Adaptive Solution for Blind Identification/Equalization Using Deterministic Maximum Likelihood

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

***Abstract*—A deterministic maximum likelihood (DML) approach is presented for the blind channel estimation problem. It is first proposed in a block version, which consists of iterating two steps, each one solving a least-squares problem either in the channel or in the symbols. In the noiseless case and under certain conditions, this algorithm gives the exact channel and the exact symbol vector with a finite number of samples. It is shown that even if the DML method has a single global minimum, the proposed iterative procedure can converge to spurious local minima. This problem can be detected (under some channel diversity conditions) by using a numerical test that is proposed in the paper.**

Based on these considerations, we extend the maximum likelihood block algorithm (MLBA) to recursive implementations [maximum likelihood recursive algorithm (MLRA)]. The MLRA is able to track variations of the system by the introduction of an exponential forgetting factor in the DML criterion. The link between the adaptive algorithm and a soft decision feedback equalizer (SDFE) is emphasized. Low-complexity versions of the recursive and adaptive algorithm are presented.

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

I. Introduction

BLIND identification/equalization is an important problem in wireless communications, either in a passive listening situation or in fast fading environments. Blind techniques present some advantages compared with the traditional training methods. 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, and thus, training sequences are not exploitable. However, when some symbols are known, "semi-blind" techniques are preferred since they are able to track system variations much more efficiently than algorithms based only on training sequences while having performance much enhanced compared with fully blind algorithms. When the known symbols are grouped, maximum likelihood (ML) permits us to obtain easily a semi-blind criterion [1]. Therefore, we choose to focus on blind criteria.

Manuscript received June 1, 2000; revised September 13, 2001. The associate editor coordinating the review of this paper and approving it for publication was Prof. Michail K. Tsatsanis.

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

M. Nikolova is with the ENST/TSI and CNRS URA 820, Paris, France (e-mail: nikolova@tsi.enst.fr).

Publisher Item Identifier S 1053-587X(02)02392-9.

The earlier approaches to blind equalization were based on higher order statistics of the received signal [2]–[5]. 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 a fast varying environment is intrinsically limited. Tong *et al.* suggested a different option [6]. They make use of time or spatial diversity at the output of the channel, which is obtained when the measured samples are oversampled or when several antennas are used. Thus, the considered system is a single input/multiple output system (SIMO). The SIMO equalization problem can be solved using second-order statistics only, as long as the subchannels do not share common zeros. Second-order techniques have the potential for estimating the required statistics with fewer samples; hence, they do not have the intrinsic limitations of higher order ones. However, whether a given method converges faster than another one relies on other properties.

Some second-order methods rely on assumptions on the statistics of the input sequence (usually, an assumption of the sequence to be white) [7]–[10]. In a fast fading environment, if only a few data samples corresponding to the same channel characteristics are available, then the statistical estimate is not reliable. In that case, the problem may be solved by treating the input as a deterministic variable. This paper focuses on this situation: The input sequence is considered as a deterministic parameter to be identified. More precisely, this paper deals with deterministic maximum likelihood (DML) methods. The good property of DML methods in a SIMO context relies on the fact that it can be obtained through a sequence of least-squares problems, as we will see. However, their main drawback lies in the difficulty to express the estimator in closed form and in the presence of local minima. This is partially solved here in the context of DML.

Among the major contributions to DML methods, we can cite the work of Hua, who proposed in [11] the two-step maximum likelihood (TSML) method. The TSML method establishes a connection between the cross relation (CR) method, which belongs to subspace methods, and the ML estimator. Around the same time, Slock developed a method denoted the iterative quadratic maximum likelihood (IQML) method [12], [13], which is similar to the TSML method. Both methods iterate two steps to estimate the channel. The performance comparison with the Cramér-Rao bound has been obtained in [11], [14], and [15]. Others DML methods are available in [16] and the references therein. DML algorithms are capable of obtaining perfect channel estimation within a finite number of samples

in the absence of noise. On the other hand, TSML and IQML give biased channel estimates [17] and may behave poorly at low SNR, even with an asymptotical number of data. Ayadi proposed, in [18], two solutions to remedy this situation, which lead to "denoised" ML algorithms. Unfortunately, these deterministic algorithms have been developed for batch processing, and their adaptive implementations are often cumbersome. The approach proposed in this paper has the attractive properties of DML with, at the same time, a structure suitable for recursive and adaptive implementation.

This paper is built on a previous work performed in the same team [19], where a block algorithm (MLBA) was proposed. Each step of the MLBA solves a least-squares problem alternately in the channel and in the symbols, whereas in previous contributions, either the symbols or the channels were not computed explicitly during the iterations. Alternating methods have also been proposed in contributions like [20] or [21], where the finite alphabet property is used. In the absence of noise, the MLBA estimates the channels and symbols perfectly, using a finite number of symbols, and the MLBA has a single global minimum [19]. Some recursive versions were proposed, including decision devices. The behavior of the algorithm was proved to be very similar to a DFE. Pité wrongly stated in [22] that the MLBA does not admit local minima in the noiseless case. Unfortunately, the implementation of the MLBA is complicated by the existence of local minima even with noiseless data. The novelty of this paper is twofold.

- A numerical test is proposed to circumvent the local minima problem. Actually, we combine the iterative algorithm with a growing window technique, and we show that under the classical assumptions of channel diversity and sufficient excitation of the symbols, we are able to check whether the obtained stationary point is the global minimum or a spurious local minimum. This property can be extended to noisy data when a large amount of data is considered.
- A recursive version of the MLBA that does not involve any hard decision is presented. Then, the error propagation problem frequently encountered with the DFE-like algorithm (and, thus, with the algorithm proposed by Gesbert [19]) is solved. Moreover, we prove, in the noiseless case, that when the recursive algorithm converges, then it converges toward the global minimum. System adaptivity is then obtained by introducing an exponential weighting factor in the criterion. The connection between this algorithm and the structure of the DFE is emphasized. The resulting algorithm is very similar to a DFE where the hard decision is replaced by a soft estimate of all symbols involved in the computation of a given channel output. Update strategies of the filters can be either of a least-squares type (RLS like) or of a stochastic gradient type (LMS like). Both of them are derived in this paper.

This paper is organized as follows. The general setup and the DML criterion, which is a quadratic minimization problem in both the channels and the symbols, are presented in Section II. For noise-free data, we recall that the global minimum of the DML criterion is unique. Then, in Section III, we derive the two-steps block iterative algorithm and show that it can converge to local minima. A recursive version is derived in Section IV. We prove that the recursive algorithm converges only toward the global minimum. A simplified version of the recursive algorithm is presented in Section IV-D. In Section V-A, we introduce a weighting factor into the criterion, and we obtain the adaptive algorithm. A comparison between the structure of our adaptive algorithm and of a DFE is proposed in Section V-B. The performance of the algorithms and comparison with existing approaches are provided in Section VI.

II. Problem Formulation

Let $x(t)$ be the continuous-time baseband signal received at the output of a noisy channel

$$x(t) = \sum_{n=-\infty}^{+\infty} \tilde{h}(t - nT)\tilde{s}(n) + b(t) \tag{1}$$

where $\tilde{h}(t - nT)$ denotes the baseband equivalent channel, including the effects of the emission and reception filters, of the channel response, and of the modulation and demodulation. The symbol sequence $\tilde{s}(n)$ is emitted with rate $1/T$, and $b(t)$ stands for some additive independent white Gaussian noise. Consider a fractionally spaced equalizer, the received continuous time signal $x(t)$ being sampled at rate T/L. For $1 \leq i \leq L$, set $\tilde{x}_i(n) = \tilde{x}((i-1)T/L + nT), b_i(n) = b((i-1)T/L + nT)$ and $\tilde{h}_i(l) = \tilde{h}((i-1)T/L + nT)$. The discrete time version of the signal model in (1) may be expressed as

$$x_i(n) = \sum_{k=-\infty}^{+\infty} \tilde{h}_i(n-k)\tilde{s}(k) + b_i(n) \quad i = 1, \ldots, L$$

where $x_i(n), \tilde{h}_i(n), \tilde{s}(n)$ and $b_i(n)$ are complex variables. This single-input/multiple-output (SIMO) model can also be used for systems involving multiple receivers. For convenience, we adopt the following notations throughout the paper.

- h, s are variables denoting any channel and any symbol sequence, respectively.
- $\tilde{h}, \tilde{s}$ are the true channels and symbols, respectively, and X is the corresponding (noisy) observation.
- $\hat{h}, \hat{s}$ are the estimates of $\tilde{h}$ and $\tilde{s}$.
- $\mathbf{s}_N(n) = [s(n)s(n-1)\ldots s(n-N-M+1)]^T$ and n is the time index.
- $\hat{\mathbf{s}}_N^{(i)}(n+i) = [\hat{s}(n+i)\hat{s}(n+i-1)\ldots\hat{s}(n+i-N-M+1)]^T$ is a vector of length $M+N$ containing symbols estimated at iteration i.

