alpha Q_B + \beta \sigma_a^2\right)^{-1} \left(\alpha^2 \,\mathrm{E}\left[\widetilde{Q}_B h^o h^{o\,H} \widetilde{Q}_B^H\right] + \frac{\beta}{M_U} \sigma_v^2 \sigma_a^2\right) \left(\alpha Q_B + \beta \sigma_a^2\right)^{-1} , \tag{7.11.4}$$

with $\beta = \frac{M_K}{M_U}$ and $\widetilde{Q}_B = \widehat{Q}_B - Q_B$; $\mathrm{E}\left[\widetilde{Q}_B h^o h^{o\,H} \widetilde{Q}_B^H\right]$ is of order $1/M_U$. The performance in that case depends on the value of α. Strictly speaking, in order to optimize the performance, it would be necessary to find the optimal α, which represents an additional computational cost and hence is typically avoided.

7.11.2 M_U infinite, M_K finite

We assume in this analysis that $M_K \geq 2N-1$, which also means that $\mathcal{A}_{TS}^H \mathcal{A}_{TS}$ is invertible. Let the Cholesky decomposition of $\mathcal{A}_{TS}^H \mathcal{A}_{TS}$ be:

$$\mathcal{A}_{TS}^H \mathcal{A}_{TS} = \left(\mathcal{A}_{TS}^H \mathcal{A}_{TS}\right)^{1/2} \left(\mathcal{A}_{TS}^H \mathcal{A}_{TS}\right)^{H/2} = \mathcal{R}^{1/2} \mathcal{R}^{H/2} \ . \tag{7.11.5}$$

Then

$$\left(\alpha M_U \widehat{Q}_B + \mathcal{A}_{TS}^H \mathcal{A}_{TS}\right)^{-1} = \mathcal{R}^{-H/2}\left(\alpha M_U \mathcal{R}^{-1/2}\widehat{Q}_B\mathcal{R}^{-H/2} + I\right)^{-1}\mathcal{R}^{-1/2} \,. \tag{7.11.6}$$

Let the eigendecomposition of $\mathcal{R}^{-1/2}\widehat{Q}_B\mathcal{R}^{-H/2} = \widehat{Q}'_B$ now be:

$$\widehat{Q}'_B = \widehat{\mathcal{W}}'\widehat{\Lambda}'\widehat{\mathcal{W}}'^H = \widehat{\mathcal{W}}'_1\widehat{\Lambda}'_1\widehat{\mathcal{W}}'^H_1 + \widehat{\mathcal{W}}'_2\widehat{\Lambda}'_2\widehat{\mathcal{W}}'^H_2 \,. \tag{7.11.7}$$

$\widehat{\Lambda}'_2 \to 0$ asymptotically in the amount of data M or in SNR. We have

$$\begin{gathered}\left(\alpha M_U \mathcal{R}^{-1/2}\widehat{Q}_B\mathcal{R}^{-H/2} + I\right)^{-1} = \left(\alpha M_U \widehat{\mathcal{W}}'\widehat{\Lambda}'\widehat{\mathcal{W}}'^H + I\right)^{-1} = \\ \widehat{\mathcal{W}}'\left(\alpha M_U\widehat{\Lambda}' + I\right)^{-1}\widehat{\mathcal{W}}'^H = \widehat{\mathcal{W}}'_1\left(\alpha M_U\widehat{\Lambda}'_1 + I\right)^{-1}\widehat{\mathcal{W}}'^H_1 + \widehat{\mathcal{W}}'_2\left(\alpha M_U\widehat{\Lambda}'_2 + I\right)^{-1}\widehat{\mathcal{W}}'^H_2 .\end{gathered} \tag{7.11.8}$$

Consider now the following cases:

- Irreducible channel:
  In that case, $\widehat{\Lambda}'_2$ has only one element: $\widehat{\Lambda}'_2 = 0$. At first order, the first term of the last equation in (7.11.8) is negligible

$$\begin{aligned}\left(\alpha M_U \mathcal{R}^{-1/2}\widehat{Q}_B\mathcal{R}^{-H/2} + I\right)^{-1} &= \widehat{\mathcal{W}}'_2\widehat{\mathcal{W}}'^H_2 \\ &= \mathcal{W}'_2\mathcal{W}'^H_2 \qquad \text{at first order} \\ &= P_{\mathcal{R}^{H/2}\mathcal{W}_2}\end{aligned} \tag{7.11.9}$$

  In that case, the first term in (7.11.3) is negligible, and the performance of the semi–blind algorithm does not depend on the (finite) value of α.

