

Blind and Semi-Blind FIR Multichannel Estimation: (Global) Identifiability Conditions

Elisabeth de Carvalho, *Member, IEEE*, Dirk Slock, *Member, IEEE*

Abstract—Two channel estimation methods are often opposed: training sequence methods which use the information induced by known symbols and blind methods which use the information contained in the received signal, and possibly hypotheses on the input symbol statistics, but without integrating the information from known symbols, if present. Semi-blind methods combine both training sequence and blind information and are more powerful than the two methods separately. We investigate the identifiability conditions for blind and semi-blind FIR multichannel estimation in terms of channel characteristics, received data length, input symbol excitation modes as well as number of known symbols for semi-blind estimation. Two models corresponding to two different cases of a priori knowledge on the input symbols are studied: the deterministic model in which the unknown symbols are considered as unknown deterministic quantities and the Gaussian model in which they are considered as Gaussian random variables. This last model includes the methods using the second-order statistics of the received data. Semi-blind methods appear superior to blind and training sequence methods, and allow the estimation of any channel with only few known symbols. Furthermore, the Gaussian model appears more robust than the deterministic one as it leads to less demanding identifiability conditions.

Index Terms—channel estimation, blind, semi-blind, multichannel, SIMO, identifiability, antenna arrays, oversampling, space-time, spatiotemporal, Gaussian input.

I. INTRODUCTION

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

Manuscript received March 20, 2000; revised May 17, 2002. The work of Elisabeth de Carvalho was supported by Laboratoires d'Electronique Philips under contract Cifre 297/95. The work of Dirk Slock was supported in part by Laboratoires d'Electronique Philips under contract LEP 95FAR008. Eurécom's research is also partially supported by its industrial partners: Ascom, Swisscom, Thales Communications, ST Microelectronics, CEGETEL, Motorola, France Télécom, Bouygues Télécom, Hitachi Europe Ltd., and Texas Instruments. The associate editor coordinating the review of this paper and approving it for publication was Prof. Alex Kot.

E. de Carvalho was with Eurecom. She is now with NewLogic Technologies, BP 125, 06903 Sophia Antipolis Cdex, France (e-mail: elisabeth.de-carvalho@newlogic.fr).

D. Slock is with the Eurecom Institute, BP 193, 06904 Sophia Antipolis Cdx, France (e-mail: dirk.slock@eurecom.fr).

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 can be occupied: in GSM, for example, 20% of the bits in a burst are used for training. Furthermore, in certain communication systems, training sequences are not available or exploitable, when synchronization between the receiver and the transmitter is not possible.

Blind equalization techniques allow the estimation of the channel or the equalizer based only on the received signal without any training symbols. The first wave of blind techniques were based on Higher-Order Statistics (HOS). The introduction of multichannels, or SIMO (Single Input Multiple Output) models where a single input symbol stream is transmitted through multiple linear channels and sampled at the symbol rate, has given rise to a whole bunch of new blind estimation techniques that do not need higher order-statistics. The most popular SOS estimation techniques suffer from a lack of robustness: channels must satisfy diversity conditions and some blind SOS (Second OS) methods can fail when the channel length is overestimated. Furthermore, the blind techniques leave an indeterminacy in the channel or the symbols, a scale or 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 is true also for training sequence based methods, especially when the sequence is too short to estimate the channel parameters. Semi-blind techniques are a solution to overcome these problems.

We assume a transmission by burst, *i.e.* the data is divided 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 some known symbols used for synchronization or as guard intervals, like in the GSM or the DECT burst. 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

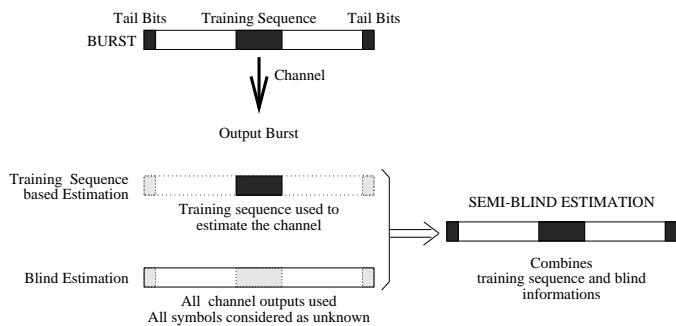


Fig. 1. Semi-Blind Principle: example of a GSM burst.

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 informations (see Fig. 1) and exploit the positive aspects of both techniques.

Semi-blind techniques, because they incorporate the information of known symbols, avoid the possible pitfalls of blind methods and with only a few known symbols, any channel, single or multiple, becomes identifiable. Furthermore, exploiting the blind information in addition to the known symbols, allows to estimate longer channel impulse responses than possible with a certain training sequence length, a feature that is of interest for the application of mobile communications in mountainous areas. For methods based on the second-order moments of the data (which we will call Gaussian methods), one known symbol is sufficient to make any channel identifiable. In addition, it allows to use shorter training sequences for a given channel length and desired estimation quality, compared to a training approach. Apart from these robustness considerations, semi-blind techniques appear also very interesting from a performance point of view, as their performance is superior to that of training sequence and 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 this paper, we study the identifiability conditions for semi-blind estimation. A treatment of the estimation performance in the form of the Cramér-Rao Bound (CRB) and a performance comparison with blind and TS estimation can be found in [2]. We will concentrate on two different models corresponding to two different forms of a priori knowledge on the unknown input symbols. The first model, the deterministic model, does not exploit any knowledge and considers the unknown input symbols as deterministic quantities. In the second model, the Gaussian model, the unknown input symbols are considered as (usually white) Gaussian random variables with known parameters. The Gaussian model could appear absurd as the input symbols are in fact discrete-valued. As a first response, it should be noticed that the Gaussian model includes the methods that are based on the first and second-order moments of the received signal and hence of the input symbols. Blind methods that consider the input symbols as i.i.d. random variables and that are based on the second-order moments of the received signal, like the prediction [3] or

the covariance matching methods [4], belong to the Gaussian category. More arguments for considering a Gaussian model are given below and in section III-B. In section III-C, the two models are put in perspective and compared to the methods exploiting the discrete alphabet nature of the input symbols.

