Blind Identification/Equalization Using Deterministic Maximum Likelihood and a Partial Prior on the Input

Florence Alberge, Mila Nikolova, and Pierre Duhamel, *Fellow, IEEE*

***Abstract*—A (semi)deterministic maximum likelihood (DML) approach is presented to solve the joint blind channel identification and blind symbol estimation problem for single-input multiple-output systems. A partial prior on the symbols is incorporated into the criterion which improves the estimation accuracy and brings robustness toward poor channel diversity conditions. At the same time, this method introduces fewer local minima than the use of a full prior (statistical) ML. In the absence of noise, the proposed batch algorithm estimates perfectly the channel and symbols with a finite number of samples.**

Based on these considerations, an adaptive implementation of this algorithm is proposed. It presents some desirable properties including low complexity, robustness to channel overestimation, and high convergence rate.

***Index Terms*—Adaptive algorithm, blind equalization, deterministic maximum likelihood method, joint estimation, prior knowledge.**

I. Introduction

BLIND identification is an important problem in many areas and especially in wireless communications. Blind techniques present some advantages compared to the traditional training methods [1], [2]. First, the reduced need for overhead information increases the bandwidth efficiency. Furthermore, in certain communication systems, the synchronization between the receiver and the transmitter is not possible; thus training sequences are not exploitable. Finally, even if some training sequence exists, the combination of trained and blind techniques can often lead to improved performances, allowing fast tracking of time-varying channels, for example.

Early approaches to blind equalization were based on higher order statistics of the received signal [3] since the second-order statistics of a scalar system output do not contain enough information to identify a nonminimum phase system. Although these algorithms are robust and reliable in many cases, estimating high-order statistics usually requires a large number of data samples. Hence, their application in fast varying environment is intrinsically limited. Tong *et al.* suggested a different option [4]. They proposed to introduce time or spatial diversity at the output. Then, the system considered is a single-input multiple-output (SIMO) system. The SIMO equalization problem can be solved using second-order statistics only, as long as the subchannels do not share common zeros. In a fast fading environment, the statistical model of the input may not be available, or there may not be enough samples to find a reliable estimate of the statistics. In this kind of scenario, the problem may be solved by treating the input as a deterministic variable. Generally, the resulting methods have the finite sample convergence property (i.e., the channel can be perfectly estimated using a finite number of samples in noiseless situations). This is a desirable property especially in packet transmission systems.

In this paper, we focus on deterministic maximum likelihood (DML) methods since they have the additional advantage of being high signal-to-noise ratio (SNR) efficient [5]. Among the major contributions to DML methods, we can cite the two-step maximum likelihood (TSML) [6] and the iterative quadratic maximum likelihood (IQML) [7], both concentrating on channel estimation. Feder *et al.* proposed in [8] a dual algorithm to IQML which aims at estimating the symbols at each step. Unfortunately, the adaptive implementation of these methods is often cumbersome. Another DML method, the maximum likelihood block algorithm (MLBA), has been proposed in [9]. The MLBA performs least squares estimation both in the channels and in the symbols in an alternating manner. This formulation permits to derive easily an adaptive algorithm (MLAA) as shown by the authors in [10]. The MLAA presents some nice properties including low-complexity in computation. However, it is not robust to the overestimation of the channel order and it has a limited ability to track time-varying channels.

In this paper, we present a new algorithm that meets the following four characteristics: adaptivity, low complexity, good speed of convergence, and robustness to the overestimation of the channel order. The proposed method consists of incorporating prior information (related to the input signal) into the DML criterion. The two first properties follow from the MLAA-like structure of the algorithm and the last characteristics are a consequence of the use of the prior. Seshadri [11] and Gosh and Weber [12] first proposed to incorporate the finite alphabet properties into DML to improve the accuracy of the estimates. Later, Talwar proposed the iterative least square with projection (ISLP) [13], which estimates the symbols first without taking the finite alphabet property into account and then projects the estimates onto the alphabet. The problem with these methods is that their convergence is not guaranteed in general and that the incorporation of the finite alphabet property often increases the number of local minima. This is partially solved here by considering only a partial prior on the symbols in order to limit the number of additional spurious local minima (different from the global one). In the proposed approach, a continuous probability distribution function is used which reflects our prior knowledge on the input.

Manuscript received April 5, 2004; revised March 26, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Helmut Boelcskei.

F. Alberge and P. Duhamel are with Supelec/LSS, 91192 Gif-Sur-Yvette Cedex, France (e-mail: alberge@lss.supelec.fr; pierre.duhamel@lss.supelec.fr).

M. Nikolova is with CMLA-ENS de Cachan and CNRS UMR 8536, 94235 Cachan Cedex, France (e-mail: nikolova@cmla.ens-cachan.fr).

Digital Object Identifier 10.1109/TSP.2005.861787

Using Bayesian theory and the classical likelihood function, we are able to derive a new criterion. Such a criterion has also been used in [14] for the purpose of binary images reconstruction. The proposed relaxation technique can also be found in [15] and [16] for application to multiuser detection in CDMA systems. However, channel estimation is not involved in these two contributions. The new algorithm, called conditional maximum likelihood batch algorithm (CMLBA), is then obtained in the same way as the MLBA (alternating minimization) and the corresponding adaptive version is derived.

Concerning the local minima problem, we prove that for a "weak" prior (to be defined later), the stationary points of the CMLBA are also stationary points of the MLBA. And, for each stationary point $(\hat{\mathbf{h}}, \hat{\mathbf{s}})$ of the MLBA, there exists a scale factor α such that $(\alpha\hat{\mathbf{h}}, \hat{\mathbf{s}}/\alpha)$ is also a stationary point of the CMLBA. Thus, the use of such a prior does not increase the number of local minima and, at the same time, it brings robustness to poor channel diversity conditions, as shown in the experimental results. For a stronger prior, the number of local minima is likely to increase. However, we show below that a local minimum is not stable through a recursive procedure, as already shown in [10]. As a result, the proposed recursive algorithm is unlikely to converge toward a local minimum.

This paper is organized as follows. Section II presents the general setup and some properties about the DML criterion. For noise-free data, we recall that the global minimum of the DML criterion is unique. The derivation of the CMLBA is available in Section III. The local minima problem is analyzed in Section IV. In Section V, we explain how to improve the quality of the estimators in the particular case of an ill-conditioned channel matrix. An adaptive version of the CMLBA is proposed in Section VI. The performance of the algorithms and comparison with existing approaches are provided in Section VII.

II. Problem Formulation

This paper addresses SIMO systems (see Fig. 1). Let $\{\tilde{s}(n)\}$ denote the symbol sequence at the input of the system and $x_i(n), 1 \le i \le L$, the ith output. The output $x_i(n)$ may be the signal picked on the ith sensor of an array (spatial diversity); or may be obtained by oversampling of a factor L the continuous time signal received on a single sensor (time diversity); or may follow from the combination of both spatial and time diversity. Such a system is described as

$$\mathrm{x}_i(n) = \sum_{k=0}^{M} \tilde{\mathrm{h}}_i(n-k)\tilde{\mathrm{s}}(k) + \mathrm{b}_i(n) \; i = 1, \dots, L$$

where $\mathbf{h}_i = [\mathbf{h}_i(0), \dots, \mathbf{h}_i(M)]^T$ is the channel impulse response, M is the maximum order of any channel, and $\{b_i(n)\}, 1 \le i \le L$, is a Gaussian independent identically distributed (i.i.d.) additive noise. Sequences $\{b_i(n)\}$ and $\{b_j(n)\} (i \ne j)$ are assumed uncorrelated. For convenience, we adopt the following notations throughout this paper.

- h, s are variables denoting any channel and any symbol sequence, respectively.

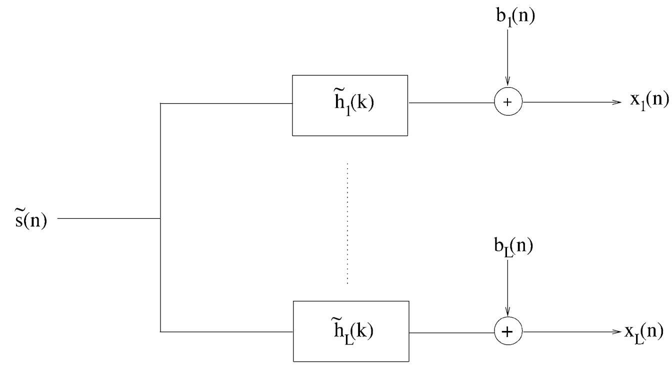

Fig. 1. Single-input/multiple-output system.

- $\tilde{\mathrm{h}}, \tilde{\mathrm{s}}$ are the true channels and the true symbols, respectively, and $\mathbf{x}(n)$ stands for the corresponding observation at time n.
- $\hat{\mathrm{h}}, \hat{\mathrm{s}}$ are the estimates of $\tilde{\mathrm{h}}$ and $\tilde{\mathrm{s}}$.
- $\mathbf{s}_N(n) = [s(n), s(n-1), \dots, s(n-N-M+1)]^T$ and n is the time index.
- $\hat{\mathbf{s}}_N^{(i)}(n+i) = [\hat{s}(n+i), \hat{s}(n+i-1), \dots, \hat{s}(n+i-N-M+1)]^T$ is a vector of length $M+N$ containing symbols estimated at iteration i.

The channel impulse responses $\tilde{\mathbf{h}}_i, 1 \le i \le L (L > 1)$ are assumed to have a finite length and M stands for the maximum order of any channel. Let $\mathbf{X}_N(n) = [\mathrm{x}_1(n), \dots, \mathrm{x}_L(n), \dots, \mathrm{x}_1(n-N+1), \dots, \mathrm{x}_L(n-N+1)]^T$ denote the vector obtained by interleaving the outputs of the different channels and $\tilde{\mathbf{h}}(k) = [\tilde{h}_1(k), \dots, \tilde{h}_L(k)]^T$. Then, the output $\mathbf{X}_N(n)$ reads

$$\mathbf{X}_N(n) = \mathcal{T}_N(\tilde{\mathbf{h}})\tilde{\mathbf{s}}_N(n) + \boldsymbol{B}_N(n) \tag{1}$$

where $\mathbf{B}_N(n)$ stands for the noise vector. In (1), operator $\mathcal{T}_N$ transforms the set of channel coefficients $\mathbf{h}(k) = [h_1(k), \dots, h_L(k)]^T, k = 0, \dots, M$ into the following $LN \times (M+N)$ generalized Sylvester matrix:

$$\mathcal{T}_N(\mathbf{h}) = \begin{pmatrix} \mathbf{h}(0) & \dots & \mathbf{h}(M) & 0 & \dots & 0 \\ 0 & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & 0 \\ 0 & \dots & 0 & \mathbf{h}(0) & \dots & \mathbf{h}(M) \end{pmatrix}.$$

Let $\mathcal{U}$ be the operator that transforms a vector $\mathbf{s}_N(n)$ into an $LN \times L(M+1)$ matrix $\mathcal{U}(\mathbf{s}_N(n))$, in such a way that

$$\mathcal{U}(\mathbf{s}_N(n))\mathbf{h} = \mathcal{T}_N(\mathbf{h})\mathbf{s}_N(n), \quad \forall \mathbf{s}_N, \quad \forall \mathbf{h}. \tag{2}$$

It can be shown that this matrix reads

$$\mathcal{U}(\mathbf{s}_N(n)) = \begin{pmatrix} I_L \bigotimes s_1(n)^T \\ I_L \bigotimes s_1(n-1)^T \\ \vdots \\ I_L \bigotimes s_1(n-N+1)^T \end{pmatrix}$$

where $\bigotimes$ stands for the Kronecker product and I_L is the $L \times L$ identity matrix. The results displayed in the paper rely on the following assumptions.

H1) $\mathcal{T}_N(\tilde{\mathbf{h}})$ is full column rank.

H2) The symbol sequence $\tilde{s}_N(n)$ has linear complexity $2M+1$ or greater [17]. The linear complexity of the sequence $\{\tilde{s}(n-k)\}_{k=0}^{k=N+M-1}$ is defined as the smallest value of c for which there exists $\{\lambda_j\}_{j=1}^{c}$ such as

$$\tilde{s}(n-i) = -\sum_{j=1}^{c} \lambda_j \tilde{s}(n-i-j) \quad i = c, \ldots, N+M-1.$$

The linear complexity measures the predictability of a finite length deterministic sequence.