- Reducible channel:

$$\left(\alpha M_U \mathcal{R}^{-1/2}\widehat{Q}_B\mathcal{R}^{-H/2} + I\right)^{-1} = \widehat{\mathcal{W}}'_2\left(\alpha M_U\widehat{\Lambda}'_2 + I\right)^{-1}\widehat{\mathcal{W}}'^H_2 \tag{7.11.10}$$

  As the non–zero elements of $\widehat{\Lambda}'_2$ are of order $\frac{1}{M_U}$, the term $\alpha M_U\widehat{\Lambda}'_2$ is not negligible w.r.t. I. The expression for the performance will then vary with α. A way to overcome this would be to force $\widehat{\Lambda}'_2$ to zero: in that case, the performance becomes independent of α. This solution would require structural knowledge about the channel, and the eigendecomposition of $\widehat{Q}_B$, which adds some complexity.

In any case, if $\widehat{\Lambda}'_2 = 0$, the estimation performance is:

$$C_{\Delta h\Delta h} = \sigma_v^2\mathcal{W}_2\left(\mathcal{W}_2^H\mathcal{A}_{TS}^H\mathcal{A}_{TS}\mathcal{W}_2\right)^{-1}\mathcal{W}_2^H \,. \tag{7.11.11}$$

If $\widehat{\Lambda}'_2 \neq 0$, the performance is inferior to the one in (7.11.11). Expression 7.11.11 can be interpreted as the performance of the estimation of $\mathrm{H}_c(z)$ by training sequence with perfect knowledge of the irreducible part $\mathbf{H}_I(z)$. Indeed, the vector channel can be decomposed as $h = T_I h_c$ where T_I is block Toeplitz with $\left[h_I^T \;\; 0_{1\times(N_c-1)m}\right]^T$ as first column. The TS criterion for h_c is:

$$\min_{h_c} \|\boldsymbol{Y}_{TS} - \mathcal{A}_{TS} T_I h_c\|^2 \tag{7.11.12}$$

and the corresponding performance of estimation of h (knowing h_I) is:

$$C_{\Delta h \Delta h} = \sigma_v^2 T_I \left(T_I^H \mathcal{A}_{TS}^H \mathcal{A}_{TS} T_I\right)^{-1} T_I^H \;. \tag{7.11.13}$$

As T_I and $\mathcal{W}_2$ have the same column space, both expressions (7.11.11) and (7.11.13) are equal.

When $\widehat{\Lambda}_2' = 0$, the channel estimate can be rewritten as:

$$\hat{h} = \widehat{\mathcal{W}}_2 \left(\widehat{\mathcal{W}}_2^H \mathcal{A}_{TS}^H \mathcal{A}_{TS} \widehat{\mathcal{W}}_2\right)^{-1} \widehat{\mathcal{W}}_2^H \mathcal{A}_{TS}^H \boldsymbol{Y}_{TS} \tag{7.11.14}$$

The projected semi–blind subspace estimate proposed in [52] is

$$\hat{h} = \widehat{\mathcal{W}}_2 \widehat{\mathcal{W}}_2^H \left(\mathcal{A}_{TS}^H \mathcal{A}_{TS}\right)^{-1} \mathcal{A}_{TS}^H \boldsymbol{Y}_{TS} \tag{7.11.15}$$

and is based on the approximation $\mathcal{A}_{TS}^H \mathcal{A}_{TS} \sim M_K \sigma_a^2 I$; its performance is however not very good because of this approximation.