Parameters will be called identifiable if they are determined uniquely by the probability distribution of the data. In the blind case, the definition differs slightly as blind estimation leaves some indeterminacies: the deterministic model can estimate the channel up to a scale factor and the Gaussian model up to a phase factor. Parameters will be called identifiable if they are identifiable up to these blind indeterminacies. Identifiability conditions for blind and semi-blind channel estimation for both models are given in terms of characteristics of the channel, received data burst length, input symbol excitation modes, as well as number of known symbols for semi-blind estimation.

Blind identifiability conditions especially for the deterministic methods can already be found in the literature [5], [6] but never in a complete way. They are often only formulated in terms of channel characteristics, the well known condition being for the channel to be irreducible, *i.e.* with no zeros. As far as the burst length and input symbol excitation modes are concerned, only sufficient conditions are given. For blind deterministic methods, we give necessary and sufficient conditions. It has to be mentioned that necessary and sufficient conditions appear in [5], but including a superfluous condition.

Let us stress that, unlike deterministic methods, blind Gaussian methods can estimate the zeros of a channel: they only cannot determine if a zero is minimum- or maximum-phase. This fact is not widely known and the irreducibility condition for Gaussian methods is sometimes imposed unnecessarily. Blind Gaussian methods appear thus more robust than blind deterministic methods.

We show that semi-blind methods can identify any channel such as irreducible channels, multichannels with zeros, monochannels, and we determine the number of known symbols required. Only a few known symbols are required, fewer than required for pure training sequence based estimation. For an irreducible channel, only 1 known symbol is sufficient in the deterministic model. Again the Gaussian model appears more robust as only 1 known symbol (not located at the edges of the burst) is sufficient for any type of channel! The semi-blind results are given for known symbols that are grouped; a small discussion is also provided for the case when they are dispersed over the burst.

We shall use the following notation and acronyms:

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

CRB	Cramer–Rao Bound
FIM	Fisher Information Matrix
ML	Maximum Likelihood

II. THE MULTICHANNEL MODEL

We consider here linear modulation of a single user signal over a linear channel with additive noise. The overall channel impulse response is modeled as FIR. We consider here the FIR multichannel case [7], [8], [9] which can arise in a number of ways as indicated in Fig. 2. The discrete-time vector received signal at symbol rate can be written as:

$$\mathbf{y}(k) = \sum_{i=0}^{N-1} \mathbf{h}(i)a(k-i) + \mathbf{v}(k) \quad (1)$$

where the $a(k)$ are the transmitted symbols, $\mathbf{y}(k) = [y_1(k) \cdots y_m(k)]^T$ and similarly for $\mathbf{v}(k)$ and the N vector channel impulse response samples $\mathbf{h}(k)$. m is the number of subchannels. In the case of real symbols [10], [11], we can assume that all quantities are real (see Fig. 2). In all cases, we can write the channel input-output relationship as

$$\begin{aligned} \mathbf{y}(k) &= \mathbf{H}A(k) + \mathbf{v}(k), \quad \mathbf{H} = [\mathbf{h}(0) \cdots \mathbf{h}(N-1)], \\ A(k) &= [a(k) \cdots a(k-N+1)]^T. \end{aligned} \quad (2)$$

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

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

$$\mathbf{Y}_M(k) = \mathcal{T}_M(h)A_M(k) + \mathbf{V}_M(k) \quad (3)$$

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

$$\mathcal{T}_M(h) = \begin{bmatrix} \mathbf{h}(0) & \cdots & \mathbf{h}(N-1) & 0 & \cdots & 0 \\ 0 & \mathbf{h}(0) & \cdots & \mathbf{h}(N-1) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{h}(0) & \cdots & \mathbf{h}(N-1) \end{bmatrix} \quad (4)$$

and

$$h = [\mathbf{h}^T(0) \cdots \mathbf{h}^T(N-1)]^T. \quad (5)$$

The channel length is assumed to be N which implies $\mathbf{h}(0) \neq 0$ and $\mathbf{h}(N-1) \neq 0$ whereas the impulse response is zero outside of the indicated range. We shall simplify the notation in (3) with $k = M-1$ to

$$\mathbf{Y} = \mathcal{T}(h)A + \mathbf{V}. \quad (6)$$

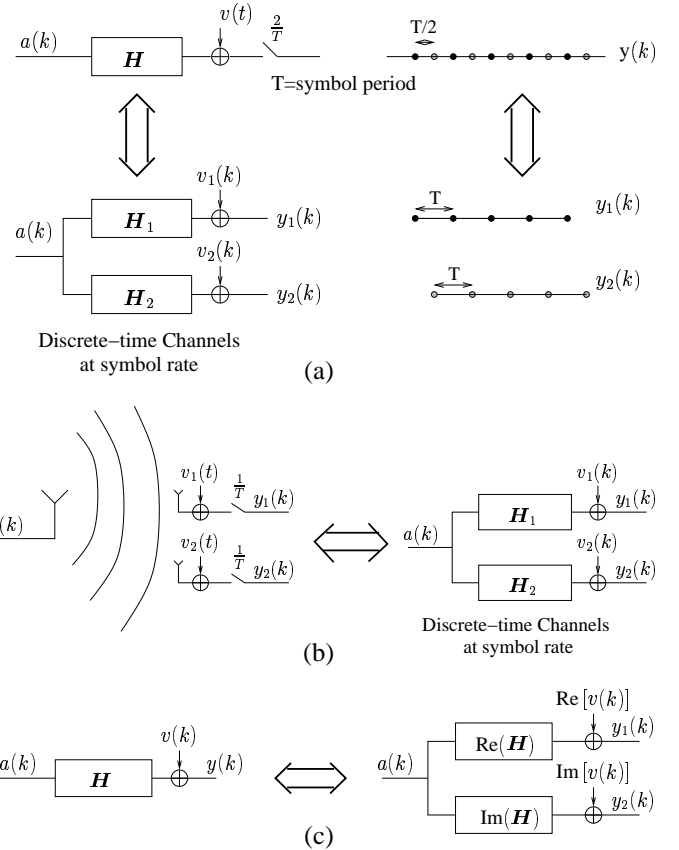


Fig. 2. Multichannel model: case of (a) oversampling, (b) multiple antennas and (c) separation of inphase and quadrature components when the input symbols are real. Example of a multichannel with 2 subchannels.