H2') When H2) is met, it can be shown that $\mathcal{U}(\tilde{\mathbf{s}}_N(n))$ is full column rank.

H3) M (maximum order of the channels) is known or correctly estimated.

H4) The emitted symbols belong to a phase-shift keying (PSK) modulation.

Assumption H1) ensures that (1) is an overdetermined system for $\tilde{\mathbf{h}}$ fixed. This assumption is most often met. However, situations with poor channel diversity conditions may occur. So, it is important to develop methods that are robust to this situation. Similarly, H2') ensures that (1) is an overdetermined system for $\tilde{\mathbf{s}}_N(n)$ fixed. Denote $\nu_M(\tilde{\mathbf{s}}_N(n))$ as the matrix defined by the equation at the bottom of the page. Then, H2) implies rank$(\nu_M(\tilde{\mathbf{s}}_N(n))) = 2M+1$. Hence, the sample covariance of the vector sequence $\tilde{\mathbf{s}}_{M+1}(n) = [\tilde{s}(n), \tilde{s}(n-1), \ldots, \tilde{s}(n-2M)]^T$ is full rank [18]. We can remark that rank$(\nu_M(\tilde{\mathbf{s}}_N(n)) = 2M+1$ implies that, necessarily, $N \geq 3M+1$.

The problem considered in this paper is to identify both $\tilde{\mathbf{h}}$ and $\tilde{\mathbf{s}}_N(n)$ based on $\mathbf{X}_N(n)$ only.

The blind equalization problem is viewed as a joint channel and symbol estimation problem. The criterion used is the DML criterion. Following [6], [9], and [7], $(\tilde{\mathbf{h}}, \tilde{\mathbf{s}}_N)$ are estimated through the minimization of the DML criterion with respect to the joint variable $(\mathbf{h}, \mathbf{s}_N(n))$

$$\begin{aligned}\mathcal{J}(\mathbf{h}, \mathbf{s}_N(n)) &= \|\mathbf{X}_N(n) - \mathcal{T}_N(\mathbf{h})\mathbf{s}_N(n)\|^2 \\ &= \|\mathbf{X}_N(n) - \mathcal{U}(\mathbf{s}_N(n))\mathbf{h}\|^2. \end{aligned} \tag{3}$$

Hence, the estimated channel and symbols read

$$(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N(n)) = \arg \min_{(\mathbf{h}, \mathbf{s}_N(n))} \mathcal{J}(\mathbf{h}, \mathbf{s}_N(n)). \tag{4}$$

In the noiseless case, the global minimum is obtained only for the true channel and symbols (up to a scale factor) [9], [19].

III. Conditional Maximum Likelihood Technique

In a fast fading environment, building reliable statistical estimates is a problem: data related to a given channel are not numerous. In such a situation, the symbols are assumed arbitrary and a deterministic method is used. But, if the system is not time-varying and if the data sequence is long enough so that the statistical estimates are reliable, then a statistical method should be used, since in that case the statistical method outperforms the deterministic one in terms of estimation accuracy. The approach proposed in this section is a tradeoff between DML and SML with the additional advantage that it can be used either when the channel is time-varying or not. In this approach, we consider the transmitted symbols to be no longer deterministic quantities but random variables that obey to an arbitrary (different from the true) statistical distribution. As a result, the obtained algorithm involves a lower computational cost than the statistical method and provides a better estimation accuracy than a DML method.

A. Derivation of the Constrained Criterion

A full use of the knowledge on the emitted symbols (their alphabet) usually introduces many local minima. Instead, we propose to consider only a partial prior. Assume that the emitted symbols belong to a PSK modulation, and consider the probability density function (pdf) shown at the bottom of the page, where Z is a normalization term, κ is a positive scalar, and $s(k)$ is the kth component of $\mathbf{s}$. The shape of the distribution function corresponding to $\kappa = 1$ and $\kappa = 10$ is plotted in Fig. 2, where real data are considered. When $s(k)$ is outside the unit circle, the probability is zero, whereas for symbols inside the unit circle, the probability increases with $\|s(k)\|$. When κ tends to zero, the shape $p(s(k))$ tends to be uniform within the unit circle and zero outside. Even if this special case corresponds to a very "weak" prior, it is of interest, since some properties of the convergence points of the algorithm can be demonstrated under this assumption. This will help in understanding the local minima problem.

Let $p(\mathbf{X}_N(n)|\mathbf{h}, \mathbf{s}_N)$ denote the likelihood function conditioned on both the channels and the symbols. The conditional likelihood function $p(\mathbf{X}_N(n), \mathbf{s}_N|\mathbf{h})$ reads

$$p(\mathbf{X}_N(n), \mathbf{s}_N \,|\, \mathbf{h}) = p(\mathbf{X}_N(n) \,|\, \mathbf{h}, \mathbf{s}_N) p(\mathbf{s}_N). \tag{6}$$

$$\nu_M(\tilde{\mathbf{s}}_N(n) = \begin{pmatrix} \tilde{s}(n) & \tilde{s}(n-1) & \cdots & \tilde{s}(n-N+M+1) \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{s}(n-2M) & \tilde{s}(n-2M-1) & \cdots & \tilde{s}(n-N-M+1) \end{pmatrix}$$

$$\begin{cases} p(s(k)) = 0 & \text{if } \|s(k)\| > 1, \\ p(s(k)) = \frac{1}{Z} e^{\kappa \|s(k)\|^2}, & \text{if } \|s(k)\| \leq 1 \quad p(\mathbf{s}) = \Pi_k p(s(k)) \end{cases} \tag{5}$$

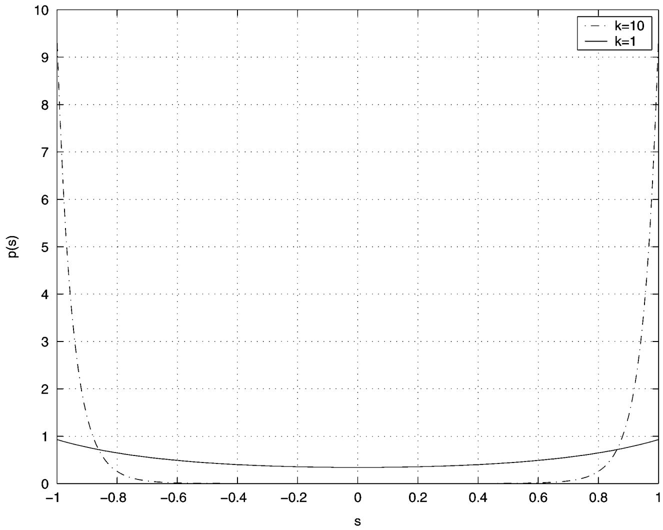

Fig. 2. Probability distribution function for $\kappa = 1, \kappa = 10$. The symbols are assumed to have real values.

$\mathbf{B}_N(n)$ is assumed Gaussian, hence $p(\mathbf{X}_N(n)|\mathbf{h}, \mathbf{s}_N)$ satisfies the following relation:

$$p(\mathbf{X}_N \,|\, \mathbf{h}, \mathbf{s}_N) = \frac{1}{(2\pi\sigma_b^2)^{LN/2}} \times \exp\left(-\frac{1}{2\sigma_b^2}\|\mathbf{X}_N - \mathcal{T}_N(\mathbf{h})\mathbf{s}_N\|^2\right). \quad (7)$$

Inserting (5) and (7) into (6), we obtain the conditional likelihood function shown at the bottom of the page, where $\gamma = 2\sigma_b^2\kappa$ and $\mathcal{E}^{M+N}$ stands for the unit disk ($\mathcal{E}^{M+N} = \{\mathbf{s}_N \in \mathbb{C}^{M+N} : \|s(k)\| \leq 1, k = 0 \ldots M+N-1\}$). The maximization of $p(\mathbf{X}_N, \mathbf{s}_N|\mathbf{h})$ is equivalent to the minimization of the following criterion on $\mathbb{C}^{L(M+1)} \times \mathcal{E}^{M+N}$:

$$\mathcal{L}_\gamma(\mathbf{h}, \mathbf{s}_N) = \|\mathbf{X}(n) - \mathcal{T}_N(\mathbf{h})\mathbf{s}_N\|^2 - \gamma\|\mathbf{s}_N\|^2 \quad (9)$$

where $\gamma > 0$. Note that a somewhat similar idea has already been exploited by Papadias in [20], where the transmitted symbols are considered as random variables that obey to a zero-mean Gaussian distribution leading to the criterion

$$\mathcal{L}_{\text{Papadias}}(\mathbf{h}, \mathbf{s}_N) = \|\mathbf{X}(n) - \mathcal{T}_N(\mathbf{h})\mathbf{s}_N\|^2 + \beta\|\mathbf{s}_N\|^2 \quad (10)$$

with $\beta > 0$. The same class of criteria is exploited in [15] except that the criterion is minimized with respect to the symbols only. The Gaussian assumption leading to (10) is unrealistic: the symbols close to zero have the highest probability. At the opposite, the pdf in (5) reflects the prior knowledge on the input.

The criterion $\mathcal{L}_\gamma(\mathbf{h}, \mathbf{s}_N)$ is a convex criterion with respect to each variable separately as long as $\gamma \leq \lambda_{\min}$, where $\lambda_{\min}$ stands for the smallest eigenvalue of $\mathcal{T}_N(\mathbf{h})^H\mathcal{T}_N(\mathbf{h})$. In the nonconvex case, the quadratic programming problem subject to linear constraints is NP-complete [21]. Moreover, checking only local optimality in constrained nonconvex programming is NP-hard [22]. This means that the computing time to obtain a solution will grow exponentially with the number of variables. Such a computational cost is unaffordable in the context under study. From now on, we shall only consider the case where $\gamma \leq \lambda_{\min}$. The readers interested by the nonconvex quadratic programming problem can refer to [23]–[25].

B. Implementation of the Method

Many solutions can be proposed to solve the constrained optimization problem in (9). Here, we follow the approach proposed by Gesbert [9] for solving the unconstrained criterion in (3). It presents two major advantages: 1) this formulation is well suited for deriving recursive solutions and 2) the prior information can be incorporated easily, which is not the case with IQML-like approaches. In this approach, a least squares estimation is performed successively in the channel and in the symbols, in an alternating manner. At each step of the iterative procedure, the channels and symbols estimates read

$$\hat{\mathbf{h}}^{(k)} = \left[\mathcal{U}\left(\hat{\mathbf{s}}_N^{(k-1)}\right)^H \mathcal{U}\left(\hat{\mathbf{s}}_N^{(k-1)}\right)\right]^{-1} \mathcal{U}\left(\hat{\mathbf{s}}_N^{(k-1)}\right)^H \mathbf{X}_N(n) \quad (11)$$

$$\hat{\mathbf{s}}_N^{(k)} = \arg\min_{\mathbf{s}_N \in \mathcal{E}^{M+N}} \mathcal{L}_\gamma\left(\hat{\mathbf{h}}^{(k)}, \mathbf{s}_N\right) \text{ with}$$
$$\mathcal{E}^{M+N} = \{\mathbf{s}_N \in \mathbb{C}^{M+N} : \|s(k)\| \leq 1\} \quad (12)$$

where matrix $\mathcal{U}(\hat{\mathbf{s}}_N^{(k)}(n))$ is assumed to be full rank $\forall k$. This class of algorithm will be referred to as CMLBA_γ. Each step diminishes the value of the criterion and thus the algorithm converges, however possibly toward a spurious local minima. The optimization problem in (12) is solved by a relaxation method as detailed below.