Furthermore, when considering expression (7.11.4), with M_K finite, we find expression (7.11.11) so the asymptotic expression in M_K and M_U is valid when M_K is finite also. So, when Λ_2' is forced to 0, there is a continuity between both expressions (7.11.4) and (7.11.11); so expression (7.11.4) is valid in any case and should be used to characterize the performance of the semi–blind criteria. This is however not true when the N_c smallest eigenvalues of $\widehat{Q}_B$ are not exactly equal to 0: in that case there is a discontinuity in the expressions, and the two analyses do not yield the same results. In general, in this last case, it may not be obvious to know in practice which analysis between M_K infinite or M_K finite is the appropriate one.

In practice, even when the criterion is correctly denoised, the performance for a small M_K is not constant w.r.t. α. For randomly chosen channels, in general, the matrix $\widehat{Q}_B$ exhibits in fact more than N_c small eigenvalues (the extra small eigenvalues are sufficiently small to influence the performance).

We consider here semi–blind SRM and we desire to force n eigenvalues of $\mathcal{Y}_B^H \mathcal{Y}_B$ to zero. Let λ_n be the largest of the n smallest eigenvalues of $\mathcal{Y}_B^H \mathcal{Y}_B$. In order to denoise, we form the matrix $\mathcal{Y}_B^H \mathcal{Y}_B - \lambda_n I$ and force its negative eigenvalues to zero. For randomly chosen channels of length 5 (so a priori irreducible channels) we force $n = 1$ to 5 eigenvalues (the smallest ones) to zero: in figure 7.15 (right), we show the resulting NMSE for 1000 runs of the channel, noise and input symbols. For a number of 3 to 5 eigenvalues forced to 0, we see that the performance is not dependent on the value of α; however the performance is degraded compared to the performance with 1 and 2 eigenvalues forced to 0. In figure 7.15 (left), we show the average distribution of the eigenvalues of $\mathcal{Y}_B^H \mathcal{Y}_B$.

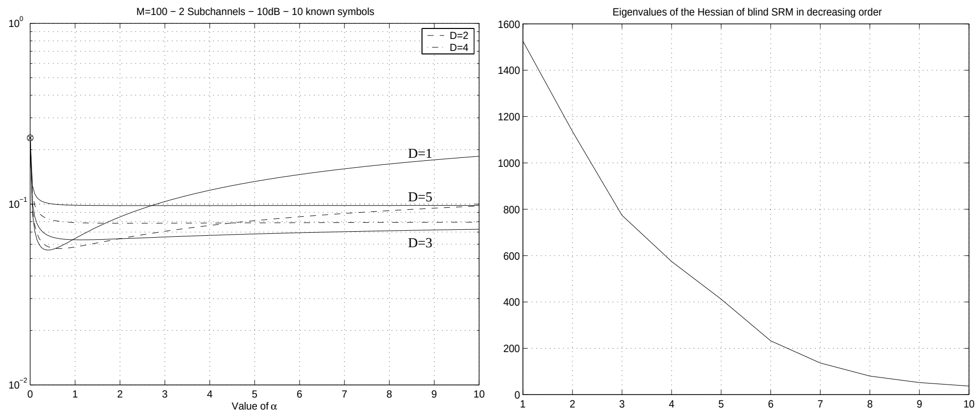

Figure 7.15. Semi–blind SRM: D=1–5 eigenvalues forced to 0.