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

$$\mathcal{T}(h)A = A_m h \quad (7)$$

where: $A_m = A_1 \otimes I_m$,

$$A_1 = \begin{bmatrix} a(M-1) & a(M-2) & \cdots & a(M-N) \\ a(M-2) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ a(0) & \cdots & \cdots & a(-N+1) \end{bmatrix}. \quad (8)$$

Sometimes, we will simplify A_m to A .

Semi-Blind Model The vector of input symbols can be written as: $A = \mathcal{P} \begin{bmatrix} A_K \\ A_U \end{bmatrix}$ where A_K are the M_K known symbols and A_U the $M_U = M+N-1-M_K$ unknown symbols. The known symbols 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 both parts in the channel output as $\mathcal{T}(h)A = \mathcal{T}_K(h)A_K + \mathcal{T}_U(h)A_U$.

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

$$\mathbf{H}(z) = \mathbf{H}_I(z)\mathbf{H}_c(z), \quad (9)$$

where $\mathbf{H}_I(z)$ of length N_I is irreducible and $H_c(z)$ of length $N_c = N - N_I + 1$ is a monochannel for which we assume $H_c(\infty) = h_c(0) = 1$ (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 $H_c(z)$ is minimum-phase.

Minimum Zero-Forcing (ZF) Equalizer Length, Effective Number of Channels The Bezout identity states that for an FIR irreducible channel, FIR ZF equalizers exist [12]. The minimum length for such a FIR ZF equalizer is

$$\underline{M} = \min \{M : \mathcal{T}_M(h) \text{ has full column rank}\} . \quad (10)$$

One may note that $\mathcal{T}_M(h)$ has full column rank for $M \geq \underline{M}$. In [13], it is shown that if the mN elements of \mathbf{H} are considered random, more precisely independently distributed with a continuous distribution, then

$$\underline{M} = \left\lceil \frac{N-1}{m-1} \right\rceil \quad \text{with probability 1,} \quad (11)$$

and $\underline{M} = 1$ when $N = 1$. In this case, the channel is irreducible w.p.1. One could consider other (perhaps more realistic) channel models. Consider e.g. a multipath channel with K paths in which the multichannel aspect comes from m antennas. Without elaborating the details, it is possible to introduce an effective number of channels m_e which in this case would equal (w.p. 1)

$$m_e = \text{rank}(\mathbf{H}) = \min \{m, N, K\} . \quad (12)$$

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 (11), 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} [12] could then be applied.

In general, for an irreducible channel, $\underline{M} \leq N-1$ [14] in which the upper bound would correspond to $m_e = 2$. Note that $m_e = 1$ corresponds to a reducible channel (in which case $\underline{M} = \infty$).

We will not define here the notion of input signal excitation modes, see for example [6], [5] for a complete definition.

III. DETERMINISTIC AND GAUSSIAN MODELS

A. Deterministic Model

In the deterministic model, both unknown input symbols and channel coefficients are assumed to be deterministic quantities. The data have a Gaussian distribution with:

$$m_Y(\theta) = \mathcal{T}_K(h)A_K + \mathcal{T}_U(h)A_U \quad C_{YY}(\theta) = \sigma_v^2 I. \quad (13)$$

Deterministic methods proceed either to the joint estimation of h and A_U or to the estimation of h with A_U considered as nuisance parameter (the estimation of h and A_U is decoupled

from the estimation of σ_v^2). The estimation is based on the received signal structure. Many blind algorithms fall into this category, among which we find:

- The least squares approach based directly on the received signal by Gürelli *et al* [15], also called Cross-Relation (CR) method [5] or Subchannel Response Matching (SRM) [16], and used for initializing deterministic ML in [6].
- The (unweighted) subspace fitting approaches initiated by Moulines *et al* [17].
- Blocking equalizers determined by linear prediction [8], [12], [18].
- The deterministic ML approaches in their blind version [8], [19], [6], [20], and in their semi-blind version [21], [22], [23].
- Two-sided Linear Prediction or Least-Squares Smoothing [24], [25], [26].

B. Gaussian Model

In the Gaussian model, the unknown input symbols are considered as i.i.d. Gaussian random variables of mean 0 and variance σ_a^2 , and the known symbols as deterministic (of mean A_K and variance 0). This model may appear inappropriate as the input symbols are in fact discrete-valued. We elaborate here further on the motivation for introducing this Gaussian model. The next section will present the advantages of the Gaussian methods.

The purpose of the Gaussian model is to take into account the first and second-order moments of the data, which appear to play a predominant role in the multichannel context and are here:

$$m_Y(\theta) = \mathcal{T}_K(h)A_K, \quad C_{YY}(\theta) = \sigma_a^2 \mathcal{T}_U(h) \mathcal{T}_U^H(h) + \sigma_v^2 I. \quad (14)$$

Already existing blind methods that base channel estimation on the second-order moments of the data, and in which the input symbols are considered as i.i.d. random variables, can be classified into the Gaussian category: certain prediction approaches [3] or the covariance matching method [4] belong to the Gaussian category and give better performance than the deterministic approaches (even when the symbols are actually discrete valued).

These methods in fact only require the second-order statistics and not the complete distribution. The Gaussian assumption is intended for ML approaches for which knowledge of the complete distribution is required. The Gaussian distribution is the simplest distribution, leading to simple derivations and allowing to incorporate the first and second-order moments of the data: $\mathbf{Y} \sim \mathcal{N}(m_Y(\theta), C_{YY}(\theta))$, the Gaussian hypothesis for the symbols leads to a Gaussian distribution for \mathbf{Y} .

A semi-blind ML method based on this model was proposed in [21], [27] and shown to give better performance than ML based on the deterministic model [22]. The blind optimally weighted covariance matching method [28] based on an asymptotically large covariance matrix is shown to have the same asymptotic (in the number of data) performance than blind Gaussian ML [29]. The Gaussian hypothesis for the sources is also regularly used in direction of arrival finding

and the associated ML proved to give better performance than the deterministic ML methods [30].

In the course of this work, we became aware of [31], [32] in which a semi-blind Gaussian ML method and corresponding CRBs have been studied. The modeling of the training sequence information in [31] is inappropriate though: instead of the training sequence, the information considered is the training sequence times an unknown zero-mean unit-variance normal variable.