The channel impulse responses $\tilde{\mathbf{h}}_i, 1 \leq i \leq L (L > 1)$ are assumed to have a finite length, and M stands for the maximum order of any channel. Let $\mathbf{X}_N(n) = [x_1(n)\ldots x_L(n)\ldots x_1(n-N+1)\ldots x_L(n-N+1)]^T$ denote the vector obtained by interlacing the outputs of the different channels, $\tilde{\mathbf{s}}_N(n) = [\tilde{s}(n)\ldots\tilde{s}(n-N-M+1)]^T$ the vector containing $N+M$ transmitted symbols, and $\tilde{\mathbf{h}}(k) = [\tilde{h}_1(k)\ldots\tilde{h}_L(k)]^T$. Then, the output $\mathbf{X}_N(n)$ reads:

$$\boldsymbol{X}_N(n) = \mathcal{T}_N(\tilde{\boldsymbol{h}})\tilde{\boldsymbol{s}}_N(n) + \boldsymbol{B}_N(n). \tag{2}$$

Here, $\mathbf{B}_N(n)$ stands for the noise vector. The noise sequences are assumed to be i.i.d., Gaussian, and mutually uncorrelated. In (2), operator $\mathcal{T}_N$ transforms a sequence of channel $\mathbf{h}(k) = [h_1(k)\dots h_L(k)]^T, k = 0,\dots,M$ into the following $LN \times (M+N)$ generalized Sylvester matrix [23]:

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

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

$$\mathcal{U}(\mathbf{s}_N(n))\mathbf{h} = \mathcal{T}_N(\mathbf{h})\mathbf{s}_N(n), \quad \forall \mathbf{s}_N, \quad \forall \mathbf{h}. \tag{3}$$

It can be shown that this matrix reads

$$\mathcal{U}(\mathbf{s}_N(n)) = \begin{pmatrix} I_L \bigotimes \mathbf{s}_1(n)^T \\ I_L \bigotimes \mathbf{s}_1(n-1)^T \\ \vdots \\ I_L \bigotimes \mathbf{s}_1(n-N+1)^T \end{pmatrix}$$

where $\bigotimes$ is the Kronecker product, I_L is the $L \times L$ identity matrix, and $\mathbf{s}_1(n) = [s(n)\ s(n-1)\ \cdots\ s(n-M)]^T$. The results displayed in the paper rely on the following assumptions.

- H1) $\mathcal{T}_N(\tilde{\mathbf{h}})$ has full column rank.
- H2) The symbol sequence $\tilde{s}_N(n)$ has linear complexity $2M+1$ or greater [24]. 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

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

  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.

Assumption **H1)** means that there is channel diversity that guarantees that (2) is an overdetermined system of equations for $\tilde{\mathbf{h}}$ fixed. Similarly, **H2')** ensures that (2) is an overdetermined system for $\tilde{\mathbf{s}}_N(n)$ fixed. Denote by $\nu_M(\tilde{\mathbf{s}}_N(n))$ the matrix

$$\begin{pmatrix} \tilde{s}(n) & \tilde{s}(n-1) & \dots & \tilde{s}(n-N+M+1) \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{s}(n-2M) & \tilde{s}(n-2M-1) & \dots & \tilde{s}(n-N-M+1) \end{pmatrix}$$

Then, **H2)** implies $\text{rank}(\nu_M(\tilde{\mathbf{s}}_N(n))) = 2M+1$. Hence, the sample covariance of the vector sequence $\tilde{\mathbf{s}}_{M+1}(n) = [\tilde{s}(n), \tilde{s}(n-1),\dots,\tilde{s}(n-2M)]^T$ is full rank, and it is seen that the linear complexity property is strongly connected with the notion of persistent excitation [25] of a sequence. We can remark that $\text{rank}(\nu_M(\tilde{\mathbf{s}}_N(n))) = 2M+1$ implies that, necessarily, $N \geq 3M+1$.

We consider the problem of identifying both $\tilde{\mathbf{h}}$ and $\tilde{\mathbf{s}}_N$ from $\mathbf{X}_N(n)$ without using any prior about the transmitted sequence or the channel. Following [11], [13], and [19], $(\tilde{\mathbf{h}}, \tilde{\mathbf{s}}_N)$ are estimated through the minimization of the DML criterion with respect to the joint variable $(\mathbf{h}, \mathbf{s}_N(n))$:

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

where the second formulation comes from (3). Hence, the estimated channels 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, $(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N(n))$ is a global minimum of $\mathcal{J}$ if and only if $\mathcal{J}(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N(n)) = 0$. The following theorem established in [19] and [22] gives a characterization of the global minimum.

Theorem 1: In the noiseless case and under **H1)** and **H2)**, $\mathcal{J}(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N(n)) = 0$ iff $\exists \alpha \in \mathbb{C}^*$ such as $\hat{\mathbf{h}} = \alpha\tilde{\mathbf{h}}$ and $\hat{\mathbf{s}}_N(n) = \tilde{\mathbf{s}}_N(n)/\alpha$.

Proof: The equality $\mathcal{T}_N(\tilde{\mathbf{h}})\tilde{\mathbf{s}}_N(n) = \mathcal{T}_N(\hat{\mathbf{h}})\hat{\mathbf{s}}_N(n)$ can be rearranged as

$$\mathcal{T}_{M+1}(\tilde{\mathbf{h}})\nu_M(\tilde{\mathbf{s}}_N(n)) = \mathcal{T}_{M+1}(\hat{\mathbf{h}})\nu_M(\hat{\mathbf{s}}_N(n)).$$

H2) ensures that $\text{rank}(\nu_M(\tilde{\mathbf{s}}_N(n))) = 2M+1$, which implies that $\text{range}(\mathcal{T}_{M+1}(\mathbf{h})\nu_M(\tilde{\mathbf{s}}_N(n))) = \text{range}(\mathcal{T}_{M+1}(\tilde{\mathbf{h}}))$, where $\text{range}(\mathbf{A})$ stands for the column space of $\mathbf{A}$. Then, $\text{range}(\mathcal{T}_{M+1}(\tilde{\mathbf{h}})) \subset \text{range}(\mathcal{T}_{M+1}(\hat{\mathbf{h}}))$. Since $\text{range}(\mathcal{T}_{M+1}(\tilde{\mathbf{h}}))$ and $\text{range}(\mathcal{T}_{M+1}(\hat{\mathbf{h}}))$ have the same dimensions, we get $\text{range}(\mathcal{T}_{M+1}(\tilde{\mathbf{h}})) = \text{range}(\mathcal{T}_{M+1}(\hat{\mathbf{h}}))$. Using [26, Th. 2], we conclude that there exists $\alpha \in \mathbb{C}$ such as $\hat{\mathbf{h}} = \alpha\tilde{\mathbf{h}}$. ■

It can be observed that Theorem 1 and the sufficient condition of identifiability presented in [27] are similar. This is not surprising since, when the noise is Gaussian, all information about the channel in the likelihood function is concentrated in the second-order moments of the observation. Theorem [27] proves that the global minimum is obtained only for the true values of the parameters up to a scalar factor. However, the implementation of DML based algorithms is most often complicated by the existence of local minima. Actually, the criterion $\mathcal{J}$ has a quadratic form in terms of $\mathbf{h}$ and $\mathbf{s}_N(n)$ separately, but $\mathcal{J}$ is nonconvex with respect to the joint variable $(\mathbf{h}, \mathbf{s}_N(n))$ and does not admit an explicit solution. Thus, even if we characterize the global minimum, we do not know whether the algorithms that will be used to minimize (4) will converge toward this global minimum.

In the next section, we recall the block algorithm proposed by Gesbert to minimize (4), and we present a characterization of the possible stationary points of this algorithm and a strategy that permits us to circumvent the local minima problem.

III. Maximum Likelihood Block Algorithm

In this section, we provide a block algorithm based on DML techniques. Usually, the local minima problem, which is frequently encountered with these methods, is solved by initializing the procedure using less efficient (in terms of

performance) techniques, which are not subject to these local minima problems. By doing so, it is hoped that even if local minima occur, they will not be close to the optimal solution so that the iterative algorithm will converge to the global minimum. This has two drawbacks: i) It is not clear that such local minima close to the global one does not exist, and ii) such a procedure is usually computationally demanding. As a substitute to this procedure, we propose a test allowing to check whether the obtained stationary point is the global minimum or a spurious local minimum. In the last case, the procedure must be reinitialized. Our test is also computationally demanding, but the objection pointed out in i) does not apply to our method. Moreover, the test emphasizes why it is pertinent to derive a recursive algorithm in this context (besides the arithmetic complexity problem). The proposed procedure is the first step toward a recursive algorithm.

A. Two-Step Iterative Algorithm

Classical ways of solving the minimization problem of (4) consist in expressing the minimizer with respect to $\mathbf{s}_N(n)$ as a function of $\mathbf{h}$ and inserting this expression in $\mathcal{J}$. Then, an iterative procedure is applied. Finally, the symbols are computed when the algorithm has converged. This formulation is not appropriate for building recursive algorithms, and we follow the approach proposed by Gesbert [19], who derives a simple iterative algorithm in two steps in which each step solves a least-squares problem alternately in $\mathbf{h}$ and in $\mathbf{s}$.