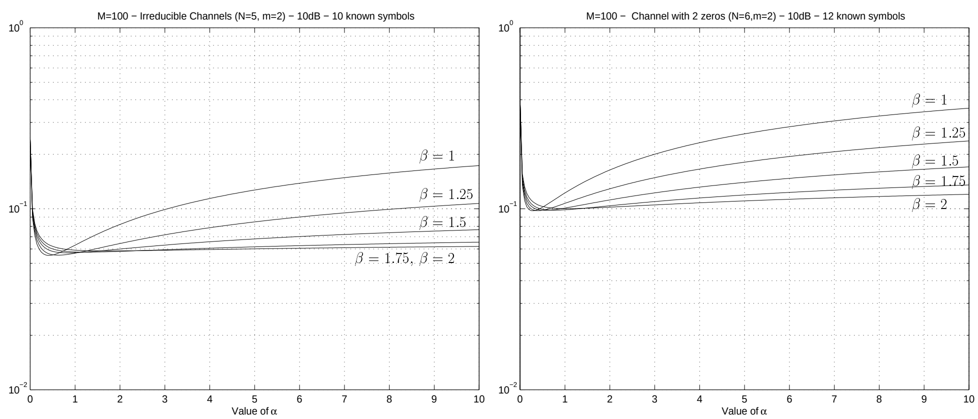

Figure 7.16. Semi–blind SRM: β=1–2.

The number of eigenvalues to force to 0 depends on various parameters such as the channel impulse response and the SNR. We propose here alternatively a more adaptive denoising. We form $\mathcal{Y}_B^H\mathcal{Y}_B - \beta\lambda_{min}(\mathcal{Y}_B^H\mathcal{Y}_B)I$, where $\beta \geq 1$ is an amplification factor, and force to 0 the negative eigenvalues. In figure 7.16 (left), we show the NMSE for 500 realizations of the channel (a priori irreducible), noise and the input symbols; in figure 7.16 (right), the channels have 2 zeros. We see that performance improves as β grows for $\beta \leq 2$: $\beta \in (1.5, 2)$ gives satisfactory results.

The proposed denoised PQML, SRM and SF based algorithms force to zero only one eigenvalue and as already stated have a performance that depends on α. However, the algorithms were constructed (weighted) such that the optimal α is approximately equal to 1: this solution is preferable (to forcing more (n)

eigenvalues to zero) because it is less complex. The β amplification mechanism just mentioned may be considered as a computationally cheap improvement though (forcing nonnegative definiteness can be done using the Cholesky decomposition).

7.11.3 Optimally Weighted Quadratic Criteria

M_U and M_K infinite

We consider a semi–blind criterion built from a quadratic blind criterion of the form $\min_h h^H \mathcal{U}_B^H \mathcal{U}_B h$ with $\mathcal{U}_B h \to 0$ (for $M_U \to \infty$ or SNR $\to \infty$). We consider the case of an infinite (large) amount of known and unknown symbols. We know the optimal weighting matrix $\mathcal{W}$ of the weighted least–squares criterion

$$\min_h \left\| \begin{array}{c} \mathcal{U}_B h \\ \boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_{TS} \end{array} \right\|_{\mathcal{W}^+} \tag{7.11.16}$$

(with a decomposition based on figure 7.7). Indeed:

$$\mathcal{W} = \mathrm{E} \left[\begin{array}{c} \mathcal{U}_B h^o \\ \boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h^o) A_{TS} \end{array} \right] \left[\begin{array}{c} \mathcal{U}_B h^o \\ \boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h^o) A_{TS} \end{array} \right]^H = \left[\begin{array}{cc} \mathcal{W}_B & 0 \\ 0 & \sigma_v^2 I \end{array} \right] \tag{7.11.17}$$

with $\mathcal{W}_B = \mathrm{E}\left[\mathcal{U}_B h^o h^{o\,H} \mathcal{U}_B^H\right]$. The optimally weighted semi–blind criterion becomes:

$$\min_h \left\{ h^H \mathcal{U}_B^H \mathcal{W}_B^+ \mathcal{U}_B h + \frac{1}{\sigma_v^2} \| \boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h) A_{TS} \|^2 \right\} \tag{7.11.18}$$

For these asymptotic conditions, finding the right scale factor α to optimize the performance of the non–weighted part of the criterion is difficult as already mentioned in section 7.11.1, however finding the right weighting matrix is easier.