Unlike in the deterministic case, the input symbols in the Gaussian model are no longer nuisance parameters for the estimation of h . The parameters to be estimated jointly are the channel coefficients and the noise variance.

The Gaussian model has been formulated here in the special case of semi-blind estimation, in which case symbols are perfectly known or unknown. It allows however a general formulation in which any prior knowledge on the input symbols at the level of first and second-order moments can be incorporated. In this case, the received signal is $\mathbf{Y} = \mathcal{T}(h)A + \mathbf{V}$ with $\mathbf{V} \sim \mathcal{N}(0, C_{VV})$ independent of $A \sim \mathcal{N}(A^o, C_{AA})$. A^o , the prior mean for the symbols, represents any prior knowledge on the symbols (and could be something else than the knowledge of symbols itself) and the covariance C_{AA} captures the remaining uncertainty.

C. Classification of Channel Identification Methods According to the A Priori Knowledge on the Unknown Input Symbols Exploited

The purpose of this section is to put both deterministic and Gaussian methods in perspective: see their respective interest and compare them to the methods exploiting the finite alphabet of the input symbols. We shall first concentrate on the blind methods: they can be classified according to the increasing a priori knowledge on the unknown input symbols exploited as follows (see Fig. 3):

- 1) No information exploited: the deterministic methods.
- 2) Second-order statistics: the Gaussian methods.
- 3) Higher-order statistics.
- 4) Finite symbol constellation alphabet: the Finite Alphabet (FA) methods (see [33], [34], [35], [36] for example, and [37] for associated semi-blind methods).
- 5) Complete Symbol Distribution: stochastic methods [38], [39].

Certain methods are situated in between the categories above; e.g. optimally weighted subspace fitting is situated in between the deterministic (subspace fitting) and Gaussian (optimal weighting) methods. In the higher-order statistics category, we can also consider techniques based on a partial exploitation of the finite alphabet, such as constant modulus techniques.

1) *Identification Indeterminacies*: The different methods are classified here according to decreasingly severe identification indeterminacies. As an example, for complex input constellations, blind deterministic methods can identify the channel up to a complex scale factor only, $\hat{h} = \alpha h^o$, with $\alpha \in \mathbb{C}$; in the Gaussian case, the channel can be identified up to a phase factor $\hat{h} = e^{j\varphi} h^o$, with $\varphi \in \mathbb{R}$; FA and stochastic methods can identify the channel up to a discrete-valued phase

factor, $\hat{h} = e^{j\varphi} h^o$, with φ taking a finite number of discrete values (depending on the symmetry properties of the symbol constellation).

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

3) *Performance*.: The above classification respects also the order of increasing performance. The FA methods are particularly powerful: indeed, a performance bound for FA methods corresponds to the case in which all the input symbols would act as training sequence, as discussed in [2]. In [36], it was explicitly shown that the blind deterministic CRB under the constraint that the unknown symbols belong to a FA (case 4 in Fig. 3) yields the TS CRB (as if all unknown symbols form a training sequence).

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

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

The above discussion was about blind estimation. The associated semi-blind versions inherit from their blind counterpart properties, advantages or disadvantages. It should be noted that the performance difference between the three classes of methods gets smaller as more and more symbols are known. Performance differences are most visible in the case of blind methods, especially for ill-conditioned channels. Semi-blind FA methods would be the ideal solution if a good initialization quality was not so important. When the training sequence is too short to give a good channel initialization or when blind methods fail, the initialization could be not good enough for FA methods to work directly. One can instead proceed in smaller steps by first using a semi-blind deterministic method to initialize a semi-blind Gaussian method, which could in

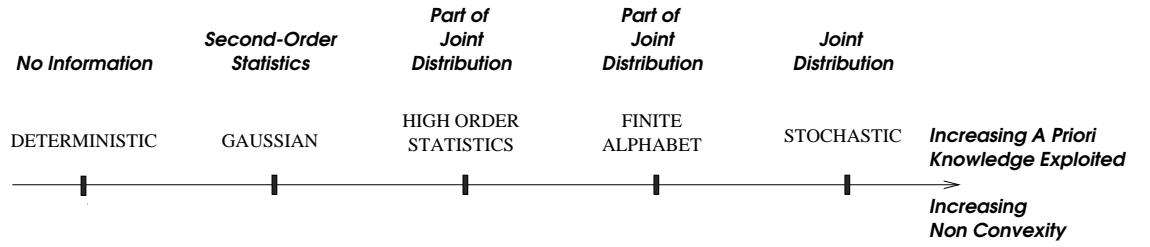


Fig. 3. Classification of channel identification methods according to the a priori knowledge on the unknown input symbols exploited.

turn be used to initialize a semi-blind FA method.

IV. IDENTIFIABILITY DEFINITION

Let θ be the parameter to be estimated and \mathbf{Y} the observations. In the regular cases (*i.e.* in the non blind cases), θ is called identifiable if [40]:

$$\forall \mathbf{Y}, f(\mathbf{Y}|\theta) = f(\mathbf{Y}|\theta') \Rightarrow \theta = \theta'. \quad (15)$$

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. The identifiability condition (15) will be for θ to equal θ' up to the blind indeterminacy.

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

V. IDENTIFIABILITY IN THE DETERMINISTIC MODEL

In the deterministic model, $\mathbf{Y} \sim \mathcal{N}(\mathcal{T}(h)A, \sigma_v^2 I)$ 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 . A_U and h are identifiable if:

$$\begin{aligned} \mathcal{T}(h)A &= \mathcal{T}(h')A' \Rightarrow \\ \begin{cases} A_U = A'_U \text{ and } h = h' & \text{for semi-blind and TS} \\ A = \frac{1}{\alpha}A' \text{ and } h = \alpha h' & \text{for blind estimation} \end{cases} \end{aligned} \quad (16)$$

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

A. TS Based Channel Identifiability

The TS case could be considered as a limiting case of either deterministic or Gaussian models in which received signal containing unknown symbols is not taken into account. We shall discuss the TS case here under the deterministic model and not duplicate this discussion for the Gaussian model.