1) Algorithm: After some initialization, one iterates the following two steps until convergence:

$$\hat{\mathbf{h}}^{(k)} = \left[\mathcal{U}\left(\hat{\mathbf{s}}_N^{(k-1)}\right)^H \mathcal{U}\left(\hat{\mathbf{s}}_N^{(k-1)}\right)\right]^{-1} \times \mathcal{U}\left(\hat{\mathbf{s}}_N^{(k-1)}\right)^H \mathbf{X}_N(n) \tag{5}$$

$$\hat{\mathbf{s}}_N^{(k)}(n) = \left[\mathcal{T}_N\left(\hat{\mathbf{h}}^{(k)}\right)^H \mathcal{T}_N\left(\hat{\mathbf{h}}^{(k)}\right)\right]^{-1} \times \mathcal{T}_N\left(\hat{\mathbf{h}}^{(k)}\right)^H \mathbf{X}_N(n). \tag{6}$$

This is the MLBA. $\mathcal{T}_N(\hat{\mathbf{h}}^{(k)})$ and $\mathcal{U}(\hat{\mathbf{s}}_N^{(k)}(n))$ are assumed to be full rank for all k. The desired solution $(\tilde{\mathbf{h}}, \tilde{\mathbf{s}}_N(n))$ verifies **H1)** and **H2')**, which justifies this restriction. Each step decreases the value of $\mathcal{J}$, and the MLBA converges, possibly toward a local minimum. The corresponding stationary points are characterized below.

2) Characterization of the Stationary Points: Let $(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N(n))$ denote a stationary point of the MLBA; then

$$\hat{\mathbf{h}} = [\mathcal{U}(\hat{\mathbf{s}}_N(n))^H \mathcal{U}(\hat{\mathbf{s}}_N(n))]^{-1} \mathcal{U}(\hat{\mathbf{s}}_N)^H \mathbf{X}_N(n) \tag{7}$$

$$\hat{\mathbf{s}}_N(n) = [\mathcal{T}_N(\hat{\mathbf{h}})^H \mathcal{T}_N(\hat{\mathbf{h}})]^{-1} \mathcal{T}_N(\hat{\mathbf{h}})^H \mathbf{X}_N(n). \tag{8}$$

The two equations above are equivalent to

$$\mathbf{K} = \hat{\mathbf{X}}_N(n) - \mathbf{X}_N(n) \in \text{Null}(\mathcal{U}(\hat{\mathbf{s}}_N(n))^H) \cap \text{Null}(\mathcal{T}_N(\hat{\mathbf{h}})^H) \tag{9}$$

where $\hat{\mathbf{X}}_N(n) = \mathcal{T}_N(\hat{\mathbf{h}})\hat{\mathbf{s}}_N(n)$ stands for the estimation of $\mathbf{X}_N(n)$, and Null$(\mathbf{A})$ denotes the null space of $\mathbf{A}$. In the noiseless case, the global minimum is obtained for $\mathbf{K} = \mathbf{0}_{LN}$; otherwise, $(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N(n))$ is a local minimum. Therefore, local minima do exist if Null$(\mathcal{U}(\hat{\mathbf{s}}_N(n))^H) \cap Null(\mathcal{T}_N(\hat{\mathbf{h}})^H) \neq \emptyset$. We introduce the matrix $\mathbf{C} = [\mathcal{T}_N(\hat{\mathbf{h}})\mathcal{U}(\hat{\mathbf{s}}_N(n))]^H$. One can easily verify that Null$(\mathbf{C}) =$ Null$(\mathcal{U}(\hat{\mathbf{s}}_N(n))^H) \cap$ Null$(\mathcal{T}_N(\hat{\mathbf{h}})^H)$. Matrix $\mathbf{C}$ cannot be full column rank if it has more columns than rows, i.e., if $LN \geq L(M+1)+M+N$. In Section II, we underlined that **H2)** requires that $N \geq 3M+1$. In general, both relations cannot hold simultaneously; then Null$(\mathbf{C})$ is not an empty set, and the MLBA does present local minima. Experience shows that fortunately, local minima seldom happen, which is logical from their characterization. The size of their "subspace" is small. We will provide a procedure allowing us to check whether the algorithm has converged toward a local minimum, as well as offering insights into possible methods that would not be sensitive to this problem. The demonstration relies on the stability of the estimate of the channel in a recursive procedure.

B. Solving the Local Minima Problem

For a slowly varying channel (with respect to the block-size $N+M$), the true channel $\tilde{\mathbf{h}}$ may remain identical during several blocks. This observation suggests that we initialize block $i+1$ with the channel $\hat{\mathbf{h}}$ of block i. If $\hat{\mathbf{h}}$ and $\tilde{\mathbf{h}}$ are proportional, the vector $\hat{\mathbf{s}}_N$ in block $i+1$ is computed in one iteration. Otherwise, it is crucial to know whether a local minimum $\hat{\mathbf{h}}$ of block i may also be a local minimum of block $i+1$ since in the latter case, initializing each block with the previous $\hat{\mathbf{h}}$ might propagate an error from one block to another. We answer partially by providing a necessary and sufficient condition for $\hat{\mathbf{h}}$ to be a local minimum different from the global minimum. Based on it, we build a simple numerical test that combines the block algorithm with a growing window technique (the BGWT, cf. Fig. 1).

Block Growing Window Technique (BGWT)

- **Step 0**: Minimize $\mathcal{J}(\mathbf{h}, \mathbf{s}_N)$ using the block algorithm (5) and (6) to obtain $(\hat{\mathbf{h}}^{(0)}, \hat{\mathbf{s}}_N^{(0)}(n))$.
- **Step** $k = 1, \ldots, K-1$, **where** $K = 3M+1$: At time k, a new symbol $\tilde{s}(n+k)$ is transmitted. The vector of emitted symbols is $\tilde{\mathbf{s}}_{N+k}(n+k)$, and its length is $M+N+k$. The minimizer of $\mathcal{J}(\mathbf{h}, \mathbf{s}_{N+k})$ is $(\hat{\mathbf{h}}^{(k)}, \hat{\mathbf{s}}_{N+k}^{(k)}(n+k))$ (which is obtained via MLBA).

At the end of these iterations, either $\hat{\mathbf{h}}^{(k)} = \hat{\mathbf{h}}$ for $k = 0, \ldots, K-1$ or $\exists i, k \in [1, \ldots, K]$, such as $\hat{\mathbf{h}}^{(k)} \neq \hat{\mathbf{h}}^{(i)}$. The consequences of these issues are formalized below.

1) Noiseless Case: The following theorem gives a rule to distinguish the global minimum from the local ones.

Theorem 2: Assume that there is no noise and that the channel $\tilde{\mathbf{h}}$ is constant over the window $[n - N - M + 1, \ldots, n + 3M + 1]$. Assume also that $\mathcal{T}_{N+k}(\hat{\mathbf{h}}), \mathcal{T}_{N+k}(\tilde{\mathbf{h}}), \mathcal{U}(\hat{\mathbf{s}}_{N+k}(n+k)), \mathcal{U}(\tilde{\mathbf{s}}_{N+k}(n+k))$ are full column rank matrices for any $k = 0, \ldots, K-1$ and that rank$(\nu_M(\tilde{\mathbf{s}}_K(n+K)) = 2M+1$.

If, all along the BGW procedure, $\hat{\mathbf{h}}^{(k)} = \hat{\mathbf{h}}$ and $\hat{\mathbf{s}}_{N+k}^{(k)}(n+k) = \hat{\mathbf{s}}_{N+k}(n+k) \neq \mathbf{0}_{M+N+k}$ for any $k = 0, \ldots, K-1$, then $\hat{\mathbf{h}}$ and $\hat{\mathbf{s}}_{N+K}(n+K)$ are the global minimizers of $\mathcal{J}(\mathbf{h}, \mathbf{s}_{N+K})$.

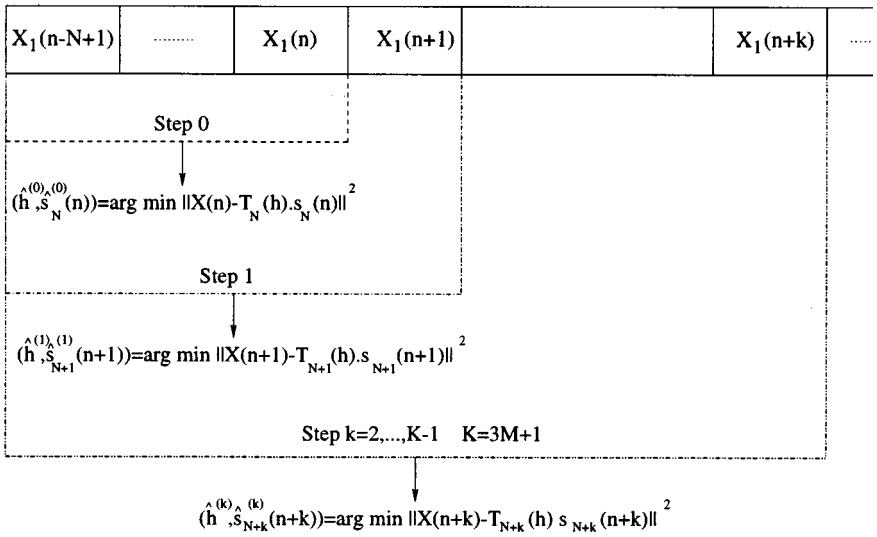

Fig. 1. Procedure BGWT: Block algorithm combined with a growing window technique.