In fact, semi–blind DML (7.9.3) is built as an optimally weighted combination of the blind and TS criteria and SRM can be seen as an approximation of this weighted criterion. We have not tested the weighted version of the SF criterion [54]: the introduction of $\mathcal{N}$ may perhaps be justified by the blind weighting matrix.

M_K finite

We force the smallest eigenvalues of $\mathcal{U}_B^H \mathcal{W}_B^+ \mathcal{U}_B$ to 0, as we did in the PQML, SRM and SF case. As seen in the previous analysis, only the null space of the asymptotic value of $\mathcal{U}_B^H \mathcal{W}_B^+ \mathcal{U}_B$ counts; this null space remains unchanged with the weighting. So in the finite M_K case, the performance of the weighted combination is the same as for the unweighted combination.

7.12 Gaussian Methods

In the Gaussian case, semi–blind criteria can be similarly built as the sum of a blind and a TS based criterion as in the deterministic case. The decomposition of the data should be different though: indeed, not only the noise in the blind and

TS parts should be decoupled but also the signal contributions. However, we have verified in simulations that the correlations in the signal part can be ignored and the same decompositions as in the deterministic case can be adopted: see [35] for the semi–blind GML example.

We consider here the example of semi–blind covariance matching. The blind covariance matching [18], [55] method fits the theoretical expression for the covariance matrix, $R_L(\theta) = \sigma_a^2 \mathcal{T}_L(h)\mathcal{T}_L^H(h) + \sigma_v^2 I$, to an estimate of the covariance matrix, $\widehat{R}_L = \frac{1}{M-L}\sum_{k=0}^{M-L-1} \boldsymbol{Y}_L(k)\boldsymbol{Y}_L^H(k)$. Let $r(\theta) = \text{vec}\{R_L(\theta)\}$, $\hat{r} = \text{vec}\{\widehat{R}_L\}$, the covariance matching criterion can be written as:

$$\min_h \left(r(\theta) - \hat{r}\right)^H \mathcal{W}_B^{+} \left(r(\theta) - \hat{r}\right) \tag{7.12.1}$$

(we consider here σ_v^2 as known). $\mathcal{W}_B$ is a weighting matrix. A good approximation of the optimal weighting matrix is [35], [56]:

$$\mathcal{W}_B^{(1)} = \frac{1}{M_U - L} R_L^T \otimes R_L \quad . \tag{7.12.2}$$

The weighted semi–bind cost function then is:

$$(M_U - L)\left(r(\theta) - \hat{r}\right)^H \left[R_L^{-T} \otimes R_L^{-1}\right]\left(r(\theta) - \hat{r}\right) + \frac{1}{\sigma_v^2}\|\boldsymbol{Y}_{TS} - \mathcal{T}_{TS}(h)A_{TS}\|^2. \tag{7.12.3}$$

We can solve here the blind part by alternating minimizations as follows. We rewrite R_L as $R_L(\theta) = \sigma_a^2 \mathcal{T}_L(h_1)\mathcal{T}_L^H(h_2) + \sigma_v^2 I$. The blind covariance matching criterion can then be written as

$$\min_{h_1,h_2}\left(\text{vec}\{\left(\mathcal{T}_L(h_1)\mathcal{T}_L^H(h_2) + \sigma_v^2 I\right) - \widehat{R}_L\}\right)^H \mathcal{W}_B^{-1}\left(\text{vec}\{\left(\mathcal{T}_L(h_1)\mathcal{T}_L^H(h_2) + \sigma_v^2 I\right) - \widehat{R}_L\}\right) \tag{7.12.4}$$

We minimize alternatively between h_1 and h_2 (each minimization problem is quadratic) until convergence.