Let $s(j)$ denote the jth component of $\mathbf{s}_N$. The partial criterion $l_\gamma^{(j)}(s(j))$ reads

$$\begin{aligned} l_\gamma^{(j)}(s(j)) &= s^*(j)(A_j - \gamma)s(j) + s^*(j)B_j \\ &\quad + B_j^* s(j) - s^*(j)C_j + C_j^* s(j) \\ &\qquad \text{with } \|s(j)\| \leq 1 \\ A_j &= T_j^{(k)^H} T_j^{(k)} \\ B_j &= T_j^{(k)^H} \sum_{l \neq j} T_l^{(k)} s_l \\ C_j &= T_j^{(k)^H} \mathbf{X}_N(n) \end{aligned}$$

$$\begin{cases} p(\mathbf{X}_N, \mathbf{s}_N|\mathbf{h}) = \frac{1}{Z\left(2\pi\sigma_b^2\right)^{LN/2}} \exp\left(-\frac{1}{2\sigma_b^2}\{\|\mathbf{X}_N - \mathcal{T}_N(\mathbf{h})\mathbf{s}_N\|^2 - \gamma\|\mathbf{s}_N\|^2\}\right), & \text{if } \mathbf{s}_N \in \mathcal{E}^{M+N} \\ p(\mathbf{X}_N, \mathbf{s}_N|\mathbf{h}) = 0, & \text{elsewhere} \end{cases} \quad (8)$$

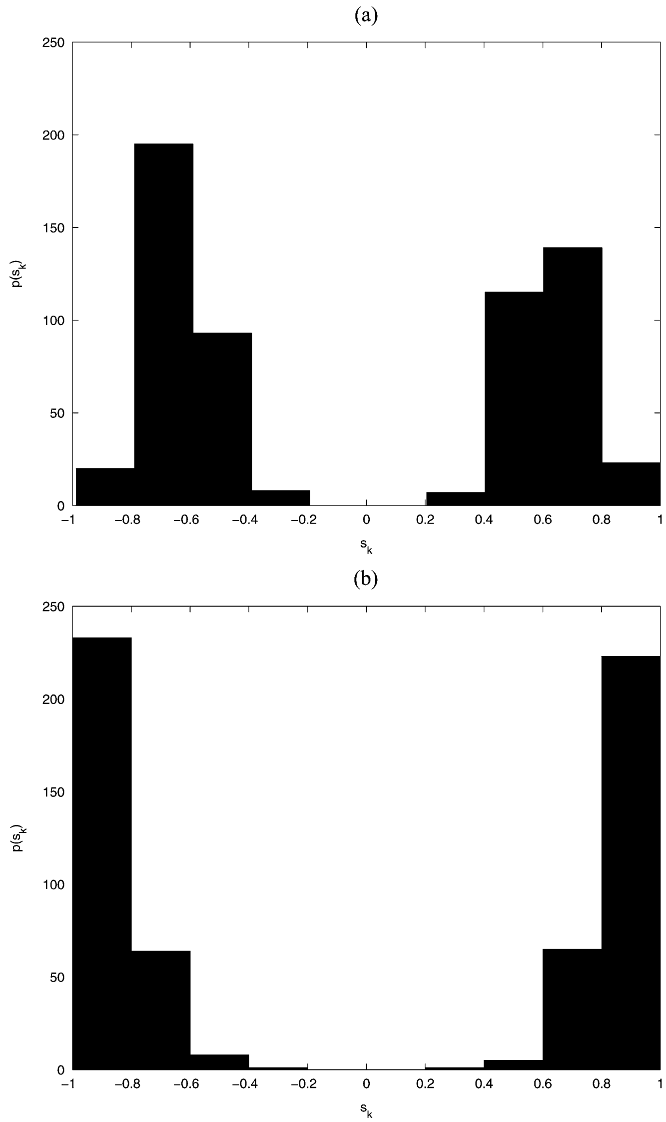

Fig. 3. Histogram of the estimated symbols computed with (a) $\text{CMLBA}_{\min}$ and (b) $\text{CMLBA}_{\max}$. SNR $= 10$ dB, 600 input symbols.

where $T_j^{(k)}$ is the jth column of $\mathcal{T}_N(\hat{\mathbf{h}}^{(k)})$. The derivative of $l_\gamma^{(j)}(s(j))$ with respect to $s^*(j)$ can be written as

$$\frac{\partial l_\gamma^{(j)}(s(j))}{\partial s^*(j)} = s(j)(A_j - \gamma) + B_j - C_j$$

where $(\partial/\partial s^*(j)) = (1/2)((\partial/\partial t_j) + (\partial/\partial u_j))$ and t_j and u_j stand, respectively, for the real and imaginary part of $s(j)$ [26]. Then at each iteration of the relaxation method, $s(j)$ is computed by the following expression:

$$s(j) = -\mathbf{P}\left(\frac{B_j - C_j}{A_j - \gamma}\right)$$

where $\mathbf{P}$ is the projection operator of $\mathbb{C}^{M+N}$ onto $\mathcal{E}^{M+N}$. The relaxation method converges as long as $(\mathcal{T}_N(\hat{\mathbf{h}}^{(k)})^H \mathcal{T}_N(\hat{\mathbf{h}}^{(k)}) - \gamma \mathbf{I}_{M+N})$ is positive definite [27]. The role played by parameter γ is illustrated in Fig. 3, where we plot the histograms of the symbols estimated via the CMLBA when $\gamma = 0$ (denoted as $\text{CMLBA}_{\min}$) and when $\gamma = \lambda_{\min}$ (denoted as $\text{CMLBA}_{\max}$). A mixed-phase channel is considered and the modulation is a binary PSK (BPSK). It is clear in Fig. 3 that a large value of γ prioritizes the extreme values of the set.

IV. The Local Minima Problem

In this section, the local minima problem is investigated. First, we prove that the global minimum of $\mathcal{L}_\gamma$ is obtained for the true parameters only. Then, a characterization of the local minima is given. The general case $(0 \le \gamma \le \lambda_{\min})$ is first considered, and finally, we concentrate on a special case: the uniform prior on the unit disk $(\gamma = 0)$, which was denoted "weak" prior in the introduction.

A. Uniqueness of the Global Minimum

The following theorem proves the identifiability property for the class of criteria $\mathcal{L}_\gamma(\mathbf{h}, \mathbf{s}_N(n))$.

Theorem 1: In the noiseless case and under H1) and H2), $(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N(n))$ is the global minimum of $\mathcal{L}_\gamma(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N(n))$ on $\mathbb{C}^{L(M+1)} \times \mathcal{E}^{M+N}$ iff $\exists \alpha \in \mathbb{C}^*$ such that $\hat{\mathbf{h}} = \alpha \tilde{\mathbf{h}}, \hat{\mathbf{s}}_N(n) = \tilde{\mathbf{s}}_N(n)/\alpha$. If $\gamma > 0$, then $\|\alpha\| = 1$.

Proof: Stating that $(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N(n))$ is the global minimum of $\mathcal{L}_\gamma(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N(n))$ on $\mathbb{C}^{L(M+1)} \times \mathcal{E}^{M+N}$ is equivalent to the following set of equations:

$$\|\mathcal{T}_N(\hat{\mathbf{h}})\hat{\mathbf{s}}_N(n) - \mathbf{X}_N(n)\| = 0 \quad (13)$$

$$\|\hat{\mathbf{s}}_N\|^2 = M + N \ (\text{if} \gamma > 0). \quad (14)$$

In the noiseless case and under H1) and H2), the global minimum is unique up to a scalar factor. Thus, (13) implies that $\exists \alpha \in \mathbb{C}^*$ such that $\hat{\mathbf{h}} = \alpha \tilde{\mathbf{h}}, \hat{\mathbf{s}}_N(n) = \tilde{\mathbf{s}}_N(n)/\alpha$ [9], [19].

The estimated symbols belong to $\mathcal{E}^{M+N}$. Thanks to (14) we conclude that, if $\gamma > 0$, then $\|\alpha\| = 1$. ■

Thus, the global minimum of $\mathcal{L}_\gamma(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N(n))$ on the set $\mathbb{C}^{L(M+1)} \times \mathcal{E}^{M+N}$ is obtained only for the true values of the parameters up to a phase displacement (for $\gamma > 0$) which ensures that the true parameters can be identified.

B. General Case

We first investigate the characterization of the stationary points of the CMLBA_γ when $0 \le \gamma \le \lambda_{\min}$. Let $(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N(n))$ denote a stationary point; then it is solution of the following set of equations:

$$\hat{\mathbf{h}} = [\mathcal{U}(\hat{\mathbf{s}}_N)^H \mathcal{U}(\hat{\mathbf{s}}_N)]^{-1} \mathcal{U}(\hat{\mathbf{s}}_N)^H \mathbf{X}_N(n) \quad \text{where } \hat{\mathbf{s}}_N \in \mathcal{E}_{M+N} \quad (15)$$

$$\hat{\mathbf{s}}_N(n) = \arg \min_{\mathbf{s}_N \in \mathcal{E}^{M+N}} \mathcal{L}_\gamma(\hat{\mathbf{h}}, \mathbf{s}_N(n)) \quad 0 \le \gamma \le \lambda_{\min}. \quad (16)$$

Equation (16) is a minimization problem subject to inequality constraints. Moreover, $\mathcal{L}_\gamma(\hat{\mathbf{h}}, \mathbf{s}_N(n))$ is a convex function of $\mathbf{s}_N(n)$ as long as $\gamma \le \lambda_{\min}$ and $\mathcal{E}^{M+N}$ is a convex set. Thus, the Kuhn and Tucker [27] relations provide a necessary and sufficient condition for $\hat{\mathbf{s}}_N(n)$ to be a solution of (16). As a result, a stationary point of the CMLBA_γ is such that

$$\hat{\mathbf{h}} = [\mathcal{U}(\hat{\mathbf{s}}_N)^H \mathcal{U}(\hat{\mathbf{s}}_N)]^{-1} \mathcal{U}(\hat{\mathbf{s}}_N)^H \mathbf{X}_N(n) \quad \text{where } \hat{\mathbf{s}}_N \in \mathcal{E}_{M+N} \quad (17)$$

$$\hat{\mathbf{s}}_N(n) = [\mathcal{T}_N(\hat{\mathbf{h}})^H \mathcal{T}_N(\hat{\mathbf{h}}) - \gamma \mathbf{I}_{M+N} + \Phi]^{-1} \mathcal{T}_N(\hat{\mathbf{h}})^H \mathbf{X}_N(n) \quad \text{where } \hat{\mathbf{s}}_N \in \mathcal{E}_{M+N} \quad (18)$$

$$\Phi = \mathrm{diag}(\phi_i)_{0\le i\le M+N-1},$$
$$\phi_i \ge 0 \quad \text{and } \phi_i(\|\hat{s}(n-i)\|^2-1) = 0\ \forall\, i=0,\dots,M+N-1 \quad (19)$$

where $\hat{s}(n-i)$ stands for the ith component of $\hat{\mathbf{s}}_N$. The difference between the stationary points of the MLBA and of the CMLBA$_\gamma$ is due to the term $\Phi - \gamma\mathbf{I}_{M+N}$ appearing in the expression of $\hat{\mathbf{s}}_N(n)$. Then, the value of the taps of Φ is of major importance. They are further characterized in the following proposition.