Proof: Assume that $\forall k \in [0; K-1]$, $(\hat{\mathbf{h}}, \hat{\mathbf{s}}_{N+k}(n+k))$ is a stationary point of $\mathcal{J}(\mathbf{h}, \mathbf{s}_{N+k})$. Then, for any $k \in [0; K-1]$, $((\partial \mathcal{J}(\mathbf{h}, \mathbf{s}_{N+k}))/\partial \mathbf{h})$ computed in $(\hat{\mathbf{h}}, \hat{\mathbf{s}}_{N+k}(n+k))$ is zero:

$$\mathcal{U}(\hat{\mathbf{s}}_{N+k}(n+k))^H[\mathcal{U}(\hat{\mathbf{s}}_{N+k}(n+k))\hat{\mathbf{h}} - \mathcal{U}(\tilde{\mathbf{s}}_{N+k}(n+k))\tilde{\mathbf{h}}] = 0. \quad (10)$$

By comparing the expression in (10) corresponding to $k-1$ with the expression relevant to k, we see that the latter is satisfied whenever

$$\mathcal{U}(\hat{\mathbf{s}}_1(n+k))^H[\mathcal{U}(\hat{\mathbf{s}}_1(n+k))\hat{\mathbf{h}} - \mathcal{U}(\tilde{\mathbf{s}}_1(n+k))\tilde{\mathbf{h}}] = 0. \quad (11)$$

$\mathcal{U}(\hat{\mathbf{s}}_1(n+k))^H$ is a $L(M+1) \times L$ matrix that has full column rank, provided that $\hat{\mathbf{s}}_1(n+k) \neq 0_{M+1}$. Then, (11) is equivalent to

$$\mathcal{U}(\hat{\mathbf{s}}_1(n+k))\hat{\mathbf{h}} - \mathcal{U}(\tilde{\mathbf{s}}_1(n+k))\tilde{\mathbf{h}} = 0. \quad (12)$$

Equation (12) holds for all $k \in [0; K-1]$, and stacking the K obtained equations, we get

$$\mathcal{U}(\hat{\mathbf{s}}_K(n+K))\hat{\mathbf{h}} - \mathcal{U}(\tilde{\mathbf{s}}_K(n+K))\tilde{\mathbf{h}} = 0.$$

Thus, the noiseless DML criterion is zero in $(\hat{\mathbf{h}}, \hat{\mathbf{s}}_K(n+K))$:

$$\mathcal{J}(\hat{\mathbf{h}}, \hat{\mathbf{s}}_K(n+K)) = ||\mathbf{X}_K(n+K) - \mathcal{U}(\hat{\mathbf{s}}_K(n+K))\hat{\mathbf{h}}||^2 = 0.$$

Under **H1)** and **H2)** (which requires $K \geq 3M+1$), the conditions of theorem [27] are satisfied, and hence, there exists α such as $\hat{\mathbf{h}} = \alpha\tilde{\mathbf{h}}$ and $\hat{\mathbf{s}}_K(n+K) = \tilde{\mathbf{s}}_K(n+K)/\alpha$. Then, at step zero, $\hat{\mathbf{h}}^{(0)} = \alpha\tilde{\mathbf{h}}$; therefore, $\hat{\mathbf{s}}_{N+K}(n+K) = \tilde{\mathbf{s}}_{N+K}(n+K)/\alpha$, which completes the proof. ■

As a consequence, the only stationary point such that the filter and the symbol estimate remain unchanged during $K = 3M+1$ consecutive steps is the global minimum. The proof above is based on the unrealistic assumption of noiseless data. In the next section, an extension to the case of noisy data is provided with the restriction that a large amount of symbols is considered.

2) Case of Noisy Data: In the noisy case and if a large number of data is considered, $\hat{\mathbf{h}}$ and $\hat{\mathbf{s}}_N$ can be read as

$$(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N) = \arg\min_{\mathbf{h}, \mathbf{s}_N} \mathcal{L}(\mathbf{h}, \mathbf{s}_N) = \arg\min_{\mathbf{h}, \mathbf{s}_N} ||\mathcal{T}_N(\tilde{\mathbf{h}})\tilde{\mathbf{s}}_N - \mathcal{T}_N(\mathbf{h})\mathbf{s}_N||^2$$

where $\mathcal{L}(\mathbf{h}, \mathbf{s}_N)$ is the noiseless DML criterion (the proof is outlined in Appendix I). Thus, for a number of symbols large enough, the proof for the noisy case is similar to the proof presented for noise-free data, and Theorem 2 holds.

The BGW procedure is not suited for working with large data sequences. On the other hand, the test provides a solution to the local minima problem that is one of the main difficulties with these methods. These remarks justify our choice to develop a low-cost recursive version of the BGWT, paying attention that the property established in this section is maintained.

IV. Maximum Likelihood Recursive Algorithm (MLRA)

The above result strongly suggests that the local minima problem could be easily solved in a recursive growing window procedure. Such a recursive algorithm can be derived from the BGWT of Section III by applying some approximations to the BGWT, which leads to a lower complexity algorithm. Moreover, we show that in the digital communication context, the update of the filters in the proposed algorithm is equivalent to a stochastic gradient-based method. The computational complexity of the resulting algorithm is reasonably small for being used in practical implementation. At the same time, we prove, under some classical assumptions, that when the proposed algorithm converges, then it converges toward the global minimum.

A. Derivation of the MLRA

Here, we modify the BGWT for obtaining a recursive algorithm such that, given the least-squares estimates of the symbols and of the filters at iteration $i-1$, we may update the estimates of these vectors at iteration i on the arrival of a new symbol.

The filters and the symbols are still computed alternately. Let $\hat{\mathbf{h}}^{(i)}$ and $\hat{\mathbf{s}}_{N+i}^{(i)}(n+i)$ denote, respectively, the channel and the $N+M+i \times 1$ symbol vector estimated at iteration i. The first simplification consists of replacing the minimization in step $i(\forall i)$ of the BGWT by the two relations

$$\hat{\mathbf{s}}_{N+i}^{(i)}(n+i) = \arg \min_{\mathbf{s}_{N+i}(n+i)} \|\mathbf{X}_{N+i}(n+i) - \mathcal{T}_{N+i}\left(\hat{\mathbf{h}}^{(i-1)}\right)\mathbf{s}_{N+i}(n+i)\|^2 \quad (13)$$

$$\hat{\mathbf{h}}^{(i)} = \arg \min_{\mathbf{h}} \|\mathbf{X}_{N+i}(n+i) - \mathcal{U}\left(\hat{\mathbf{s}}_{N+i}^{(i)}(n+i)\right)\mathbf{h}\|^2. \quad (14)$$

The equations above are quite similar to the equations in the first iteration of step i of the BGWT. Equation (13) solves a least-squares problem in $\hat{\mathbf{s}}_{N+i}^{(i)}(n+i)$, whose length is $N+M+i$. Hence, the computational complexity of (13) increases quickly with i. Considering that it is unlikely that the most recent received samples have a strong impact on the estimate of symbols that have been emitted long ago, we update only the last P symbols. Hence, P is a fundamental parameter to be determined, which will drive a complexity/efficiency tradeoff. Implicitly, the other symbols are thus supposed to be correctly estimated. At iteration i, $\hat{\mathbf{s}}_{N+i}^{(i)}(n+i)$ is split into two parts:

$$\left[\underbrace{\hat{\mathbf{s}}_{P+1-M}^{(i)}(n+i)}_{\text{Updated at iteration } i} \;\vdots\; \underbrace{\hat{\mathbf{s}}_{N+i-P-1}^{(i)}(n+i-P-1)}_{\text{Not updated at and after iteration } i}\right].$$

Now, we consider separately the minimization w.r.t. the symbols and the filter.