We show the NMSE given by the criterion (7.12.3) (which corresponds to (7.12.4) for the blind part with the (quasi–)optimal weighting $\mathcal{W}_B^{(1)}$ given in (7.12.2)) in figure 7.17; α scales the blind part of the criterion. We also show the performance corresponding to the weighting matrices $\mathcal{W}_B^{(2)} = \frac{1}{M_U-L}\text{diag}(R_L) \otimes \text{diag}(R_L)$ and $\mathcal{W}_B^{(3)} = \frac{1}{M_U-L}\left(\frac{\|h\|^2}{m}I\right) \otimes \left(\frac{\|h\|^2}{m}I\right) = \frac{1}{M_U-L}\frac{\|h\|^4}{m^2}I$ ($\frac{\|h\|^2}{m}$ is the mean value of the diagonal elements of R_L), as well as $\mathcal{W}_B^{(4)} = \frac{1}{M_U-L}I$. In this simulation, we can observe the quasi–optimality of the value $\alpha = 1$ and the relative insensitivity of the performance around $\alpha = 1$ for the weighted criteria (7.12.3) and the criteria corresponding to the weightings $\mathcal{W}_B^{(2)}$ and $\mathcal{W}_B^{(3)}$, this is not the case for $\mathcal{W}_B^{(4)}$. The weighting in (7.12.3) gives better results than the weightings $\mathcal{W}_B^{(2)}$ and $\mathcal{W}_B^{(3)}$ however.

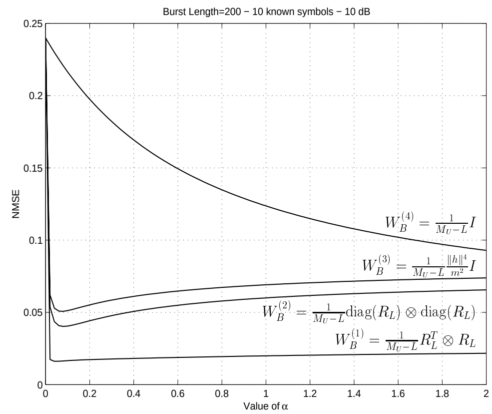

Figure 7.17. Semi–blind covariance matching built as a linear combination of blind covariance matching and training sequence based criteria.

7.13 Conclusion

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, which we discussed here, and also of methods that estimate the symbols directly such as the one in [32]. 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 symbols, *i.e.* a training sequence, low complexity 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 criterion. 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 with 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.

The last part of this chapter dealt with the construction of a semi–blind criterion built as a linear combination of a given blind criterion and a TS based criterion. We have proposed the examples of SRM, SF and covariance matching. Furthermore, a performance study of quadratic semi–blind criteria have been presented: we have seen conditions for the semi–blind criterion to be insensitive to the value of the weights in the linear combination and nearly optimal for the weight values that were proposed.

BIBLIOGRAPHY

[1] Raymond Steele, *Mobile Radio Communications*, Pentech Press Limited, 1992.

[2] 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.

[3] H. Trigui and D. Slock, "Performance Bounds for Cochannel Interference Cancellation within the Current GSM Standard," EURASIP Signal Processing, Elsevier Science, to appear, 1999.

[4] 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.

[5] 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.

[6] D.T.M. Slock and C.B. Papadias, "Blind Fractionally-Spaced Equalization Based on Cyclostationarity," in *Proc. Vehicular Technology Conf.*, Stockholm, Sweden, June 1994.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] L. Tong and Q. Zhao, "Blind Channel Estimation by Least Squares Smoothing," in *Proc. ICASSP 98 Conf.*, Seattle, USA, May 1998.

[17] 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.

[18] 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.

[19] 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.

[20] D. Hatzinakos and C.L. Nikias, ""Blind Equalization Based on Higher-Order Statistics (H.O.S.)"," in *Blind Deconvolution*, S. Haykin, Ed., pp. 181–258. Prentice–Hall, 1994.

[21] 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.

[22] 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.

[23] E. de Carvalho and D.T.M. Slock, "Blind and Semi–Blind FIR Multichannel Estimation: Identifiability Conditions," Submitted to IEEE Transactions on Signal Processing.

[24] B. Hochwald and A. Nehorai, "On Identifiability and Information-Regularity in Parameterized Normal Distributions," *Circuits, Systems ans Signal Processing*, vol. 16, no. 1, 1997.