Proposition 1: Let $(\hat{\mathbf{h}}, \hat{\mathbf{s}}N(n))$ be a stationary point of the CMLBA$_\gamma$ and Φ the matrix defined in (17)–(19). Then the following relation holds:

$$\mathrm{trace}(\Phi) = \gamma\|\hat{\mathbf{s}}_N\|^2. \quad (20)$$

Proof: $(\hat{\mathbf{h}}, \hat{\mathbf{s}}N(n))$ is a stationary point of the CMLBA. Then (17)–(19) imply that the following two relations are met:

$$\hat{\mathbf{h}}^H\mathcal{U}(\hat{\mathbf{s}}_N)^H[\mathcal{T}_N(\hat{\mathbf{h}})\hat{\mathbf{s}}_N(n) - \mathbf{X}_N(n)] = 0 \quad (21)$$
$$\hat{\mathbf{s}}_N(n)^H\mathcal{T}_N(\hat{\mathbf{h}})^H[\mathcal{T}_N(\hat{\mathbf{h}})\hat{\mathbf{s}}_N(n) - \mathbf{X}_N(n)] = \hat{\mathbf{s}}_N(n)^H(\gamma\mathbf{I}_{M+N} - \Phi)\hat{\mathbf{s}}_N(n). \quad (22)$$

The left terms of the two equations above are strictly equivalent. Thus (21) and (22) boil down to

$$\gamma\|\hat{\mathbf{s}}_N(n)\|^2 = \hat{\mathbf{s}}_N(n)^H\Phi\hat{\mathbf{s}}_N(n) = \sum_{i=0}^{M+N-1}\phi_i\|\hat{s}(n-i)\|^2.$$

Let I be the set defined as $I = \{i = 0,\dots,M+N-1 : \|s(n-i)\|^2 = 1\}$. The conditions in (19) imply that $\phi_i = 0$ for $i \notin I$. Thus

$$\sum_{i\in I}\phi_i = \mathrm{trace}(\Phi) = \gamma\|\hat{\mathbf{s}}_N(n)\|^2.$$

■

This result illustrates the role of the term $\gamma\|\mathbf{s}_N(n)\|$ in the criterion. Actually, for $\gamma > 0$, the relation $\mathrm{trace}(\Phi) = \gamma\|\hat{\mathbf{s}}_N(n)\|^2$ leads to $\mathrm{trace}(\Phi) > 0$ ($\hat{\mathbf{s}}_N(n) = 0$ is not an acceptable solution). Since, by definition $\phi_i \ge 0$, then it exists $i_0 \in [0,\dots M+N-1]$ such that $\phi_{i_0} > 0$ and consequently $\|\hat{s}(n-i_0)\|^2 = 1$. So, when $\gamma > 0$, there is at least one component of the estimated symbol vector that belongs to the unit circle. The parameter γ permits to push the estimated symbols to the frontier of the set. The special case $\gamma = 0$ is considered in the next paragraph.

C. Special Case: Uniform Prior on the Unit Disk

This case is of special interest, since it forces the symbols to belong to the unit disk which brings robustness to poor channel diversity conditions (see Section VII-A). Furthermore, this so-called "weak" prior is shown below not to introduce additional local minima compared to the unconstrained case. When $\gamma = 0$, (20) becomes $\mathrm{trace}(\Phi) = 0$. All the taps of Φ are nonnegative; then we get $\Phi = 0_{M+N}$. Thus, the stationary points $(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N(n))$ of the CMLBA$_{\min}$ (CMLBA$_\gamma$ with $\gamma = 0$) are a solution of the following set of equations:

$$\begin{cases} \hat{\mathbf{h}} = [\mathcal{U}(\hat{\mathbf{s}}_N)^H\mathcal{U}(\hat{\mathbf{s}}_N)]^{-1}\mathcal{U}(\hat{\mathbf{s}}_N)^H\mathbf{X}_N(n)\ \hat{\mathbf{s}}_N \in \mathcal{E}^{M+N} \\ \hat{\mathbf{s}}_N(n) = [\mathcal{T}_N(\hat{\mathbf{h}})^H\mathcal{T}_N(\hat{\mathbf{h}})]^{-1}\mathcal{T}_N(\hat{\mathbf{h}})^H\mathbf{X}_N(n)\ \hat{\mathbf{s}}_N \in \mathcal{E}^{M+N}. \end{cases} \quad (23)$$

This means that, once the CMLBA$_{\min}$ has converged, the solution defined above belongs to the set of local minima of the MLBA, corresponding to a specific α (scale factor), which reflects our prior knowledge up to some degree. On the other side, the stationary points of the MLBA are a solution of the system of (23) except that $\hat{\mathbf{s}}_N \in \mathbb{C}^{M+N}$. If $(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N)$ is a stationary point of the MLBA with $\hat{\mathbf{s}}_N \notin \mathcal{E}^{M+N}$, then $(\alpha\hat{\mathbf{h}}, \hat{\mathbf{s}}_N/\alpha)$ with $\alpha = \max(\|\hat{s}(n-i)\|)_{0\le i\le M+N-1}$ is a stationary point of the CMLBA. Hence, the constraint given by the "weak" prior does not add any local minima to the algorithm. The difference in the set of local minima between the MLBA and the CMLBA lies in the value of the scale factor.

V. CMLBA$_\gamma$ and Ill-Conditioned Filtering Matrix

When some roots of the L subchannel impulse responses are close to each other, the corresponding Sylvester matrix $\mathcal{T}_N(\mathbf{h})$ is hardly full column rank. Thus, $\lambda_{\min}$ (smallest eigenvalue of $\mathcal{T}_N(\mathbf{h})^H\mathcal{T}_N(\mathbf{h})$) is almost zero. In that case, CMLBA$_{\max}$ is equivalent to CMLBA$_{\min}$ and $p(s_k)$ is a uniform pdf. In this section, we explain how we can introduce a strong prior information even when the filtering matrix is badly conditioned. The key point of the method is given by Theorem 2.

Theorem 2: Let $\mathbf{A_1}$ and $\mathbf{A_2}$ be two submatrices of a matrix $\mathbf{A}$ such that $\mathbf{A} = [\mathbf{A_1} \vdots \mathbf{A_2}]$. Then

$$\lambda_{\min}^{A} \le \lambda_{\min}^{A_1} \text{ and } \lambda_{\min}^{A} \le \lambda_{\min}^{A_2}$$

where $\lambda_{\min}^{A}$ (respectively, $\lambda_{\min}^{A_i}$) stands for the smallest eigenvalue of $\mathbf{A}^H\mathbf{A}$ (respectively, $\mathbf{A_i}^H\mathbf{A_i}, i = 1,2$).

Proof: Proof is obvious. ■

If we consider, in the minimization problem, a partition of $\mathcal{T}_N(\mathbf{h})$ instead of considering the whole matrix, then the reduced minimization problem (in terms of number of variables) is at least as well conditioned as the initial problem. Thus, the proposed method consists in partitioning the symbol vector to be estimated and then estimating alternately each part while the rest is fixed to the value obtained at the previous iteration. For simplicity's sake, we give the update equations in the case where the symbol vector is split into two parts. Generalization to others partitions is straightforward since the subvectors have equal lengths (except for the last one if the length of $\mathbf{s}_N$ is not proportional to the number of partitions) and consecutively ordered elements. Let $\mathbf{s}_N = [\mathbf{u}^H\mathbf{v}^H]^H$ where the length of $\mathbf{u}$ (respectively, $\mathbf{v}$) is N_1 (respectively, N_2). The matrix $\mathcal{T}_N(\mathbf{h})$ is also split into two submatrices as: $\mathcal{T}_N(\mathbf{h}) = [\mathcal{T}_N^{1\to N_1}(\mathbf{h})\ \mathcal{T}_N^{N_1\to N_1+N_2}(\mathbf{h})]$ ($\mathcal{T}_N^{1\to N_1}(\mathbf{h})$ contains the N_1 first columns of $\mathcal{T}_N(\mathbf{h})$). Then, at each step of the iterative procedure, the channel and symbol estimates can be read

$$\hat{\mathbf{h}}^{(k)} = \left[\mathcal{U}\left(\hat{\mathbf{s}}_N^{(k-1)}\right)^H\mathcal{U}\left(\hat{\mathbf{s}}_N^{(k-1)}\right)\right]^{-1}\mathcal{U}\left(\hat{\mathbf{s}}_N^{(k-1)}\right)^H\mathbf{X}_N(n) \quad (24)$$

$$\hat{\mathbf{u}}^{(k)} = \arg\min_{\mathbf{u}\in\mathcal{E}^{N_1}}\mathcal{L}_{\gamma_1}\left(\hat{\mathbf{h}}^{(k)}, \begin{bmatrix}\mathbf{u}\\ \hat{\mathbf{v}}^{(k-1)}\end{bmatrix}\right) \quad (25)$$

$$\hat{\mathbf{v}}^{(k)} = \arg\min_{\mathbf{v}\in\mathcal{E}^{N_2}}\mathcal{L}_{\gamma_2}\left(\hat{\mathbf{h}}^{(k)}, \begin{bmatrix}\hat{\mathbf{u}}^{(k)}\\ \mathbf{v}\end{bmatrix}\right) \quad (26)$$

$$\hat{\mathbf{s}}_N^{(k)} = \left[\left(\hat{\mathbf{u}}^{(k)}\right)^H\left(\hat{\mathbf{v}}^{(k)}\right)^H\right]^H \quad (27)$$

with $\gamma_1 = \lambda_{\min}((\mathcal{T}_N^{1\to N_1}(\hat{\mathbf{h}}))^H \mathcal{T}_N^{1\to N_1}(\hat{\mathbf{h}}))$ and $\gamma_2 = \lambda_{\min}((\mathcal{T}_N^{N_1\to N_1+N_2}(\hat{\mathbf{h}}))^H \mathcal{T}_N^{N_1\to N_1+N_2}(\hat{\mathbf{h}}))$. The relevance of the method, in this context, is demonstrated in Section VII.

VI. Adaptive Algorithms

In wireless communication, the channel is time-varying. Thus it is important to develop algorithms able to track those variations. This section is devoted to the derivation of an adaptive algorithm based on CMLBA$_\gamma$.

A. Weighted Criterion

The adaptivity property is obtained by introducing an exponential weighting factor into the definition of the criterion $\mathcal{L}_\lambda$. Let $\mathcal{L}_\lambda^{(W)}$ denote the weighted criterion defined as

$$\mathcal{L}_\lambda^{(W)}(\mathbf{h}, \mathbf{s}_N) = \sum_{t=n-N+1}^{t=n} \lambda^{n-t} \|\mathbf{X}_1(t) - \mathcal{T}_1(\mathbf{h})\mathbf{s}_1\|^2 - \gamma \|\mathbf{s}_N\|^2 \quad (28)$$

where $\lambda \in [0;1]$ is the forgetting factor, which ensures that old data are forgotten. Using a matrix formulation, we obtain

$$\mathcal{L}_\lambda^{(W)}(\mathbf{h}, \mathbf{s}_N) = \left\| \Lambda_N^{1/2} [\mathbf{X}_N(n+k) - \mathcal{T}_N(\mathbf{h})\mathbf{s}_N] \right\|^2 - \gamma \|\mathbf{s}_N\|^2 \quad (29)$$

where $\Lambda_N = \mathrm{diag}([\underbrace{1 \dots 1}_{L} \vdots \underbrace{\lambda \dots \lambda}_{L} \vdots \dots \vdots \underbrace{\lambda^{N-1} \dots \lambda^{N-1}}_{L}])$. Note that the forgetting factor is not applied to $\gamma \|\mathbf{s}_N\|^2$ since this term is related to the prior knowledge which is not time-dependent.

B. Derivation of CMLAAγ

The proposed algorithm is such that the updated estimates of channels and symbols at iteration k are calculated based on both their estimates at iteration $k-1$ and newly arrived data. The proposed approach was first presented by the authors in [10], where the recursive and adaptive versions of the MLBA are derived. The outlines of the method are recalled below, and the update equations for the CMLAAγ are then presented.

Let $\hat{\mathbf{h}}^{(k)}$ and $\hat{\mathbf{s}}_{N+k}^{(k)}(n+k)$ denote, respectively, the channel and the $N+M+k \times 1$ symbol vector estimated at iteration k. The adaptive algorithm is obtained from a growing window procedure after the following simplifications.