We recall here the identifiability conditions for TS based channel estimation. From (7), $\mathcal{T}(h)A = Ah$: h is determined uniquely if and only if 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 $M \geq N$.*
- (ii) *Number of input symbol modes $\geq N$.*

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

B. Blind Channel Identifiability

The deterministic blind identifiability definition (16) corresponds to what is called strict identifiability in [41]. The authors of [5], [6] define identifiability based on the Cross-Relation (CR) method: a channel is said CR-identifiable if the channel can be identified uniquely (up to a scale factor) by the noise-free CR method. In [6], identifiability is based on the (complex) FIM matrix: a channel is said identifiable if the FIM has exactly one singularity. In [6], [41], those three identifiability forms were found to be equivalent. [5], [6] give sufficient conditions, and necessary conditions separately for the channel, the burst length and the symbol modes for the CR-identifiability (extended to the FIM and strict identifiability in [6], [41]). In [5], necessary and sufficient conditions on the channel and the modes (but not on the burst length though) are also given and a coupled relation between the channel and the input symbols modes appears, which usefulness is not guaranteed.

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

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

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

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

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

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

(ii): Condition (ii) says that the number of equations ($= mM$)

should be greater than the number of unknowns: $Nm-1$ unknowns for \mathbf{H} , $M+N-1$ for the unknown symbols.

(iii): A proof of condition (iii) can be found in [41]. \square

Sufficient conditions [DetB] *In the deterministic model, the m -channel $\mathbf{H}(z)$ and the input symbols A are blindly identifiable if*

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

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

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

Proof: see Appendix I. \square

These conditions express the fact that one should have enough data with the right properties to be able to completely describe the signal (or noise) subspace. The proof is based on subspace fitting results. An alternative proof based on linear prediction and blocking equalizers has been given in [42].

Note that the sufficient conditions above are sufficient conditions for the subspace fitting method. A priori, sufficient conditions for identifiability as in (15) could be weaker than the sufficient conditions for the subspace fitting method. These conditions appear to be sufficient for all the deterministic methods listed in section III-C except for SRM [6].

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

C. Semi-Blind Channel Identifiability

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

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

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

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

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

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

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

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

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

Proof: See Appendix III. \square

For an irreducible channel, 1 known symbol is sufficient. For a monochannel, $2N-1$ grouped known symbol are sufficient. If $2N-1$ grouped known symbols containing

N independent modes are available, condition (ii) becomes superfluous.

Identifiability is also guaranteed with the same number of known symbols in the case where the known symbols are not grouped. We shall omit the details of the proof here, but to give some elements of the proof, it can be shown that FIM regularity holds under conditions very similar to [DetSB], which implies local identifiability. In order to have global identifiability, the burst length should be larger however.

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

D. Semi-Blind Robustness to Channel Length Overestimation

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

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

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

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

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

Proof: See Appendix IV. \square

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

VI. IDENTIFIABILITY IN THE GAUSSIAN MODEL

A. Gaussian Model

The parameters to be estimated are the channel coefficients and the noise variance: $\theta = [h^H \sigma_v^2]^H$. 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.

B. Blind Channel Identifiability

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

$$C_{YY}(h, \sigma_v^2) = C_{YY}(h', \sigma_v'^2) \Rightarrow h' = e^{j\varphi} h, \sigma_v'^2 = \sigma_v^2. \quad (17)$$

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

Blind identifiability conditions based on the second-order statistics of the noise-free outputs of a FIR multichannel

driven by a white stationary input sequence were given in [43], [44]. Only conditions on the channel are given: in [43], [44], a channel is said blindly identifiable up to a phase factor if the channel is irreducible. In fact, it is possible to identify blindly the channel based on the second-order moments even for a reducible channel, it is only not possible to determine if the zeros are minimum or maximum-phase. We give conditions on the channel and the correlation sequence length. (The conditions on the input symbols are that they are white).

1) *Irreducible Channel:*

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

- (i) $\mathbf{H}(z)$ is irreducible.
- (ii) Burst length $M \geq \underline{M} + 1$

Proof: When condition (ii) is verified, $\mathcal{T}_U(h)$ is (strictly) tall and σ_v^2 can then be uniquely identified as the minimal eigenvalue of $C_{YY}(\theta)$. $\mathbf{H}(z)$ can then be identified up to a phase factor from the denoised covariance matrix $C_{YY}(\theta) - \sigma_v^2 I$ by linear prediction [9]: under conditions (i) and (ii), one can find $P(z)$, the multivariate prediction filter of order \underline{M} and $\mathbf{h}(0)$ (the first coefficient of \mathbf{H}) up to a phase factor from the denoised covariance matrix, and they are related to $\mathbf{H}(z)$ via the relationship:

$$P(z)\mathbf{H}(z) = \mathbf{h}(0). \quad (18)$$

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

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

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

2) *Reducible Channel:* Let $\mathbf{H}(z)$ be a reducible channel: $\mathbf{H}(z) = \mathbf{H}_I(z)\mathbf{H}_c(z)$.

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

- (i) $H_c(z)$ is minimum-phase.
- (ii) $M \geq \max(\underline{M}_I + 1, N_c - N_I + 1)$.

Proof: Under condition (ii), $\mathcal{T}(h_I)$ is strictly tall and σ_v^2 can be identified as the minimal eigenvalue of $C_{YY}(\theta)$. The irreducible part \mathbf{H}_I can be identified up to a scale factor thanks to the deterministic method described in section V-B [12] provided that $M \geq \underline{M}_I + 1$: let $h'_I = \alpha h_I$ be this estimate of h_I . $(\mathcal{T}^H(h'_I)\mathcal{T}(h'_I))^{-1} \mathcal{T}^H(h'_I) [C_{YY}(\theta) - \sigma_v^2 I] \mathcal{T}(h'_I) (\mathcal{T}^H(h'_I)\mathcal{T}(h'_I))^{-1} = \sigma_a^2 \mathcal{T}(\alpha^{-1} h_c) \mathcal{T}^H(\alpha^{-*} h_c)$. $\alpha^{-1} \mathbf{H}_c(z)$ can now be identified up to a phase factor by spectral factorization provided that $\alpha H_c(z)$ or hence $H_c(z)$ is minimum-phase and $\mathcal{T}(h_c)\mathcal{T}^H(h_c)$ contains the N_c non-zero correlations, i.e. $M + N_I - 1 \geq N_c$ or $M \geq N_c - N_I + 1$. \square

3) *Monochannel Case:* In the monochannel case, the noise variance σ_v^2 cannot be estimated and so 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 .