1) Minimization With Respect to the Symbols: Since only P symbols are updated at iteration i, the minimization with respect to the symbols in (13) reduces to

$$\hat{\mathbf{s}}_{P+1-M}^{(i)}(n+i) = \underset{\mathbf{Z}\in\mathbb{C}^{P+1}}{\operatorname{argmin}} \left\| \mathbf{X}_{P+1}(n+i) - \mathcal{T}_{P+1}\left(\hat{\mathbf{h}}^{(i-1)}\right)\begin{bmatrix}\mathbf{Z}\\ \hat{\mathbf{s}}_0^{(i-1)}(n+i-P-1)\end{bmatrix}\right\|^2. \quad (15)$$

Matrix $\mathcal{T}_{P+1}(\hat{\mathbf{h}}^{(i-1)})$ can be split into two submatrices:

$$\mathcal{T}_{P+1}\left(\hat{\mathbf{h}}^{(i-1)}\right) = \left[\underbrace{\mathcal{T}_{P+1}^{\text{Left}}\left(\hat{\mathbf{h}}^{(i-1)}\right)}_{P+1} \;\vdots\; \underbrace{\mathcal{T}_{P+1}^{\text{Right}}\left(\hat{\mathbf{h}}^{(i-1)}\right)}_{M}\right]. \quad (16)$$

By combining (15) and (16), the estimated symbol vector at iteration i can be calculated as

$$\hat{\mathbf{s}}_{P+1-M}^{(i)}(n+i) = \left[\left[\mathcal{T}_{P+1}^{Left}\left(\hat{\mathbf{h}}^{(i-1)}\right)\right]^H \mathcal{T}_{P+1}^{\text{Left}}\left(\hat{\mathbf{h}}^{(i-1)}\right)\right]^{-1} \times \left[\mathcal{T}_{P+1}^{\text{Left}}\left(\hat{\mathbf{h}}^{(i-1)}\right)\right]^H \left[\mathbf{X}_{P+1}(n+i). - \mathcal{T}_{P+1}^{\text{Right}}\left(\hat{\mathbf{h}}^{(i-1)}\right) \times \hat{\mathbf{s}}_0^{(i-1)}(n+i-P-1)\right] \quad (17)$$

where $\times$ stands for the multiplication operator.

2) Minimization With Respect to the Filter: The filter $\hat{\mathbf{h}}^{(i)}$ is obtained using (14). It can be computed recursively from $\hat{\mathbf{h}}^{(i-1)}$. Let $\mathbf{R}^{(i)} = \mathcal{U}(\hat{\mathbf{s}}_{N+i}^{(i)}(n+i))^H \mathcal{U}(\hat{\mathbf{s}}_{N+i}^{(i)}(n+i))$ denote the block diagonal covariance matrix of the estimated symbols. Then, the update of $\hat{\mathbf{h}}^{(i)}$ is performed thanks to

$$\hat{\mathbf{h}}^{(i)} = \hat{\mathbf{h}}^{(i-1)} + \left[\mathbf{R}^{(i)}\right]^{-1} \times \left\{\mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)^H \times \left[\mathbf{X}_{P+1}(n+i) - \mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)\hat{\mathbf{h}}^{(i-1)}\right] - \mathcal{U}\left(\hat{\mathbf{s}}_{P}^{(i-1)}(n+i-1)\right)^H \left[\mathbf{X}_P(n+i-1) - \mathcal{U}\left(\hat{\mathbf{s}}_{P}^{(i-1)}(n+i-1)\right)\hat{\mathbf{h}}^{(i-1)}\right]\right\}. \quad (18)$$

More details on the derivation of this equation is given in Appendix B. The MLRA consists of (17) and (18). Matrix $\mathbf{R}^{(i)}$ turns out to be a block diagonal matrix with diagonal $\underbrace{[(N+i)\mathbf{C}_{M+1} \cdots (N+i)\mathbf{C}_{M+1}]}_{L}$ and $\mathbf{C}_{M+1}$ the $M+1 \times M+1$ empirical covariance matrix of the estimated symbols. Hence, for $P = 0$, (18) reduces to the classical equations of L recursive least-squares (RLS) algorithms (each subchannel being updated separately).

3) Comparison Between the MLRA and the BGWT: The computational complexity of the MLRA is largely reduced compared to the BGWT for the following reasons.

1) The minimization with respect to the joint variable in step $i(\forall i)$ of the BGWT (iterative procedure) is replaced by the minimization of a criterion with respect to each variable separately, which coincides with the first iteration of step i in the BGWT.
2) At iteration i, the BGWT computes $N+i$ symbols whose length increases with i, whereas only a fixed number P (independent of i) symbols are computed in the MLRA.
3) In the MLRA, $\hat{\mathbf{h}}^{(i)}$ is updated recursively from $\hat{\mathbf{h}}^{(i-1)}$, which is done without any approximation, whereas, in the BGWT, the whole channel computation is performed in a single step without having any benefit from the channel computations done in the previous steps.

Unfortunately, possible divergence problems may occur. These diverging situations are essentially due to the choice of P, which should be greater than M (order of the channel). This point is illustrated in Figs. 7 and 8.

B. Initialization

In recursive implementations, the computation usually starts with known initial conditions and makes use of the information contained in the new data samples to update the previous estimates. Here, neither the symbols nor the channel are known. Hence, we need to find a reliable initial estimate. A similar problem is encountered in TSML [11] or IQML [12]. Generally, the problem is solved by making use of an initialization procedure such as the subspace algorithm, for example. Here, we propose to initialize the MLRA with $(\hat{\mathbf{h}}^{(0)}, \hat{\mathbf{s}}_N^{(0)}(n))$ defined as the stationary point of the MLBA (5) and (6) over a block of size $N+M$. The MLBA starts with a randomly chosen initialization point. The choice of N reflects a tradeoff between the accuracy of the estimates and the involved computational cost. Experience shows that choosing N about $10M$ leads to a reasonable tradeoff. In any case, this always corresponds to

$N > 3M+1$, which ensures that $(\tilde{\mathbf{h}}, \tilde{\mathbf{s}}_N)$ is the unique global minimum of $\mathcal{J}(\mathbf{h}, \mathbf{s}_N)$ (cf. Theorem 1).

C. Convergence of the Recursive Algorithm

Our main result concerning the convergence of the MLRA is as follows.

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

Proof: If the MLRA converges, then $\exists n_0$ such as $\forall k \geq n_0, \hat{\mathbf{h}}^{(k)} = \hat{\mathbf{h}}^{(n_0)}$ and $\hat{\mathbf{s}}_{N+k}^{(k)}(n+k) = \hat{\mathbf{s}}_{N+k}$. At iteration k, the estimated filter satisfies the following relation:

$$\hat{\mathbf{h}}^{(k)} = \arg\min_{\mathbf{h}} \mathcal{J}\left(\mathbf{h}, \hat{\mathbf{s}}_{N+k}^{(k)}(n+k)\right). \tag{19}$$

Equation (19) is equivalent to

$$\mathcal{U}\left(\hat{\mathbf{s}}_{N+k}^{(k)}(n+k)\right)^H \mathcal{U}\left(\hat{\mathbf{s}}_{N+k}^{(k)}(n+k)\right)\hat{\mathbf{h}}^{(k)} = \mathcal{U}\left(\hat{\mathbf{s}}_{N+k}^{(k)}(n+k)\right)^H \mathbf{X}_{N+k}(n+k). \tag{20}$$

We split the previous expression into two terms, and we obtain

$$\begin{aligned}&\mathcal{U}\left(\hat{\mathbf{s}}_1^{(k)}(n+k)\right)^H \mathcal{U}\left(\hat{\mathbf{s}}_1^{(k)}(n+k)\right)\hat{\mathbf{h}}^{(k)} \\ &\quad -\mathcal{U}\left(\hat{\mathbf{s}}_1^{(k)}(n+k)\right)^H \mathbf{X}_1(n+k) \\ &= \mathcal{U}\left(\hat{\mathbf{s}}_{N+k-1}^{(k)}(n+k-1)\right)^H \mathbf{X}_{N+k}(n+k-1) \\ &\quad - \mathcal{U}\left(\hat{\mathbf{s}}_{N+k-1}^{(k)}(n+k-1)\right)^H \mathcal{U}\left(\hat{\mathbf{s}}_{N+k-1}^{(k)}(n+k-1)\right)\hat{\mathbf{h}}^{(k)}.\end{aligned} \tag{21}$$

Since $\forall k \geq n_0, \hat{\mathbf{h}}^{(k)} = \hat{\mathbf{h}}^{(n_0)}$ and $\hat{\mathbf{s}}_{N+k}^{(k)}(n+k) = \hat{\mathbf{s}}_{N+k}$ and using (19) taken at time $k-1$, we conclude that the right member of (21) is zero. Then, (21) reduces to

$$\mathcal{U}(\hat{\mathbf{s}}_1(n+k))^H \left[\mathcal{U}(\hat{\mathbf{s}}_1(n+k))\hat{\mathbf{h}}^{(n_0)} - \mathbf{X}_1(n+k)\right] = 0. \tag{22}$$

Matrix $\mathcal{U}(\hat{\mathbf{s}}_1^{(k)}(n+k))^H$ is full column rank as long as $\mathbf{s}_1^{(k)}(n+k) \neq \mathbf{0}_{M+1}$. Then, we get

$$\mathcal{U}(\hat{\mathbf{s}}_1(n+k))\hat{\mathbf{h}}^{(n_0)} - \mathbf{X}_1(n+k) = 0. \tag{23}$$

Equation (23) holds $\forall k \geq n_0$. We stack the equations obtained for $k, k+1, \ldots k+K$ with $K \geq 3M+1$, and we obtain

$$\mathcal{U}(\hat{\mathbf{s}}_K(n+K))\hat{\mathbf{h}}^{(n_0)} - \mathbf{X}_K(n+K) = 0.$$

Under **H1)** and **H4)**, the conditions of Theorem 1 are satisfied; hence, $\hat{\mathbf{h}}^{(n_0)}$ and $\hat{\mathbf{s}}_K(n+K)$ are the true values of the parameters up to a scalar factor. ∎

Once again, the theorem above, which has been established for noiseless data, can be extended to noisy data with the restriction that a large amount of data are considered (see Appendix A).