S1) The iterative minimization with respect to the joint variable is replaced by a minimization with respect to each variable separately. So, at step k, we compute

$$\hat{\mathbf{s}}_{N+k}^{(k)}(n+k) = \arg \min_{\mathbf{s}_{N+k} \in \mathcal{E}^{M+N+k}} \mathcal{L}_\gamma^{(W)}\left(\hat{\mathbf{h}}^{(k-1)}, \mathbf{s}_{N+k}\right) \quad (30)$$

$$\hat{\mathbf{h}}^{(k)} = \arg \min_{\mathbf{h}} \mathcal{L}_\gamma^{(W)}\left(\mathbf{h}, \hat{\mathbf{s}}_{N+k}^{(k)}(n+k)\right). \quad (31)$$

S2) At iteration k, (30) updates $M+N+k$ symbols. Hence, the computational complexity involved in (31) grows with k. Here, we propose to compute, at iteration k, the new emitted symbol, and we also update the next P (independent of k) symbols in the delay line where P is a crucial parameter to be determined. Implicitly, the previous symbols are supposed correctly estimated, which is often met since no decision device is introduced. Then (30) is replaced by

$$\begin{aligned}\hat{\mathbf{s}}_{P+1-M}^{(k)}&(n+k) \\ &= \arg \min_{Z \in \mathcal{E}^{P+1}} \mathcal{L}_\gamma^{(W)} \\ &\quad \times \left(\hat{\mathbf{h}}^{(k-1)}, \begin{bmatrix} \mathbf{Z} \\ \hat{\mathbf{s}}_0^{(k-1)}(n+k-P-1) \end{bmatrix}\right) \quad (32) \\ &= \arg \min_{Z \in \mathcal{E}^{P+1}} \Bigg\| \Lambda_{P+1}^{1/2} [\mathbf{X}_{P+1}(n+k) - \mathcal{T}_{P+1}(\mathbf{h}) \\ &\quad \times \begin{bmatrix} \mathbf{Z} \\ \hat{\mathbf{s}}_0^{(k-1)}(n+k-P-1) \end{bmatrix}\Bigg]\Bigg\|^2 \\ &\quad - \gamma \left\| \begin{bmatrix} \mathbf{Z} \\ \hat{\mathbf{s}}_0^{(k-1)}(n+k-P-1) \end{bmatrix} \right\|^2. \quad (33)\end{aligned}$$

Note that for the minimization problem in (33), the maximum value of γ is the minimum eigenvalue of $(\mathcal{T}_{P+1}^{1\to P+1})^H \mathcal{T}_{P+1}^{1\to P+1}$, where $\mathcal{T}_{P+1}^{1\to P+1}$ is the submatrix of $\mathcal{T}_{P+1}$ containing its P+1 first columns. Remember that, according to Theorem 2, $\lambda_{\min}(\mathcal{T}_{P+1}^{1\to P+1})^H \mathcal{T}_{P+1}^{1\to P+1}) \geq \lambda_{\min}(\mathcal{T}_{P+1}^H \mathcal{T}_{P+1})$. Then, even if $\mathcal{T}_{P+1}$ is badly conditioned, γ is not necessarily close to zero.

S3) The estimated channel $\hat{\mathbf{h}}^{(k)}$ is updated recursively from $\hat{\mathbf{h}}^{(k-1)}$, which is done without any approximation.

In the following, we derive the update equations for CMLAA$_\gamma$ (CML *adaptive* algorithm). We consider separately the minimization with respect to the symbols and the minimization with respect to the channel.

1) Minimization With Respect to the Symbols: The optimization problem in (33) is solved by a relaxation method. Let $s(j)$ denote the jth component of $\hat{\mathbf{s}}_{P+1-M}^{(k),(i)}(n+k)$, where k is the iteration number of CMLAA$_\gamma$ and i is the iteration number of the relaxation method; and let $T_{j,\lambda}^{(k-1)}$ denote the jth column of $\Lambda_{P+1}^{1/2} \mathcal{T}_{P+1}(\hat{\mathbf{h}}^{(k-1)})$. Then $\hat{\mathbf{s}}_{P+1-M}^{(k)}(n+k)$ is the stationary point obtained through the following iterative algorithm:

While $\|\hat{\mathbf{s}}_{P+1-M}^{(k),(i)}(n+k) - \hat{\mathbf{s}}_{P+1-M}^{(k),(i-1)}(n+k)\| > \epsilon$ do
  For $j = 1 : P+1$
    $A_j = T_{j,\lambda}^{(k-1)^H} T_{j,\lambda}^{(k-1)}$
    $B_j = T_{j,\lambda}^{(k-1)^H} \sum_{l=1, l\neq j}^{l=P+1+M} T_{l,\lambda}^{(k-1)} s(l)$
    $C_j = T_{j,\lambda}^{(k-1)^H} \Lambda_{P+1}^{1/2} \mathbf{X}_{P+1}(n)$
    $s(j) = -\mathbf{P}((B_j - C_j / A_j - \gamma))$
  end
end

2) Minimization With Respect to the Channel: The update of the filter for CMLAA$_\gamma$ can be performed using (34) shown at the bottom of the next page (for more details, see [10]), where $\mathbf{R}^{(i)} = \mathcal{U}(\hat{\mathbf{s}}_{N+i}^{(i)}(n+i))^H \Lambda_{N+i} \mathcal{U}(\hat{\mathbf{s}}_{N+i}^{(i)}(n+i))$. Simplifications of these equations are provided in [10], where the recursive least squares (RLS)-like algorithm above is turned into a least mean square (LMS)-like algorithm with almost no loss in performance.

C. Initialization

CMLAA$_\gamma$ needs of course a reliable initialization. A similar problem is encountered in TSML [6] or IQML [28]. Generally, the problem is solved by the use of an initialization procedure such as the subspace algorithm for example. Here, we propose to initialize the CMLAA$_\gamma$ with the solution given by the corresponding batch algorithm to a minimization problem over a block of size N. In any case, we take $N > 3M + 1$ to ensure that $(\tilde{\mathbf{h}}, \tilde{\mathbf{s}}_N)$ is the only global minimum of the considered criterion [see H2)].

D. Properties of CMLAA$_\gamma$

CMLAA$_\gamma$ exhibits some desirable properties for tracking the channel parameters in practical contexts, such as GSM, for example. This properties are summarized below.

- The introduction of the a priori knowledge into the criterion improves the convergence speed (see Section VII-B) of the algorithm as well as its tracking capabilities.
- The computational cost is moderate. Indeed, the most demanding part is the minimization with respect to the symbols. In CMLAA$_\gamma$, the length of the symbol vector to be estimated is P at each iteration (against $N + M$ in the CMLBA). We will see in the simulation part that a good choice for P is the channel order M. The estimation of the channel is performed by an LMS-like algorithm.
- CMLAA$_\gamma$ appears to be robust to the overestimation of the channel (see Section VII-B). This property is mandatory for using the proposed algorithm in practical applications.

E. Convergence of CMLAA$_\gamma$

In this section, we point out a result concerning the convergence of the CMLAA$_\gamma$ established for $\lambda = 1$ (which means that the adaptivity property is lost). The main result is formalized below.

Theorem 3: In the noiseless case, if CMLAA$_\gamma$ $(\lambda = 1)$ converges, if the assumptions H1) and H2) are met, and if for all k situated after the convergence $\hat{\mathbf{s}}_1^{(k)}(n + k) \neq \mathbf{0}_{M+1}$, then CMLAA$_\gamma$ converges toward the global minimum.

Proof: The proof is identical to that of [10, Theorem 3]. ■

This result is the consequence of the stability of the global minimum (instability of a local minimum) during a recursive procedure proved in [10, Theorem 2]. This result has first been established in the noiseless case. However, if we express the DML criterion as a function of the channel only $(\|\mathcal{T}_N(\mathbf{h})\mathbf{s}_N(n) - \mathbf{X}_N(n)\|^2 = \|\mathcal{P}^\perp \mathbf{X}_N(n)\|^2$ where $\mathcal{P}^\perp$ is the projection matrix on the range of $\mathcal{T}_N(\mathbf{h})$), it can be proved that when N tends to infinity, the noisy DML criterion tends to the noiseless one [29]. Then, the proof is also relevant in the noisy case as far as the number of data $(M + N)$ is large enough. In other words, the theorem states that the global minimum is the only stationary point in a recursive procedure.

VII. Simulation

To gain more insights about the results obtained in the previous sections, we present some numerical evaluations. The performance of the algorithms is measured by the normalized root mean square error in decibels

$$\mathrm{NRMSE}_{dB}(\mathbf{h}) = 20\log_{10}\left(\frac{1}{\|\tilde{\mathbf{h}}\|}\sqrt{\frac{1}{N_r}\sum_{i=1}^{N_r}\left\|\hat{\alpha}^{(i)}\hat{\mathbf{h}}^{(i)} - \tilde{\mathbf{h}}\right\|^2}\right)$$

where $\hat{\mathbf{h}}^{(i)}$ stands for the estimated channel from the ith trial, $\tilde{\mathbf{h}}$ is the true channel, and $\hat{\alpha}^{(i)} = \arg\min_\alpha \|\alpha\hat{\mathbf{h}}^{(i)} - \tilde{\mathbf{h}}\|^2$. N_r denotes the number of Monte Carlo runs. Noise samples are generated from i.i.d. zero-mean Gaussian random sequences with variance σ^2. The symbols belong either to a PSK modulation or to a 8-quadrature amplitude modulation (QAM). Figs. 4–7 concern the batch algorithm, whereas Figs. 8–10 concern the adaptive algorithms.

A. Batch Algorithms

In this section, we present some simulation studies of the MLBA and of CMLBA$_\gamma$ where only the extreme values of γ are considered (i.e., $\gamma = 0, \gamma = \lambda_{\min}(\mathcal{T}_N(\hat{\mathbf{h}})^H\mathcal{T}_N(\hat{\mathbf{h}}))$). These two algorithms will be referred to as CMLBA$_{\min}$ and CMLBA$_{\max}$, respectively. The performance of these algorithms is compared against those of TSML [6], the multistep linear prediction algorithm (MLPA) [30], and the (joint order detection and channel estimation via least squares smoothing (J-LSS) [31]. We also consider the method described in Section V. We call it CMLBA$_{\max}(N_P)$, where N_P stands for the number of subvectors of $\mathbf{s}_N$ updated separately [for example, the algorithm in (25) and (26) is CMLBA$_{\max}$ (2)].

1) Description of the Multipath Channels: In our simulations, we considered the following channels commonly used in the literature.

$$\begin{cases}\hat{\mathbf{h}}^{(i)} = \hat{\mathbf{h}}^{(i-1)} + \left[\mathbf{R}^{(i)}\right]^{-1}\Big\{\mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)^H \Lambda_{P+1}\left[\mathbf{X}_{P+1}(n+i) - \mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)\hat{\mathbf{h}}^{(i-1)}\right] \\ -\lambda\mathcal{U}\left(\hat{\mathbf{s}}_P^{(i-1)}(n+i-1)\right)^H \Lambda_P\left[\mathbf{X}_P(n+i-1) - \mathcal{U}\left(\hat{\mathbf{s}}_P^{(i-1)}(n+i-1)\right)\hat{\mathbf{h}}^{(i-1)}\right]\Big\} \\ \left[\mathbf{R}^{(i)}\right]^{-1} = \mathbf{A} + \mathbf{A}\mathcal{U}\left(\hat{\mathbf{s}}_P^{(i-1)}(n+i-1)\right)^H\Big[\frac{\Lambda_P}{\lambda} \\ -\mathcal{U}\left(\hat{\mathbf{s}}_P^{(i-1)}(n+i-1)\right)\mathbf{A}\mathcal{U}\left(\hat{\mathbf{s}}_P^{(i-1)}(n+i-1)\right)^H\Big]^{-1} \times \mathcal{U}\left(\hat{\mathbf{s}}_P^{(i-1)}(n+i-1)\right)\mathbf{A} \\ \mathbf{A} = \left[\lambda\mathbf{R}^{(i-1)}\right]^{-1} - \left[\lambda\mathbf{R}^{(i-1)}\right]^{-1}\mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)^H \times \Big[\Lambda_{P+1} \\ +\mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)\left[\lambda\mathbf{R}^{(i-1)}\right]^{-1}\mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)^H\Big]^{-1} \times \mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)\left[\lambda\mathbf{R}^{(i-1)}\right]^{-1}\end{cases} \tag{34}$$

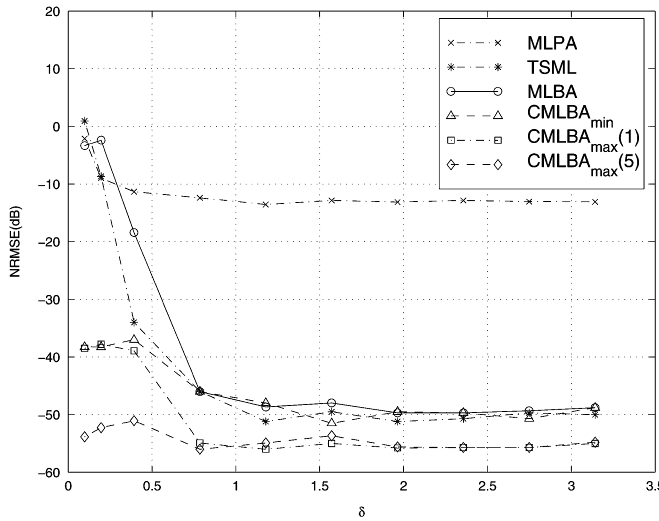

Fig. 4. Performance comparison versus δ (100 Monte Carlo runs, $M+N=32$ BPSK input symbols).