C. *Semi-Blind Channel Identifiability*

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

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

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

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

2) *Identifiability for an Irreducible Channel:*

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

- (i) $\mathbf{H}(z)$ is irreducible.
- (ii) At least 1 non-zero known symbol (located anywhere) appears.

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

VII. CONCLUSIONS

In this paper, we have introduced the principle of semi-blind FIR channel estimation (see also [32], [37] where the semi-blind concept has been introduced independently) and pointed out its advantages w.r.t. training sequence and blind estimation. Two models were studied, the deterministic and Gaussian models, which include numerous already existing blind and semi-blind methods. Their respective interest was presented and compared to methods exploiting the finite alphabet of the input symbols. Identifiability conditions for the two models

were given in terms of channel characteristics, burst length, input symbol excitation modes and number of known symbols for semi-blind estimation in the case of grouped known symbols. The semi-blind approach appears more robust than blind estimation, as it allows the estimation of any channel with only a few known symbols. In the deterministic case, 1 known symbol is required for an irreducible channel, $2N_c - 1$ for a reducible channel and $2N - 1$ known symbols for a monochannel (notations of section II). The Gaussian model only requires 1 known symbol (not located at the edges of the burst) and is hence more robust than the deterministic model. This Gaussian model appears especially interesting in the multiuser context (SDMA) [45].

APPENDIX I PROOF OF SUFFICIENT CONDITIONS [DET B]

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

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

$$\mathcal{X} = \begin{bmatrix} \mathbf{x}(M-1) & \cdots & \mathbf{x}(\underline{M}) \\ \vdots & \ddots & \vdots \\ \mathbf{x}(M-\underline{M}-1) & \cdots & \mathbf{x}(0) \end{bmatrix}, \quad (19)$$

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

and related as

$$\mathcal{X} = \mathcal{T}_{\underline{M}+1}(h)\mathcal{A}. \quad (20)$$

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

$$P_{\mathcal{X}}^{\perp} = P_{\mathcal{T}_{\underline{M}+1}(h)}^{\perp} \quad (21)$$

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

$$P_{\mathcal{X}}^{\perp} \mathcal{T}_{\underline{M}+1}(\hat{h})\hat{A} = 0 \Rightarrow P_{\mathcal{T}_{\underline{M}+1}(h)}^{\perp} \mathcal{T}_{\underline{M}+1}(\hat{h}) = 0 \Leftrightarrow \text{range} \{ \mathcal{T}_{\underline{M}+1}(\hat{h}) \} \subset \text{range} \{ \mathcal{T}_{\underline{M}+1}(h) \}. \quad (22)$$

Now, in Appendix II¹ (with $M = \underline{M}+1$ here) it is shown that this implies $\hat{h} = \alpha h$ where α is some complex scalar. Now also A can be estimated up to a scale factor:

$\hat{A} = \left(\mathcal{T}^H(\hat{h})\mathcal{T}(\hat{h}) \right)^{-1} \mathcal{T}^H(\hat{h})\mathbf{X} = A/\alpha$ (i.e. the output of the MMSE zero-forcing equalizer built from \hat{h}).

APPENDIX II

CHANNEL IDENTIFIABILITY FROM THE SIGNAL SUBSPACE

Theorem 1 (Subspace Fitting): Let h and h' be causal channel impulse responses of length N and N' respectively. If h is irreducible, then for $M > \underline{M}$

$$\text{range} \{ \mathcal{T}_M(h') \} \subset \text{range} \{ \mathcal{T}_M(h) \} \Rightarrow \begin{cases} \mathbf{H}'(z) = \mathbf{H}(z)\alpha(z) & , N' \geq N \\ h' = 0 & , N' < N \end{cases} \quad (23)$$

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

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

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

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

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

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

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

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

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

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

APPENDIX III

PROOF OF SUFFICIENT CONDITIONS [DETSB]

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

Consider the $m(\underline{M}_I+1) \times (M-\underline{M}_I)$ data matrix $\mathcal{X} = \mathcal{T}_{\underline{M}_I+1}(h_I)\mathcal{T}_{\underline{M}_I+N_I}(h_c)\mathcal{A}$. Then $P_{\mathcal{X}} = P_{\mathcal{T}_{\underline{M}_I+1}(h_I)}$ if and

¹The proof in Appendix II is a shorter alternative to the proof in Appendix A of [17], generalized to an extended range of signal space dimension M .

only if $\mathcal{T}_{\underline{M}_I+N_I}(h_c)\mathcal{A}$ has full row rank. Condition (i) expresses that the number of columns of this last quantity should be greater than its number of rows, plus the fact that in general $M \geq M_K - N + 1$, which gets combined with condition (iii). Let p be the number of modes of A (which are assumed to be unrepeated, the extension to the case of higher multiplicity being straightforward [6]): $a(k) = \sum_{i=1}^p \alpha_i z_i^k$. It can be shown that \mathcal{A} can be decomposed as

$$\mathcal{A} = \mathcal{M}_1 \mathcal{M}_2 \mathcal{M}_3 = \begin{bmatrix} 1 & \cdots & 1 \\ z_1^{-1} & \cdots & z_p^{-1} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ z_1^{-(\underline{M}_I+N-1)} & \cdots & z_p^{-(\underline{M}_I+N-1)} \end{bmatrix} \begin{bmatrix} \alpha_1 & 0 & \cdots & \cdots & 0 \\ 0 & \alpha_2 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \alpha_p \end{bmatrix} \begin{bmatrix} z_1^{M-1} & z_1^{M-2} & \cdots & z_1^{\underline{M}_I} \\ z_2^{M-1} & z_2^{M-2} & \cdots & z_2^{\underline{M}_I} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ z_p^{M-1} & z_p^{M-2} & \cdots & z_p^{\underline{M}_I} \end{bmatrix} \quad (24)$$

so that we can write

$$\mathcal{T}(h_c)\mathcal{A} = \mathcal{B}_1 \mathcal{B}_2 \mathcal{M}_2 \mathcal{M}_3 \quad \text{with} \quad (25)$$

$$\mathcal{T}(h_c)\mathcal{M}_1 = \mathcal{B}_1 \mathcal{B}_2 = \begin{bmatrix} 1 & \cdots & 1 \\ z_1^{-1} & \cdots & z_p^{-1} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ z_1^{-\underline{M}_I-N_I+1} & \cdots & z_p^{-\underline{M}_I-N_I+1} \end{bmatrix} \begin{bmatrix} \mathbf{H}_c(z_1) & 0 & \cdots & \cdots & 0 \\ 0 & \mathbf{H}_c(z_2) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \mathbf{H}_c(z_p) \end{bmatrix}. \quad (26)$$