[25] 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.

[26] 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.

[27] J. Manton, Y. Hua, Y. Zheng, and C. Zhang, "Semi-Blind Identification of Finite Impulse Response Channels," in *Proc. ICASSP 98 Conf.*, Seattle, USA, May 1998.

[28] J.H. Manton, W.D. Neumann, P.T. Norbury, and Y. Hua, "Algebraic Analysis of Linear Precoding for Blind Channel Identification," Submitted, for IEEE Transactions on Information Theory.

[29] 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.

[30] E. de Carvalho and D.T.M. Slock, "Cramér–Rao Bounds for Blind Multichannel Estimation," Submitted to IEEE Transactions on Information Theory.

[31] COST 207, "Digital Land Mobile Radio Communications," Tech. Rep., Office for Official Publications of the European Communities, 1989, Luxembourg, final report.

[32] A.-J. Van der Veen, "Algebraic Methods for Deterministic Blind Beamforming," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 1951–1968, Oct. 1998.

[33] D. Gesbert, P. Duhamel, and S. Mayrargue, "Blind least–squares criteria for joint sata/channel estimation," in *Proc. IEEE Digital Processing Workshop*, 1996.

[34] E. Pitié and P. Duhamel, "Bilinear Methods for Blind Channel Equalization: (No) Local Minimum Issue," in *Proc. ICASSP 98 Conf.*, Seattle, USA, May 1998.

[35] Elisabeth de Carvalho, *Blind and Semi–Blind Multichannel Estimation and Equalization for Wireless Communications*, Ph.D. thesis, ENST, Paris, 1999.

[36] 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.

[37] V. Radionov and S. Mayrargue, "Semi-blind approach to second order identification of SIMO-FIR channel driven by finite-alphabet sequence," in 13^{th} *international Conference on Digital Signal Processing*, Santorini, Greece, Jul 1997, vol. 1, pp. 115–118.

[38] M. Ghosh and C.L. Weber, ""Maximum-likelihood blind equalization"," *Optical Engineering*, vol. 31, pp. 1224–1228, June 1992.

[39] Dempster, N.M. Laird, and D.B. Rubin, "Maximum Likelihood from Incomplete Data via EM algorithm," *J. Royal Statisti. Soci.*, vol. 39, 1977.

[40] 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.

[41] 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.

[42] 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.

[43] Elisabeth de Carvalho and Dirk Slock, "Semi–Blind Maximum Likelihood Methods for FIR Multichannel Identification," In preparation, for IEEE Transactions on Signal Processing.

[44] M. Kristensson, M. Jansson, and B. Ottersten, "Modified IQML and Weighted Subspace Fitting without Eigendecomposition," *Signal Processing*, vol. 79, no. 1, pp. 29–44, Nov. 1999.

[45] 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.

[46] A. Gorokhov and P. Loubaton, "Semi–Blind Second–Order Identification of Convolutive Channels," in *Proc. ICASSP 97 Conf.*, Munich, Germany, April 1997.

[47] 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.

[48] 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.

[49] L.A. Baccala and S. Roy, "A New Time-Domain Blind Identification Method," *IEEE Signal Processing Letters*, vol. 1, no. 6, pp. 89–91, June 1994.

[50] G. Li and Z. Ding, "A Semi-Blind Channel Identification Method for GSM Receivers," in *Proc. ICASSP 98 Conf.*, Seattle, WA, May 12-15 1998.

[51] 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.

[52] V. Buchoux, O. Cappé, A. Gorokhov, and E. Moulines, "On the Performance of Semi-Blind Subspace Based Channel Estimation," Accepted for publication, IEEE Transactions on Signal Processing.

[53] 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.

[54] 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.

[55] M. Kristensson and B. Ottersten, "Asymptotic Comparison of two Blind Channel Identification Algorithms," in *Proc. SPAWC 97 Conf.*, Paris, France, April 1997.

[56] 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.