(a)

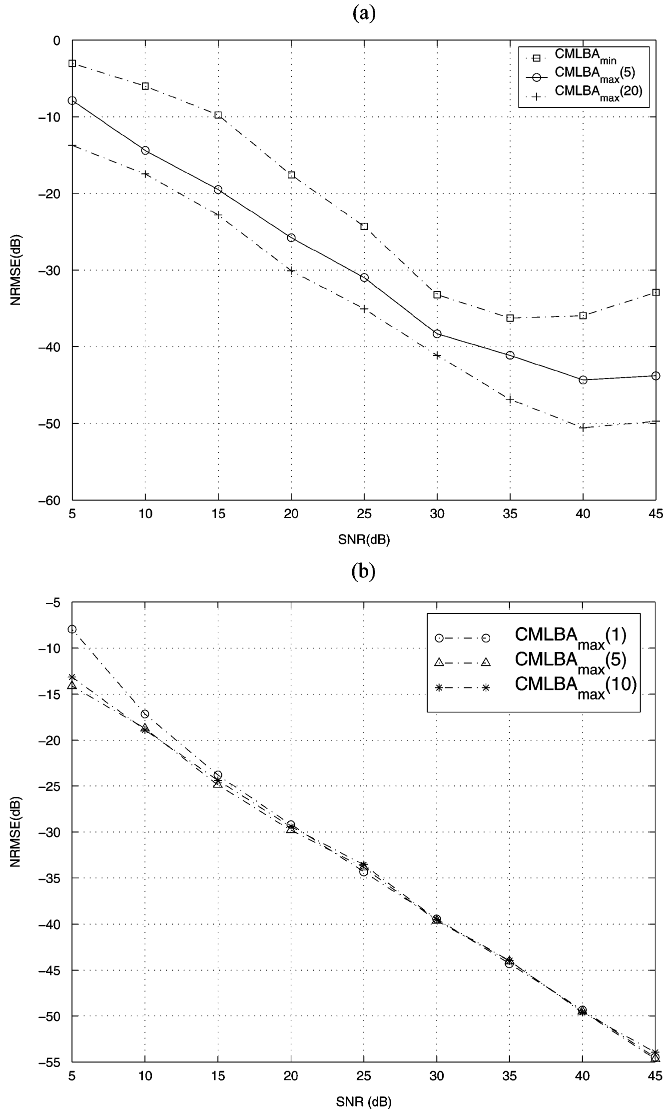

Fig. 5. Performance comparison. (a) $\mathbf{h}^{\mathrm{Hua}}$, 100 Monte Carlo runs, $M+N=58$ BPSK input symbols. (b) $\mathbf{h}^{\mathrm{Xu}}$, 100 Monte Carlo runs, $M+N=50$ QPSK input symbols.

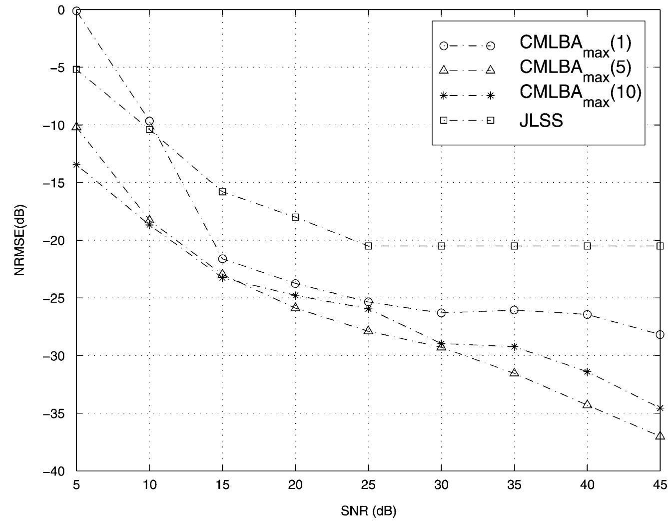

Fig. 6. Performance comparison. $\mathbf{h}^{\mathrm{Tong}}$, 100 Monte Carlo runs, $M+N=100$ QPSK input symbols.

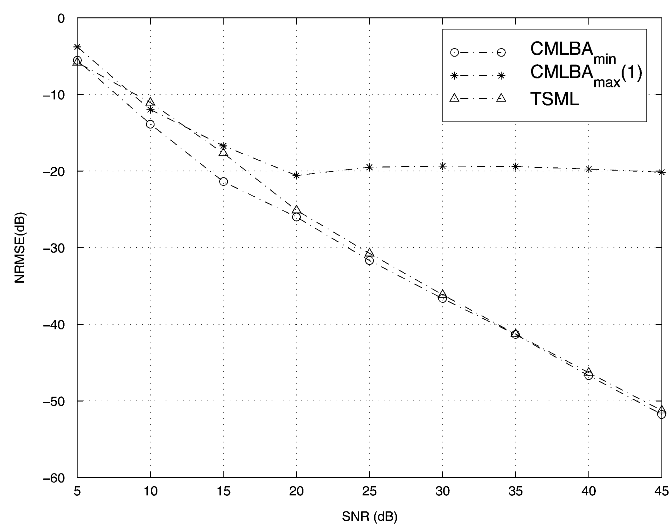

Fig. 7. Performance comparison. $\mathbf{h}^{\mathrm{Hua}}$, 100 Monte Carlo runs, $M+N=58$ 8-QAM input symbols.

- $\mathbf{h}^{\mathbf{Hua}}$: This two-channel system was first used by Hua [6]. The corresponding channel response is given by

$$\tilde{\mathbf{h}}_1^{\mathbf{Hua}}(z) = (1 - e^{j\theta_1}z^{-1})(1 - e^{-j\theta_1}z^{-1})$$
$$\tilde{\mathbf{h}}_2^{\mathbf{Hua}}(z) = \left(1 - e^{j(\theta_1+\delta)}z^{-1}\right)\left(1 - e^{-j(\theta_1+\delta)}z^{-1}\right) \quad (35)$$

where θ_1 and $\theta_1+\delta$ represent the angular position of zeros on the unit circle and δ is the distance between the zeros of the two channels. We choose to use this channel since it permits to evaluate the influence of the channel diversity. Secondly, Hua also used this channel to compare the performance of the TSML against the performance of the cross-relation algorithm [32] and of the subspace algorithm [33], and he compared both algorithms to the Cramer–Rao bound. Moreover, Zhao also used this channel to evaluate the performance of the adaptive least squares smoothing algorithm [34]. Then, by using the experimental conditions in [34], we make a fair comparison with existing approaches.

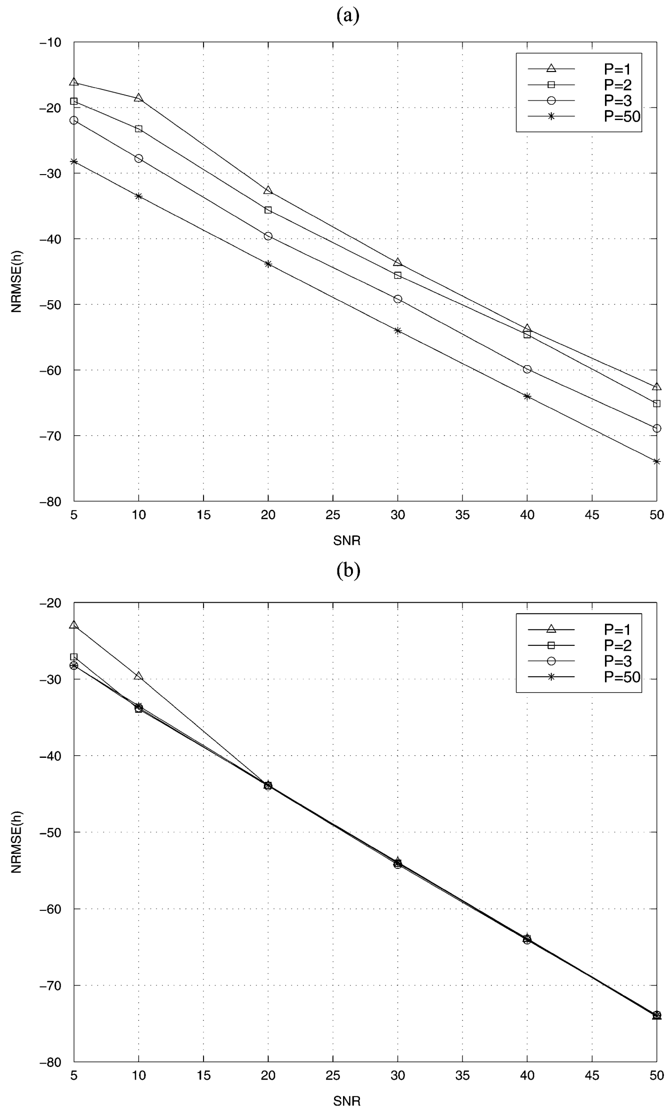

Fig. 8. NRMSE($\mathbf{h}$) versus SNR obtained with (a) $\mathrm{CMLAA}_{\min}$ and with (b) $\mathrm{CMLAA}_{\max}$ for various values of P $(\lambda = 1), N = 50$.

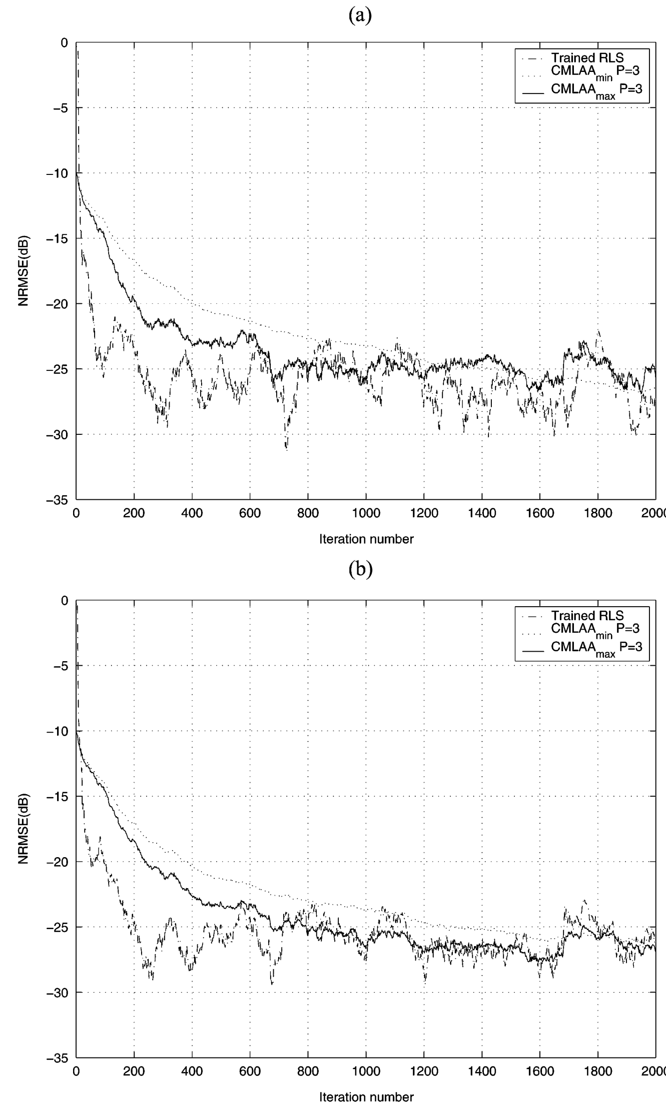

Fig. 9. NRMSE($\mathbf{h}$) versus the iteration number. SNR $= 10$ dB. Channel order (a) correctly estimated and (b) overestimated.