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

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

APPENDIX IV

PROOF OF SUFFICIENT CONDITIONS [DETSBR]

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

with $N'_I + N'_c - 1 = N'$. To the irreducible h'_I corresponds a minimum ZF equalizer length \underline{M}'_I . Consider

$$\mathcal{X} = \mathcal{T}_{\underline{M}'_I+1}(h_I)\mathcal{T}(h_c)\mathcal{A} = \mathcal{T}_{\underline{M}'_I+1}(h'_I)\mathcal{T}(h'_c)A'. \quad (27)$$

Assume for a moment that the conditions are satisfied for $\mathcal{T}(h_c)\mathcal{A}$ to have full row rank; we shall see below what this entails. Then (27) implies

$$\text{range}\{\mathcal{X}\} = \text{range}\{\mathcal{T}_{\underline{M}'_I+1}(h_I)\} \subset \text{range}\{\mathcal{T}_{\underline{M}'_I+1}(h'_I)\}. \quad (28)$$

According to Appendix I, this implies

$$\begin{cases} h_I = 0 & , N'_I > N_I \\ h'_I = \alpha h_I & , N'_I = N_I \\ h_I \text{ reducible} & , N'_I < N_I \end{cases} \quad (29)$$

Hence necessarily $h'_I = \alpha h_I$ and $\underline{M}'_I = \underline{M}_I$ so that $\mathcal{T}(h_c)\mathcal{A}$ has full row rank under conditions (i) – (ii). Since $\mathcal{T}_{\underline{M}_I}(h_I)$ has full column rank, (27) implies $\mathcal{T}(h_c)\mathcal{A} = \alpha \mathcal{T}(h'_c)A'$. Let's denote $h_d = [h_c^T \ 0 \cdots 0]^T$ and $A_d = [A^T \ 0 \cdots 0]^T$, h_d and A_d being of the same length as h'_c and A' respectively. Then we can also write

$$\mathcal{T}(h_d)A_d = \mathcal{T}(\alpha h'_c)A' \quad (30)$$

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

REFERENCES

- [1] R. Steele, *Mobile Radio Communications*. Pentech Press Limited, 1992.
- [2] E. de Carvalho and D. Slock, "Cramér-Rao Bounds for Semi-blind, Blind and Training Sequence based Channel Estimation," in *Proc. SPAWC 97 Conf.*, Paris, France, April 1997.
- [3] 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.
- [4] G. Giannakis and S. 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] D. Slock, "Blind Fractionally-Spaced Equalization, Perfect-Reconstruction Filter Banks and Multichannel Linear Prediction," in *Proc. ICASSP 94 Conf.*, Adelaide, Australia, April 1994.
- [9] D. Slock and C. Papadakis, "Blind Fractionally-Spaced Equalization Based on Cyclostationarity," in *Proc. Vehicular Technology Conf.*, Stockholm, Sweden, June 1994.
- [10] M. Kristensson, B. Ottersten, and D. Slock, "Blind Subspace Identification of a BPSK Communication Channel," in *Proc. of the 30th Asilomar Conference on Signals, Systems & Computers*, Pacific Grove, CA, Nov. 1996.
- [11] A. van der Veen, "Analytical Method for Blind Binary Signal Separation," *IEEE Transactions on Signal Processing*, vol. 45, no. 4, pp. 1078–1082, April 1997.
- [12] D. Slock, "Blind Joint Equalization of Multiple Synchronous Mobile Users Using Oversampling and/or Multiple Antennas," in *Proc. 28th Asilomar Conference on Signal, Systems & Computers*, Pacific Grove, CA, Nov. 1994.