D. Simplified Recursive Algorithm

In the recursive algorithm, the channels are updated thanks to (18), which is equivalent to

$$\begin{aligned}\hat{\mathbf{h}}^{(i)} = \hat{\mathbf{h}}^{(i-1)} + \mu_i \left[\mathbf{S}^{(i)}\right]^{-1} \Big\{ &\mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)^H \\ &\times \left[\mathbf{X}_{P+1}(n+i) - \mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)\hat{\mathbf{h}}^{(i-1)}\right] \\ &- \mathcal{U}\left(\hat{\mathbf{s}}_P^{(i-1)}(n+i-1)\right)^H \Big[\mathbf{X}_P(n+i-1) \\ &- \mathcal{U}\left(\hat{\mathbf{s}}_P^{(i-1)}(n+i-1)\right)\hat{\mathbf{h}}^{(i-1)}\Big]\Big\}\end{aligned} \tag{24}$$

where $\mu_i = 1/(N+i)$, and where

$$\mathbf{S}^{(i)} = \operatorname{diag}\left[\underbrace{\frac{1}{N+i}\sum_{l=1-N}^{i} \mathcal{S}_l \ldots \frac{1}{N+i}\sum_{l=1-N}^{i} \mathcal{S}_l}_{L \text{ blocks}}\right].$$

Each block $\mathcal{S}_l$ is a $M+1 \times M+1$ matrix. Let $\mathcal{S}_l^{a,b}$ denote the element of $\mathcal{S}_l$ at line a and column b; then, $\mathcal{S}_l^{a,b} = \hat{s}^{(i)*}(n+l-a+1)\hat{s}^{(i)}(n+l-b+1)$. We assume that the sequence $\{\hat{s}(k)\}$ is ergodic; then, $\lim_{i\to\infty}(1/N+i)\sum_{l=1-N}^{i}\mathcal{S}_l = \mathbf{Cov}_{M+1}(\hat{\mathbf{s}}^{(i)})$, where $\mathbf{Cov}_{M+1}(\hat{\mathbf{s}}^{(i)})$ is the $M+1 \times M+1$ covariance matrix of the estimated symbols. We proved (in Section IV-C) that, in the noiseless case, if the recursive algorithm converges, then it converges toward the global minimum. Therefore, for i after the convergence, $\hat{\mathbf{s}}_{N+i}^{(i)}(n+i) = \tilde{\mathbf{s}}_{N+i}(n+i)/\alpha$ and $\alpha \in \mathbb{C}^*$, which leads to

$$\mathbf{Cov}_{M+1}\left(\hat{\mathbf{s}}^{(i)}\right) = \frac{1}{||\alpha||^2}\mathbf{Cov}_{M+1}(\tilde{\mathbf{s}}).$$

It is generally assumed that $\{\tilde{s}(k)\}$ is a sequence of i.i.d., complex, circular, random variables with zero mean [16]

$$E[\tilde{s}(k)] = 0 \quad E[\tilde{s}(k)^*\tilde{s}(j)] = \sigma_s^2\delta(k-j) \quad E[\tilde{s}(k)\tilde{s}(j)] = 0.$$

Note that in the digital communication context, these hypotheses are most often met [26], [28]. Consequently, $\mathbf{Cov}_{M+1}(\hat{\mathbf{s}}^{(i)})$ is a diagonal matrix, namely

$$\mathbf{Cov}_{M+1}(\hat{\mathbf{s}}) = \frac{\sigma_s^2}{||\alpha||^2}\mathbf{I}_{M+1}.$$

Therefore, for a time index i large enough, $\mathbf{Cov}_{M+1}(\hat{\mathbf{s}}^{(i)}) \approx (\sigma_s^2/||\alpha||^2)\mathbf{I}_{M+1}$, and $\mathbf{S}^{(i)} \approx (\sigma_s^2/||\alpha||^2)\mathbf{I}_{L(M+1)}$. Replacing the new expression for $\mathbf{S}^{(i)}$ in (24), we obtain

$$\begin{aligned}\hat{\mathbf{h}}^{(i)} = \hat{\mathbf{h}}^{(i-1)} + \rho_i \Big\{ &\mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)^H \Big[\mathbf{X}_{P+1}(n+i) \\ &- \mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)\hat{\mathbf{h}}^{(i-1)}\Big] - \mathcal{U}\left(\hat{\mathbf{s}}_P^{(i-1)}(n+i-1)\right)^H \\ &\times \left[\mathbf{X}_P(n+i-1) - \mathcal{U}\left(\hat{\mathbf{s}}_P^{(i-1)}(n+i-1)\right)\hat{\mathbf{h}}^{(i-1)}\right]\Big\}\end{aligned} \tag{25}$$

where $\rho_i = ((||\alpha||^2)/((N+i)\sigma_s^2))$. Equation (25) is a member of the stochastic gradient-based algorithms with decreasing step-size parameter (inversely proportional to i) and does not require any longer the inversion of the covariance matrix. In practical situations, this decreasing step strategy should not be used, but the connection with LMS-like algorithms was worth pointing out. It is clear, after these considerations, that LMS-based algorithms will perform very much like RLS-like algorithms.

V. Maximum Likelihood Adaptive Algorithm (MLAA)

Adaptivity is an important feature for tracking the variations of time-varying channels, as well as for providing current optimal solutions, without introducing a large delay due to block processing. The MLAA is obtained by introducing an exponential weighting factor into the DML criterion and by writing the corresponding recursions. Under the same assumptions as for the MLRA, it converges toward the global minimum. The MLAA is closely connected with a soft decision feedback equalizer (SDFE). This link is emphasized at the end of the section.

A. Derivation of the MLAA

The adaptivity feature can be obtained by introducing an exponential weighting factor into the definition of the criterion $\mathcal{J}(\mathbf{h}, \mathbf{s}_{N+i})$. Thus, the new criterion can be written

$$\mathcal{J}_\lambda(\mathbf{h}, \mathbf{s}_{N+i}) = \sum_{t=n-N+1}^{n+i} \lambda^{n+i-t} ||\mathbf{X}_1(t) - \mathcal{T}_1(\mathbf{h})\mathbf{s}_1||^2 \quad (26)$$

where $\lambda \in [0;1]$. The use of the weighting factor is intended to ensure that data in the distant past is forgotten. Such an algorithm is able to track the variations of the channel in a nonstationary environment. Using a matrix formulation, $\mathcal{J}_\lambda(\mathbf{h}, \mathbf{s}_{N+i})$ can be read as

$$\mathcal{J}_\lambda(\mathbf{h}, \mathbf{s}_{N+i}) = \left\| \Lambda_{N+i}^{1/2} [\mathbf{X}_{N+i}(n+i) - \mathcal{T}_{N+i}(\mathbf{h})\mathbf{s}_{N+i}] \right\|^2 \quad (27)$$

where

$$\Lambda_{N+i} = \text{diag}([\underbrace{1 \cdots 1}_{L} \vdots \underbrace{\lambda \cdots \lambda}_{L} \vdots \cdots \vdots \underbrace{\lambda^{N+i-1} \cdots \lambda^{N+i-1}}_{L}]).$$

We replace $\mathcal{J}(\mathbf{h}, \mathbf{s}_{N+i})$ with $\mathcal{J}_\lambda(\mathbf{h}, \mathbf{s}_{N+i})$ in Section IV-A, and the MLAA is obtained in the same way as the MLRA. The update of the symbols for the MLAA is given by the expression

$$\begin{aligned} &\hat{\mathbf{s}}_{P+1-M}^{(i)}(n+i) \\ &\quad = \left[\left[\mathcal{T}_{P+1}^{\text{Left}} \left(\hat{\mathbf{h}}^{(i-1)} \right) \right]^H \Lambda_{P+1} \mathcal{T}_{P+1}^{\text{Left}} \left(\hat{\mathbf{h}}^{(i-1)} \right) \right]^{-1} \\ &\qquad \times \left[\mathcal{T}_{P+1}^{\text{Left}} \left(\hat{\mathbf{h}}^{(i-1)} \right) \right]^H \Lambda_{P+1} \left[\mathbf{X}_{P+1}(n+i) \right. \\ &\qquad \left. - \mathcal{T}_{P+1}^{\text{Right}} \left(\hat{\mathbf{h}}^{(i-1)} \right) \times \hat{\mathbf{s}}_0^{(i-1)}(n+i-P-1) \right]. \end{aligned} \quad (28)$$