- $\mathbf{h}^{\mathbf{Xu}}$: The channel responses values are given in Table I. This set of channels was first used in [35]. It considers a two-ray multipath model with delay at 0 and 1.1 baud periods. The channel model $\mathbf{h}^{\mathbf{Xu}}$ simulates a wireless environment with long delay multipath.
- $\mathbf{h}^{\mathbf{Tong}}$: The channel response values are given in Table II. The channel $\mathbf{h}^{\mathbf{Tong}}$ was first used in [31]. This channel has severe intersymbol interference. It is also close to violate the identifiability condition. Moreover, $\mathbf{h}^{\mathbf{Tong}}$ has small head and tails taps. This channel is used to test the robustness of our algorithms against the overestimation of the channel order.

2) Comments and Observations:

- Fig. 4 plots the $\mathrm{NRMSE}_{\mathrm{dB}}(h)$ against δ (relative positions of the zeros), the SNR is set to 45 dB. This SNR is quite unrealistic; however, this choice permits a fair comparison between our algorithms and TSML since TSML is biased at low SNR whereas MLBA-like algorithms are not. We also compare these methods with the MLPA [30]. We observe the following.
  - — The TSML and the MLBA have comparable performance (NRMSE≈ -50 dB when $\delta > 0.7$ and poor performance when $\delta < 0.7$).
  - — The MLPA is a generalization of the linear prediction algorithm. The MLPA exploits the channel structure completely and provides more statistical efficiency in channel identification (compared to classical LPA [28], [36]). This method has also been developed as the outer-product decomposition algorithm [37], [38]. It was also extended by Tugnait *et al.* for multiple-input multiple-output [39]. The MLPA presents poor results compared to the other methods. Actually, the simulation was run with $M + N = 32$ input symbols, whereas the MLPA exhibits good results with hundreds of input symbols. For short-burst applications, LPA-like algorithms cannot be used.
  - — The $\mathrm{CMLBA}_{\min}$ outperforms TSML and MLBA for poor diversity conditions. Remember that $\mathrm{CMLBA}_{\min}$

TABLE I
CHANNEL RESPONSE OF $\mathbf{h}^{\mathrm{Xu}}$

i	$\mathbf{h}_i^{\mathbf{Xu}}(0)$	$\mathbf{h}_i^{\mathbf{Xu}}(1)$	$\mathbf{h}_i^{\mathbf{Xu}}(2)$	$\mathbf{h}_i^{\mathbf{Xu}}(3)$
1	0	-1.280 - 0.301j	1.617 + 2.385j	0.178 + 0.263j
2	-1.023 - 0.501j	0.106 + 1.164j	1.477+1.850j	-0.482 - 0.523j
3	0	-0.282 + 0.562j	0.371 - 1.001j	0.041 - 0.110j
4	-0.227 + 0.487j	0.031 - 0.211j	0.336 - 0.866j	-0.110 + 0.271j

TABLE II
CHANNEL RESPONSE OF $\mathbf{h}^{\mathrm{Tong}}$

i	$\mathbf{h}_i^{\mathbf{Tong}}(0)$	$\mathbf{h}_i^{\mathbf{Tong}}(1)$	$\mathbf{h}_i^{\mathbf{Tong}}(2)$	$\mathbf{h}_i^{\mathbf{Tong}}(3)$	$\mathbf{h}_i^{\mathbf{Tong}}(4)$	$\mathbf{h}_i^{\mathbf{Tong}}(5)$
1	-0.0031 - 0.0017j	-0.0109 - 0.0025j	0.1522 + 0.0705j	0.3789 + 0.5930j	-0.0301 - 0.0348j	-0.0032 - 0.0017j
2	-0.0016 - 0.0047j	-0.0263 - 0.0433j	0.4409 + 0.4736j	0.0766 + 0.2168j	-0.0042 - 0.0154j	-0.0017 - 0.0044j

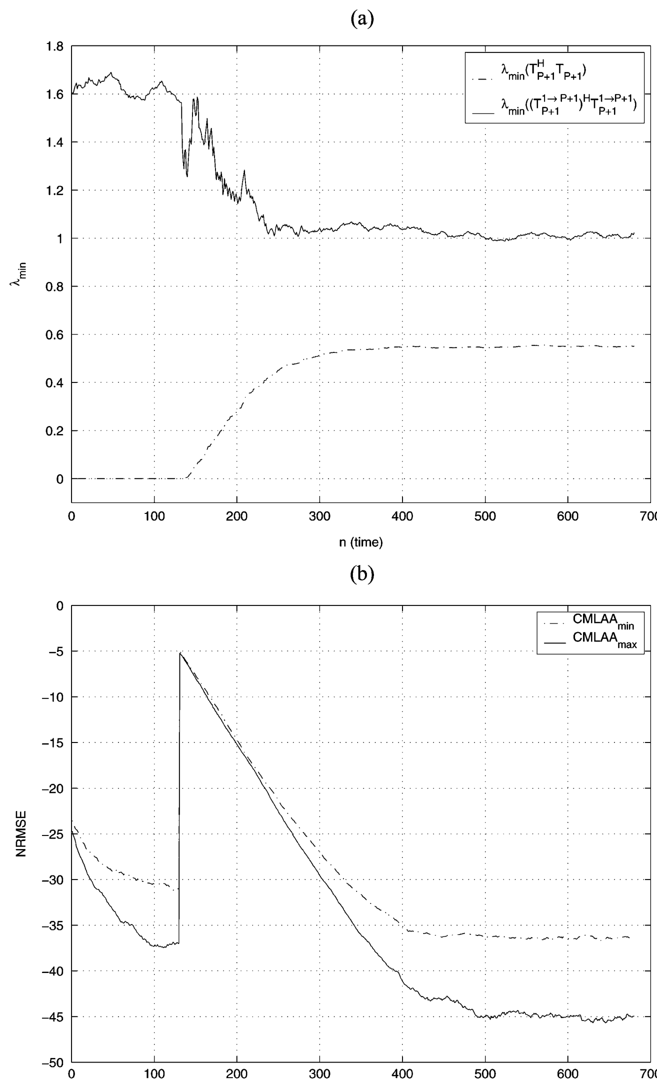

Fig. 10. Channel tracking performance (SNR $= 30$ dB, 20 Monte Carlo runs).

has the same local minima as MLBA. Thus this improvement (in terms of estimation accuracy) is not counterbalanced by extra local minima.

— The NRMSE of CMLBA$_{\max}$ is about 5 dB less than the NRMSE of CMLBA$_{\min}$, of MLBA, and of TSML when δ is large enough (thanks to the prior introduced by γ). When $\delta \rightarrow 0$, then $\gamma = \lambda_{\min} \rightarrow 0$ and CMLBA$_{\max} \rightarrow$ CMLBA$_{\min}$. We are not surprised to see that the NRMSE of CMLBA$_{\max}$ gets closer to the NRMSE of CMLBA$_{\min}$ when δ tends toward zero.

— CMLBA$_{\max}$ (5) was built to achieve good performance even when the channels share common zeros. We observe that the NRMSE of the CMLBA$_{\max}$ (5) is about -55 dB whatever δ may be.

- Fig. 5 shows a comparison of the methods in various environments. In (a), $\mathbf{h}^{\mathbf{Hua}}$ is considered with $\delta = \pi/100$ and $\theta = 0$. For such a channel the smallest eigenvalue of $\mathcal{T}_N(\tilde{\mathbf{h}})^H\mathcal{T}_N(\tilde{\mathbf{h}})$ is close to zero; then CMLBA$_{\min}$ and CMLBA$_{\max}$ are identical and the prior knowledge is lost (we do not plot the curve corresponding to CMLBA$_{\max}$). First, we can remark that the constraint of CMLBA$_{\min}$ yields robustness to the lack of channel diversity. It is interesting to note that this improvement arises without increasing the number of local minima (see Proposition 1). The minimal eigenvalue is likely to increase when the partition grows. As expected, CMLBA$_{\max}$ (20) outperforms CMLBA$_{\max}$ (5) and CMLBA$_{\min}$ even at low SNR.

  We repeat the above simulation using the set of channels $\mathbf{h}^{\mathbf{Xu}}$. This time, the smallest eigenvalue of $\mathcal{T}_N(\tilde{\mathbf{h}})^H\mathcal{T}_N(\tilde{\mathbf{h}})$ is 0.2737. Thus, in CMLBA$_{\max}$ (1), the value of $\lambda_{\min}$ is no longer negligible, which leads to very good estimates of the channel. In such an environment, the partitions do not bring improvements.
- Fig. 6 shows robustness to the overestimation of the channel order and to severe intersymbol interference. The channel used for this simulation is $\mathbf{h}^{\mathbf{Tong}}$; it presents severe interference intersymbol. Moreover, the first and last taps of each subchannel have very small amplitude. We choose to run the CMLBA$_{\max}$ with $M = 5$ to test the robustness of the method toward an overestimation of the channel order. The experimental conditions are identical to those in [31]. As can be seen, the method presents good results and outperforms J-LSS. The channel $\mathbf{h}^{\mathrm{Tong}}$

TABLE III
PATH PROFILE FOR CHANNEL EQUALIZATION TESTS (COST-GSM MODEL)

Path	1	2	3	4	5	6	7	8	9	10	11	12
Delay (μs)	0	0.2	0.4	0.6	0.8	1.2	1.4	1.8	2.4	3	3.2	5
Attenuation (dB)	-4	-3	0	-2	-3	-5	-7	-5	-6	-9	-11	-10

is ill conditioned; thus the quality of the estimation is improved by the partitions.

- Fig. 7 shows the behavior of the method with a non-PSK constellation. Throughout this paper, we have supposed that the emitted symbols belong to a PSK modulation H4). This assumption was at the origin of the prior knowledge introduced in the proposed criterion. The efficiency of the prior (for PSK constellations) was exhibited in the previous simulations, but what happens when a non-PSK constellation is used? Here, the input symbols belong to an 8-QAM constellation. We use $\mathbf{h}^{\mathbf{Hua}}$ with $\delta = 3\pi/8$ and $\theta = 0, M + N = 58$. Not surprisingly, it appears that $\text{CMLBA}_{\max}$ is not convenient for 8-QAM constellations. TSML and $\text{CMLBA}_{\min}$ present similar performances. However, (4) suggests that $\text{CMLBA}_{\min}$ is better to use especially for ill conditioned channels.

B. Adaptive Algorithms

In order to check the relevance of the approximations used to derive the CMLAA, we analyze the choice of the parameter P (number of symbols updated at each iteration). Then, we focus on the applicability of the proposed algorithm first by testing its robustness to an overestimation of the channel order (Fig. 9) and second by evaluating its parameter tracking performance (Fig. 10).

Fig. 8 shows the choice of P. In this simulation, the NRMSE is computed for $\text{CMLAA}_{\min}$ and for $\text{CMLAA}_{\max}$ (with $\lambda = 1$) for various pairs (P, SNR). The NRMSE is averaged over 50 Monte Carlo runs and is computed at the one-thousandth iteration. In (b), the value of the mean squared error obtained with $P = 50$ remains very close to the value obtained with $P = 3$. In (a), we observe that replacing $P = 3$ by $P = 50$ leads to the same improvement as replacing $P = 1$ by $P = 3$. Therefore, the accuracy of the channel estimate seems to be mainly influenced by the symbols in the delay line of the channels. Moreover, the computational cost of the method increases with P. In our simulation, the order of the channel is $M = 3$. Choosing P equal to the channel order appears to be a good compromise.

Fig. 9 shows robustness to channel overestimation. A 12-path propagation channel is considered, simulated according to the model of Clarke [40]. The path profile is shown in Table III. For this simulation, the channel order is set to $M = 3$, there are two subchannels, and the modulation is BPSK. The SNR, this time, is set to a more reasonable value of 10 dB.

We present the NRMSE versus the iteration number for $\text{CMLAA}_{\min}$, for $\text{CMLAA}_{\max}$, and for trained RLS. Fig. 9(a) shows the results obtained when the channel order is correctly estimated ($\hat{M} = 3$). From this figure, we can see that $\text{CMLAA}_{\max}$ outperforms $\text{CMLAA}_{\min}$ even if the algorithm does not perform a minimization with respect to the joint variable in each iteration (see Section VI-B). We can notice that the behavior of both algorithms could be improved by iterating the minimization with respect to each variable. Fig. 9(b) shows the results obtained when the channel order is overestimated ($\hat{M} = 4$). Both $\text{CMLAA}_{\min}$ and $\text{CMLAA}_{\max}$ appear to be robust to the overestimation of the channel order. This is due to the constraint introduced into the criterion.