- [13] G. Hari Kumar 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.
- [14] G. 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.
- [15] M. Gürelli and C. Nikias, "A New Eigenvector-Based Algorithm for Multichannel Blind Deconvolution of Input Colored Signals," in *Proc. ICASSP*, 1993, pp. 448–451.
- [16] L. Baccala and S. Roy, "A New Time-Domain Blind Identification Method," *IEEE Signal Processing Letters*, vol. 1, no. 6, pp. 89–91, June 1994.
- [17] E. Moulines, P. Duhamel, J. Cardoso, and S. Mayrargue, "Subspace Methods for the Blind Identification of Multichannel FIR filters," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 516–526, Feb. 1995.
- [18] D. Slock and C. Papadias, "Further Results on Blind Identification and Equalization of Multiple FIR Channels," in *Proc. ICASSP 95 Conf.*, Detroit, Michigan, May 1995.
- [19] E. de Carvalho and D. Slock, "Maximum-Likelihood Blind Equalization of Multiple FIR Channels," in *Proc. ICASSP 96 Conf.*, Atlanta, USA, May 1996.
- [20] E. Pitié and P. Duhamel, "Bilinear Methods for Blind Channel Equalization: (No) Local Minimum Issue," in *Proc. ICASSP 98 Conf.*, Seattle, USA, May 1998.
- [21] E. de Carvalho and D. Slock, "Maximum-Likelihood FIR Multi-Channel Estimation with Gaussian Prior for the Symbols," in *Proc. ICASSP 97 Conf.*, Munich, Germany, April 1997.
- [22] —, "Asymptotic Performance of ML Methods for Semi-Blind Channel Estimation," in *Proc. Asilomar Conference on Signals, Systems & Computers*, Pacific Grove, CA, Nov. 1997.
- [23] J. Ayadi, E. de Carvalho, and D. Slock, "Blind and Semi-Blind Maximum Likelihood Methods for FIR Multichannel Identification," in *Proc. ICASSP 98 Conf.*, Seattle, USA, May 1998.
- [24] M. 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.
- [25] L. Tong and Q. Zhao, "Blind Channel Estimation by Least Squares Smoothing," in *Proc. ICASSP 98 Conf.*, Seattle, USA, May 1998.
- [26] J. Zhu, Z. Ding, and X.-R. Cao, "Column-Anchored Zeroforcing Blind Equalization for Multiuser Wireless FIR Channels," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 3, p. 411, March 1999.
- [27] E. de Carvalho and D. Slock, "Semi-Blind Maximum-Likelihood Estimation with Gaussian Prior for the Symbols using Soft Decisions," in *48th Annual Vehicular Technology Conference*, Ottawa, Canada, May 1998.
- [28] M. Kristensson, "On parameter estimation in wireless communications, sensor array processing and spectral analysis," Ph.D. dissertation, Royal Institute of Technology (KTH), Sweden, 1998.
- [29] E. de Carvalho and D. Slock, "A Fast Gaussian Maximum-Likelihood Method for Blind Multichannel Estimation," in *Signal Processing Advances in Wireless Communications (SPAWC)*, Annapolis, Maryland, USA, May 1999.
- [30] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, "Exact and Large Sample ML Techniques for Parameter Estimation and Detection in Array Processing," in *Radar Array Processing*, S. Haykin, J. Litva, and T. Shepherd, Eds. Springer-Verlag, 1993.
- [31] J. Bapat, "Partially blind identification of fir channels for qam signals," Ph.D. dissertation, Pennsylvania State University, Aug. 1996.
- [32] —, "Partially blind estimation: ML-based approaches and Cramér-Rao bound," *Signal Processing*, vol. 71, no. 3, pp. 265–277, Dec. 1998.
- [33] D. Yellin and B. Porat, "Blind Identification of FIR Systems Excited by Discrete-Alphabet Inputs," *IEEE Transactions on Signal Processing*, vol. 41, no. 3, pp. 1331–1339, March 1993.
- [34] 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.
- [35] D. Slock and E. de Carvalho, "Burst Mode Non-Causal Decision-Feedback Equalization and Blind MLSE," in *Proc. GLOBECOM 96 Conf.*, London, Great Britain, November 1996.
- [36] B. Sadler, R. Kozick, and T. Moore, "Bounds on Bearing and Symbol Estimation with Side Information," *IEEE Transactions on Signal Processing*, vol. 49, no. 4, pp. 822–834, April 2001.
- [37] V. Radionov and S. Mayrargue, "Semi-blind approach to second order identification of SIMO-FIR channel driven by finite-alphabet sequence," in *13th International Conference on Digital Signal Processing*, vol. 1, Santorini, Greece, Jul 1997, pp. 115–118.
- [38] 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.
- [39] H. Cirpan and M. Tsatsanis, "Stochastic Maximum-Likelihood Methods for Semi-Blind Channel Estimation," *IEEE Signal Processing Letters*, vol. 5, no. 1, pp. 21–24, Jan. 1998.
- [40] B. Hochwald and A. Nehorai, "On Identifiability and Information-Regularity in Parameterized Normal Distributions," *Circuits, Systems and Signal Processing*, vol. 16, no. 1, 1997.
- [41] Y. Hua and M. Wax, "Strict Identifiability of Multiple FIR Channels Driven by an Unknown Arbitrary Sequence," *IEEE Transactions on Signal Processing*, vol. 44, no. 3, pp. 756–759, March 1996.
- [42] E. de Carvalho and D. Slock, "Identifiability Conditions for Blind and Semi-Blind Multichannel Estimation," in *European Association for Signal Processing EUSIPCO 98*, Island of Rhodes, Greece, September 1998.
- [43] L. Tong, G. Xu, and T. Kailath, "Blind Identification and Equalization Based on Second-Order Statistics: A Time Domain Approach," *IEEE Transactions on Information Theory*, vol. 40, no. 2, pp. 340–349, March 1994.
- [44] Z. Ding and Y. Li, "On Channel Identification Based on Second-Order Cyclic Spectra," *IEEE Transactions on Signal Processing*, vol. 42, no. 5, pp. 1260–1264, May 1994.
- [45] L. Deneire, E. de Carvalho, and D. 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.



Elisabeth de Carvalho received the M.Sc. degree in electrical engineering from the National Institute of Telecommunications (INT), Evry, France, in 1994 and the Ph.D. degree in electrical engineering from the Ecole Nationale Supérieure des Télécommunications (Tlcom Paris), Paris, France, in 1999. From 1995 to 1998, she was with Laboratoire d'Electronique PHILIPS supported by a CIFRE scholarship. Her Ph.D. work was on receiver algorithms for wireless communications, with emphasis on blind and semi-blind equalization. From October 1999 to

June 2001, she was a postdoctoral fellow at Star Lab, Stanford University, with the group of Prof. J.M. Cioffi, working on signal processing issues for DSL communications. From October 1999 till September 2000, she was the recipient of a Lavoisier scholarship from the French Ministry of Foreign Affairs. From 2001 to 2002, she was with GlobespanVirata, Santa Clara, CA. She is now with NewLogic Technologies, Sophia Antipolis, France. Her research interests are in the area of signal processing for digital communications, synchronization and equalization for wireless and DSL communications.



Dirk Slock received an engineering degree from the University of Gent, Belgium in 1982. In 1984 he was awarded a Fulbright scholarship for Stanford University USA, where he received the MS in Electrical Engineering, MS in Statistics, and PhD in Electrical Engineering in 1986, 1989 and 1989 respectively. While at Stanford, he developed new fast recursive least-squares (RLS) algorithms for adaptive filtering. In 1989-91, he was a member of the research staff at the Philips Research Laboratory Belgium. In 1991, he joined the Eurecom Institute where he is now

professor. At Eurecom, he teaches statistical signal processing and speech coding for mobile communications. His research interests include DSP for mobile communications (antenna arrays for (semi-blind) equalization/interference cancellation and spatial division multiple access, space-time processing and coding, channel estimation) and adaptation techniques for audio processing. More recently he is focusing on receiver design, downlink antenna array processing and source coding for third generation systems, and introducing spatial multiplexing in existing wireless systems. In 2000, he cofounded SIGtone, a start-up developing music signal processing products. He is also active as a consultant on xDSL and DVB-T systems. He received one best journal paper award from the IEEE-SP and one from EURASIP in 1992. He is the coauthor of two IEEE-Globecom98 best student paper awards. He was an associate editor for the IEEE-SP Transactions in 1994-96.