As to the update of the filter, we obtain a set of equations that are similar to (18):

$$\begin{aligned} \hat{\mathbf{h}}^{(i)} = \hat{\mathbf{h}}^{(i-1)} &+ \left[\mathbf{R}_\lambda^{(i)} \right]^{-1} \left\{ \mathcal{U} \left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i) \right)^H \Lambda_{P+1} \right. \\ &\times \left[\mathbf{X}_{P+1}(n+i) - \mathcal{U} \left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i) \right) \hat{\mathbf{h}}^{(i-1)} \right] \\ &- \lambda \mathcal{U} \left(\hat{\mathbf{s}}_P^{(i-1)}(n+i-1) \right)^H \Lambda_P \left[\mathbf{X}_P(n+i-1) \right. \\ &\left. \left. - \mathcal{U} \left(\hat{\mathbf{s}}_P^{(i-1)}(n+i-1) \right) \hat{\mathbf{h}}^{(i-1)} \right] \right\} \end{aligned} \quad (29)$$

where $\mathbf{R}_\lambda^{(i)} = \mathcal{U}(\hat{\mathbf{s}}_{N+i}^{(i)}(n+i))^H \Lambda_{N+i} \mathcal{U}(\hat{\mathbf{s}}_{N+i}^{(i)}(n+i))$ will be obtained recursively from $\mathbf{R}_\lambda^{(i-1)}$. This algorithm will be referred to as the maximum likelihood adaptive algorithm (MLAA). The MLAA also needs a good initialization that is easily obtained by running a block algorithm on a very short window in a manner very similar to the block algorithm of Section III. Arguments very similar to those of Section IV-C can be applied to the MLAA, which prove the following.

i) In the noiseless case and asymptotically in the noisy case, if **H1)** and **H4)** are met, if the channel is slowly varying, and if the MLAA converges, then it converges toward the global minimum of the weighted criterion $\mathcal{J}_\lambda^\infty = \lim_{i \to \infty} \arg\min_{\mathbf{h}, \mathbf{s}_N} ||\Lambda_{N+i}^{1/2} [\mathbf{X}_{N+i}(n+i) - \mathcal{T}_{N+i}(\mathbf{h})\mathbf{s}]||^2$. To the best of our knowledge, this is the first time that we can distinguish the local minima from the global one in that kind of algorithm.
ii) Moreover, Λ_{N+i} is a diagonal matrix, and its diagonal elements are all nonzero. Then, $\mathcal{J}_\lambda^\infty = 0$ is equivalent to $\mathcal{J}^\infty = 0$. Therefore, if the previous assumptions are met, the MLAA converges toward the global minimum of $\mathcal{J}^\infty$.

In the case of the MLAA, we do not know $\lim_{i \to \infty} \mathbf{R}^{(i)}$. The consequence is that we cannot prove that the MLAA is equivalent to a stochastic gradient method. This question will be addressed through simulation in Section VI.

B. Link With an SDFE

Here, we emphasize that the structure of the MLAA is very close to the structure of an SDFE. In [19], Gesbert has proposed an adaptive algorithm [the channel symbol algorithm (CSA)], based on least-squares techniques that aims at minimizing the criterion $\mathcal{J}_\lambda(\mathbf{h}, \mathbf{s})$ (27). We first recall some properties of the CSA.

1) Link Between the CSA and a DFE: For each iteration of the CSA, we have

$$\begin{aligned} \bar{s}(n+i) = \arg\min_{z \in \mathbb{C}} &||\mathbf{X}_1(n+i) \\ &- \mathcal{T}_1 \left(\hat{\mathbf{h}}^{(i-1)} \right) \begin{bmatrix} z \\ \hat{\mathbf{s}}_0(n+i-1) \end{bmatrix} \Big\|^2 \end{aligned} \quad (30)$$

$$\hat{s}(n+i) = g(\bar{s}(n+i)) \quad (31)$$

$$\hat{\mathbf{h}}^{(i)} \text{ updated via RLS.} \quad (32)$$

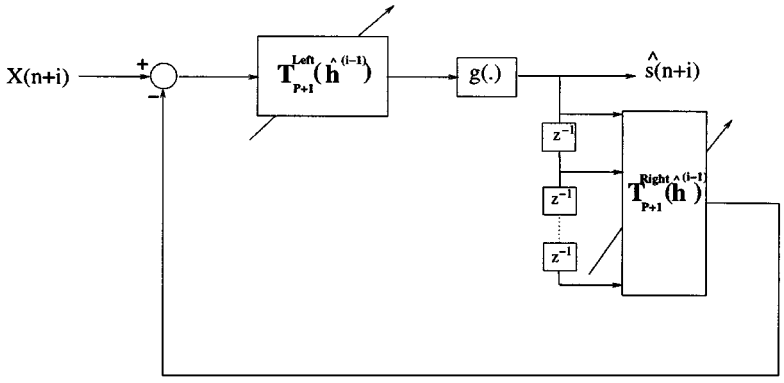

Fig. 2. Decision feedback update of the symbols in the CSA.

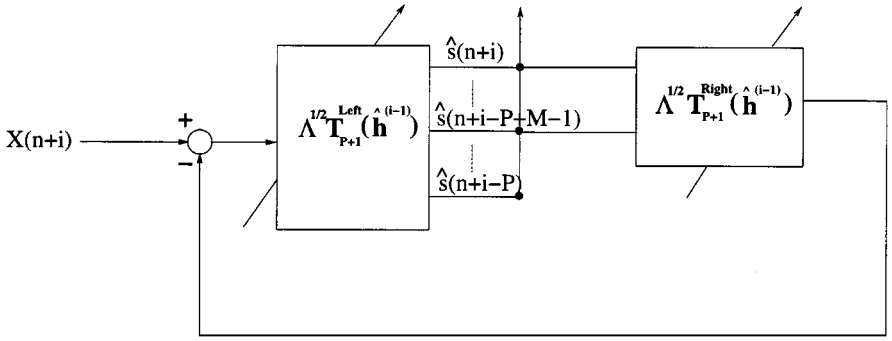

Fig. 3. Decision feedback update of the symbols in the MLAA.

The operator $g(\cdot)$ in (31) is a decision device. The link between the CSA and the DFE structure is shown in [19]. It is summarized below. An explicit expression for $\overline{s}(n+i)$ is obtained as

$$\overline{s}(n+i) = \frac{\left[\mathcal{T}_1^{\text{Left}}\left(\hat{\mathbf{h}}^{(i-1)}\right)\right]^H}{\left(\mathcal{T}_1^{\text{Left}}\left(\hat{\mathbf{h}}^{(i-1)}\right)\right)^H \mathcal{T}_1^{\text{Left}}\left(\hat{\mathbf{h}}^{(i-1)}\right)} \times \left(\mathbf{X}_1(n+i) - \mathcal{T}_1^{\text{Right}}\left(\hat{\mathbf{h}}^{(i-1)}\right)\hat{\mathbf{s}}_0(n+i-1)\right). \quad (33)$$

The decision feedback structure of (33) and (31) is shown in Fig. 2. The main difference between the structure of the CSA and of a DFE is the following: the presence of a feedforward filter in the DFE and the presence, in the CSA, of a "spatial" filter $(\mathcal{T}_1^{\text{Left}}(\hat{\mathbf{h}}^{(i-1)}))^H$, which combines the signals before the decision. Gesbert has also underlined the similarities between our criterion $\mathcal{J}_\lambda(\mathbf{h}, \mathbf{s}_{N+i})$ and a decision-directed criterion $\mathcal{J}_{\mathcal{DD}}$ [29] defined as

$$\mathcal{J}_{\mathcal{DD}} = \sum_{t=n-N+1}^{n+i} \lambda^{n+i-t} \|\hat{s}(t) - \overline{s}(t)\|^2.$$

Replacing (30) in the previous relation, we obtain

$$\mathcal{J}_{\mathcal{DD}} = \sum_{t=n-N+1}^{n+i} \lambda^{n+i-t} \times \left\| \frac{\left(\mathcal{T}_1^{\text{Left}}\left(\hat{\mathbf{h}}^{(i-1)}\right)\right)^H}{\left\|\mathcal{T}_1^{\text{Left}}\left(\hat{\mathbf{h}}^{(i-1)}\right)\right\|^2} (\mathbf{X}_1(t) - \mathcal{T}_1(\hat{\mathbf{h}})\hat{\mathbf{s}}_1(t)) \right\|^2.$$

The difference between the criterion $\mathcal{J}_{\mathcal{DD}}$ and $\mathcal{J}_\lambda$ lies only in the presence of the term $(((\mathcal{T}_1^{\text{Left}}(\hat{\mathbf{h}}^{(i-1)}))^H)/(\|\mathcal{T}_1^{\text{Left}}(\hat{\mathbf{h}}^{(i-1)})\|^2))$. This so-called CSA is thus very similar to the kind of algorithm one would obtain by implementing the equivalent of a DFE in a SIMO context: The adaptive algorithm is driven by the decision device. The additional properties of the SIMO system make such an algorithm more useful than in the SISO situation, as explained in [19].