Fig. 10 shows tracking of the channel parameters. The simulation presented here has first been experimented with by Zhao in [34]. It concerns the case where the channel order and the channel parameters have a sudden change. The initial channel is the one used for testing the batch algorithms (35) with $\delta = \pi$ and $\theta_1 = \pi/10$. The channel order and the channel parameters change at time $n = 151$ where we add zeros $Z_1 = 1$ and $Z_2 = -1$ to the two subchannels, respectively. In the simulation, the estimated channel order is fixed to $\hat{M} = 3$. The NRMSE convergence of $\text{CMLAA}_{\min}$ and of $\text{CMLAA}_{\max}$ is shown in Fig. 10(b), where we can see the ability of both algorithms to track the channel variations. In (a), the smallest eigenvalues of the full matrix $(\lambda_{\min}(\mathcal{T}_{P+1}^H \mathcal{T}_{P+1}))$ and of the truncated matrix $(\lambda_{\min}((\mathcal{T}_{P+1}^{1\to P+1})^H \mathcal{T}_{P+1}^{1\to P+1}))$ are plotted. When $n < 151$, we try to estimate a channel of order 3, whereas the order of the channel to be estimated is 2; that is the reason why $\lambda_{\min}(\mathcal{T}_{P+1}^H \mathcal{T}_{P+1}) \approx 0$. In $\text{CMLAA}_{\max}, \gamma$ is equal to $\lambda_{\min}((\mathcal{T}_{P+1}^{1\to P+1})^H \mathcal{T}_{P+1}^{1\to P+1})$, which is not null. $\text{CMLAA}_{\max}$ takes more advantage of the prior information than $\text{CMLAA}_{\min}$ even when the subchannels share common zeros.

VIII. CONCLUSION

In this paper, a maximum likelihood approach to solve the joint blind channel identification and blind symbol estimation problem was presented. We demonstrated the improvement of the estimation accuracy by the use of a prior knowledge. Moreover, the proposed batch algorithm presents the finite-sample convergence property.

Based on this block algorithm, an adaptive version is derived by exploiting the recursive procedure proposed for solving the local minima problem. A nice advantage inherent to the use of the prior is that it brings robustness to the overestimation of the channel order thus our method does not require the channel order to be known or well estimated. Thanks to the forgetting factor, the algorithm is able to track efficiently the changes in the channel parameters. At each iteration, the number of symbols to be updated is limited to the length of the channel. Moreover, the update of the filters can be performed by stochastic gradient techniques as shown in [10], which renders the CMLAA computationally nonexpensive. Current work on the application of these algorithms under practical situations is currently undertaken, and will be reported.

REFERENCES

[1] Z. Ding and Y. Li, *Blind Equalization and Identification*, ser. Signal Processing and Communications Series: Marcel Dekker, 2001.

[2] J. K. Tugnait, "Blind equalization techniques," in *Encyclopedia of Telecommunications*. New York: Wiley, 2002.

[3] G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, "Channel estimation and equalization using higher-order statistics," in *Signal Processing Advances in Wireless and Mobile Communications: Trends in Channel Estimation and Equalization*. Englewood Cliffs, NJ: Prentice-Hall, 2000.

[4] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channels," in *Proc. 25th Asilomar Conf. Signals, Systems, and Computers*, Nov. 1991, pp. 856–860.

[5] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, pp. 1951–1968, 1998.

[6] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. Signal Process.*, vol. 44, pp. 661–672, Mar. 1996.

[7] D. Slock and C. B. Papadias, "Further results on blind identification and equalization of multiple FIR channels," in *Proc. ICASSP*, 1995, pp. 1964–1967.

[8] M. Feder and J. A. Catipovic, "Algorithms for joint channel estimation and data recovery-application to equalization in underwater communications," *IEEE J. Ocean. Eng.*, vol. 16, pp. 42–55, Jan. 1991.

[9] D. Gesbert, P. Duhamel, and S. Mayrargue, "Blind least-squares approaches for joint data/channel estimation," in *IEEE DSP Workshop*, Sept. 1996.

[10] F. Alberge, P. Duhamel, and M. Nikolova, "Adaptive solution for blind identification/equalization using deterministic maximum likelihood," *IEEE Trans. Signal Process.*, vol. 50, pp. 923–936, Apr. 2002.

[11] N. Seshadri, "Joint data and channel estimation using fast blind treillis search techniques," in *Proc. Globecom*, 1990, pp. 1659–1663.

[12] M. Ghosh and C. L. Weber, "Maximum-likelihood blind equalization," *Opt. Eng.*, vol. 31, no. 6, pp. 1224–1228, Jun. 1992.

[13] S. Talwar, M. Viberg, and A. Paulraj, "Blind estimation of multiple co-channel digital signals using an antenna array," *IEEE Signal Process. Lett.*, vol. 1, pp. 29–31, Feb. 1994.

[14] M. Nikolova, "Estimation of binary images by minimizing convex criteria," in *Proc IEEE Int. Conf. Image Processing*, vol. 2, Oct. 1998, pp. 108–112.

[15] W. K. Ma, T. N. Davidson, K. M. Wong, Z.-Q. Luo, and P. C. Ching, "Quasi maximum likelihood multiuser detection using semi-definite relaxation with application to synchronous CDMA," *IEEE Trans. Signal Process.*, vol. 50, no. 4, pp. 912–922, 2002.

[16] P. H. Tan, L. K. Rasmussen, and T. J. Lim, "Constrained maximum likelihood detection in CDMA," *IEEE Trans. Commun.*, vol. 49, no. 1, pp. 142–153, 2001.

[17] R. E. Blahut, *Algebraic Methods for Signal Processing and Communication Coding*. New York: Springer-Verlag, 1992.

[18] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[19] E. Pité and P. Duhamel, "Bilinear methods for blind channel equalization: (no) local minimum issue," in *Proc. ICASSP*, May 1998, pp. 2113–2116.

[20] C. B. Papadias, "Methods for blind equalization and identification of linear channels," Ph.D. dissertation, ENST, Paris, France, 1995.

[21] K. G. Murty and S. N. Kabadi, "Some NP-complete problems in quadratic and nonlinear programming," *Math. Program.*, vol. 39, pp. 117–129, 1987.

[22] P. M. Pardalos and G. Schnitger, "Checking local optimality in constrained quadratic programming is np-hard," *OR Lett.*, vol. 7, pp. 33–35, 1988.

[23] R. Horst and H. Tuy, *Global Optimization (Deterministic Approaches)*. Berlin, Germany: Springer-Verlag, 1993.

[24] I. Bomze and G. Danninger, "A finite algorithm for solving general quadratic problem," *J. Global Opt.*, vol. 4, pp. 1–16, 1994.

[25] O. Barrientos and R. Correa, "An algorithm for global minimization of linearly constrained quadratic functions," *J. Global Opt.*, vol. 16, pp. 77–93, 2000.

[26] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1991.

[27] P. G. Ciarlet, *Introduction to Numerical Linear Algebra and Optimization*: Cambridge Univ. Press, 1989.

[28] D. Slock, "Blind fractionaly-spaced equalization, perfect reconstruction filter banks, and multichannel linear prediction," in *Proc. ICASSP*, 1994, pp. 585–588.

[29] E. de Carvalho, "Identification de canal et egalization aveugles et semi-aveugles pour les communications mobiles," Ph.D. dissertation, ENST, Paris, France, 1999.

[30] D. Gesbert and P. Duhamel, "Robust blind channel identification and equalization based on multi-step predictors," in *Proc. ICASSP*, Apr. 1997, pp. 3621–3624.

[31] L. Tong and Q. Zhao, "Joint order detection and blind channel estimation by least squares smoothing," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2345–2355, 1999.

[32] H. Liu, G. Xu, and L. Tong, "A deterministic approach to blind channel identification of multi-channel FIR systems," in *Proc. ICASSP*, Apr. 1994.

[33] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Process.*, vol. 43, pp. 516–526, Feb. 1995.

[34] Q. Zhao and L. Tong, "Adaptive blind channel estimation by least squares smoothing," *IEEE Trans. Signal Process.*, vol. 47, pp. 3000–3012, Nov. 1999.

[35] G. Xu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2982–2993, 1995.

[36] K. A. Meraim, P. Duhamel, D. Gesbert, and P. Loubaton, "Prediction error methods for time-domain blind identification of multichannel FIR filters," in *Proc. ICASSP*, May 1995.

[37] Z. Ding, "An outer-product decomposition algorithm for multichannel blind identification," in *Proc. 8th IEEE Workshop Stat. Signal Array Processing*, June 1996, pp. 132–135.

[38] ——, "Matrix outer-product decomposition method for blind multiplechannel identification," *IEEE Trans. Signal Process.*, vol. 45, pp. 3053–3061, Dec. 1997.

[39] J. K. Tugnait and B. Huang, "Multistep linear predictors-based blind identification and equalization of multiple-input multiple output channels," *IEEE Trans. Signal Process.*, vol. 48, pp. 26–38, Jan. 2000.

[40] R. H. Clarke, "A statistical theory of mobile radio reception," *Bell Syst. Tech. J.*, 1968.

Florence Alberge was born in Albi, France, in 1971. She received the engineering degree from Ecole Nationale Supérieure de l'Electronique et de ses Applications, France, and the master's degree from the University of Cergy, France, in 1996. She received the Agregation de Physique and Ph.D. degrees from the Ecole Nationale Supérieure des Télécommunications, Paris, France, in 1998 and 1999, respectively.

Since 2000, she has been with the Laboratoire des Signaux et Systèmes, Gif sur Yvette, France, and with the University of Paris-Sud, Orsay, France, as an Assistant Professor. Her main research interests are in the field of signal processing for communications.

Mila Nikolova is a Researcher with the National Center for Scientific Research (CNRS), France. She is also with the Center for Mathematics and their Applications (CMLA), ENS de Cachan, France. Her research interests include inverse problems, mathematical image and signal processing, and variational problems and their analysis.

Pierre Duhamel (SM'87–F'98) was born in France in 1953. He received the Eng. degree in electrical engineering from the National Institute for Applied Sciences (INSA) Rennes, France, in 1975 and the Dr.Eng. and Doctorat ès Sciences degrees from Orsay University, Orsay, France, in 1978 and 1986, respectively.

From 1975 to 1980, he was with Thomson-CSF, Paris, France, where his research interests were in circuit theory and signal processing, including digital filtering and analog fault diagnosis. In 1980, he joined the National Research Center in Telecommunications (CNET), Issy les Moulineaux, France, where his research activities were first concerned with the design of recursive charge-coupled device filters. Later, he worked on fast algorithms for computing Fourier transforms and convolutions and applied similar techniques to adaptive filtering, spectral analysis, and wavelet transforms. From 1993 to 2000, he was a Professor with Ecole Nationale Supérieure des Télécommunications, Paris, with research activities focused on signal processing for communications. He was Head of the Signal and Image Processing Department from 1997 to 2000. He is now with the Laboratoire de Signaux et Systemes, Gif sur Yvette, France, where he is developing studies in signal processing for communications (including equalization, iterative decoding, multicarrier systems) and signal/image processing for multimedia applications, including source coding, joint source/channel coding, watermarking, and audio processing.

Dr. Duhamel was Co-General Chair of the 2001 International Workshop on Multimedia Signal Processing, Cannes, France, and will be Co-Technical Chair of ICASSP'06, Toulouse, France. He was Chairman of the IEEE DSP Committee from 1996 to 1998 and a member of the SP for Com Committee until 2001. He was an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1989 to 1991, an Associate Editor of IEEE SIGNAL PROCESSING LETTERS, and a Guest Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING Special Issue on Wavelets. He was an IEEE Distiguished Lecturer for 1999. He was a Corecipient of the Best paper Award from the IEEE TRANSACTIONS ON SIGNAL PROCESSING in 1998. He received the Grand Prix France Telecom award from the French Science Academy in 2000.