2) Link Between the MLAA and a Soft DFE: In the MLAA, we do not have exactly the same structure as in the CSA: Both the number of symbols computed at each iteration and the decision function are different. Actually, the CSA computes, at each iteration, one and only one symbol, whereas the MLAA updates the $P+1$ first symbols in the delay line. Therefore, possible errors made during the first estimation can be corrected, and the error propagation phenomenon frequently observed in the DFE is limited. The absence of hard decision device in the MLAA permits us to preserve a linear estimation of the data, which can thus be considered a "soft decision device," the estimate being refined during the time the symbol is seen in the delay line (or even less if the arithmetic complexity is of major importance). The corresponding scheme is outlined in Fig. 3.

VI. Simulations

To gain more insights about the results obtained in the previous sections, we present evaluations and simulations.

A. Block Algorithm

We first evaluate the performance (in terms of MSE) of the MLBA, and we compare them with the performance of the

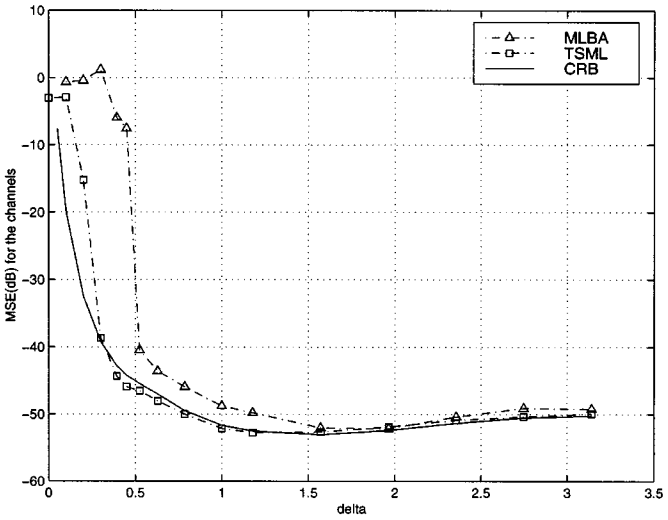

Fig. 4. MSE for the channels versus δ (diversity indicator)—Block algorithms.

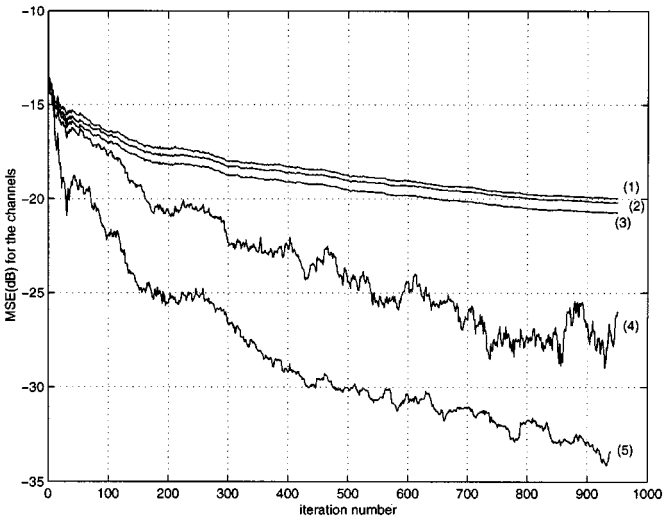

Fig. 5. MSE for the channels versus the iteration number. (1) MLRA $P = 4$. (2) MLRA $P = 10$. (3) MLRA $P = 25$. (4) MLAA $P = 10$ and $\lambda = 1 - 1/(6L(M+1))$. (5) The BGWT.

TSML. For both algorithms, we investigate the influence of the channel diversity. We consider a channel $\tilde{\mathbf{h}}$ involving two subchannels of order $M = 2$ associated with the following transfer function:

$$\tilde{\mathbf{H}}_1(z) = 1 - 2\cos(\theta)z^{-1} + z^{-2}$$
$$\tilde{\mathbf{H}}_2(z) = 1 - 2\cos(\theta + \delta)z^{-1} + z^{-2}.$$

The parameter δ is the distance between the zeros of the subchannels. Therefore, δ is an indicator of the channel diversity. The additive noise is assumed Gaussian with zero mean and variance σ_b^2. Each simulation is driven by one symbol sequence belonging to the binary alphabet $\{-1; +1\}$. The performance are measured in terms of mean-square error defined as

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

where $\hat{\mathbf{h}}^i$ stands for the ith run estimate of $\tilde{\mathbf{h}}$. N_r denotes the number of Monte Carlo runs. Here, $N_r = 500$. The CRB has been computed thanks to the formulae presented in [11].

• Fig. 4 shows the $\mathrm{MSE}_{\mathrm{dB}}(h)$ against δ (relative positions of the zeros), and the SNR is set to 45 dB. This SNR is quite unrealistic; however, this choice combines two advantages. First, it permits a meaningful comparison between MLBA and TSML with the CRB since the DML methods are known to be nonefficient at low SNR [30]. Second, it also permits a fair comparison between MLBA and TSML since TSML is biased at low SNR, whereas MLBA is not. The simulation shows that the TSML and the MLRA are close to the Cramér-Rao bound for good channel diversity conditions. The TSML offers more robustness to the diversity conditions, but this drawback of the MLBA can be overcome thanks to the introduction of a convex constraint like in [31]. The resulting algorithm is able to estimate the true channels and symbols, even when the subchannels are not coprime. A full paper is in preparation.

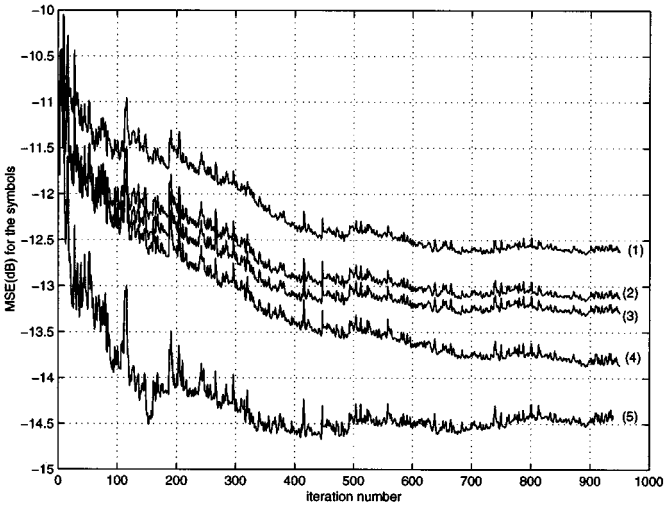

Fig. 6. MSE for the symbols versus the iteration number. (1) MLRA $P = 4$. (2) MLRA $P = 10$. (3) MLRA $P = 25$. (4) MLAA $P = 10$ and $\lambda = 1 - 1/(6L(M+1))$. (5) The BGWT.

B. Adaptive/Recursive Algorithm

In this section, we first check the relevance of the approximations used to derive the MLRA, the MLAA, and their simplified versions. First of all, we compare the performance of the MLRA and of the MLAA to the ones of the BGWT (Figs. 5 and 6). Then, the choice of the best value for P is analyzed (Figs. 7 and 8). Fig. 9 suggests that the MLAA can be turned into a LMS-like algorithm without loss of performance.

• Figs. 5 and 6 compare the MLAA and the MLRA for different values of P with the BGWT. The true channel taps $\tilde{\mathbf{h}}$ are shown in Table I. The SNR was set to 10 dB, and all algorithms have been initialized with $(\hat{\mathbf{h}}^{(0)}, \hat{\mathbf{s}}_N^{(0)})$ obtained from the MLBA ran for $N = 50$. The BGWT is optimal since at each iteration, the criterion is minimized, whereas only the first step of the minimization process is performed in the MLRA. This remark, combined with the fact that the MLRA is almost equivalent to the

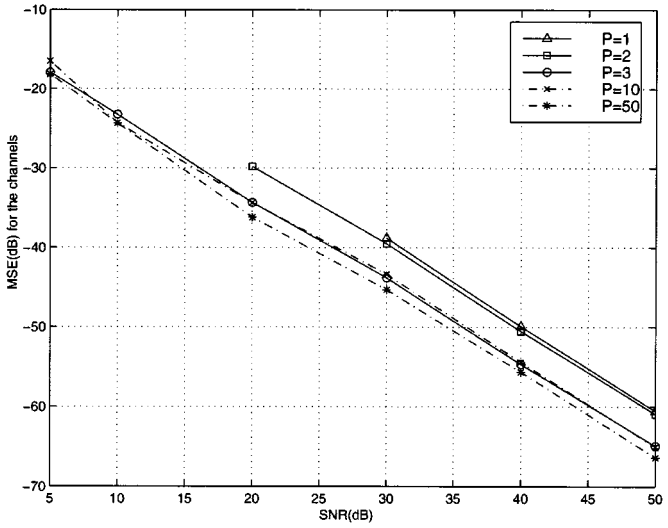

Fig. 7. MSE for the channels versus SNR obtained with the MLRA for various values of P.

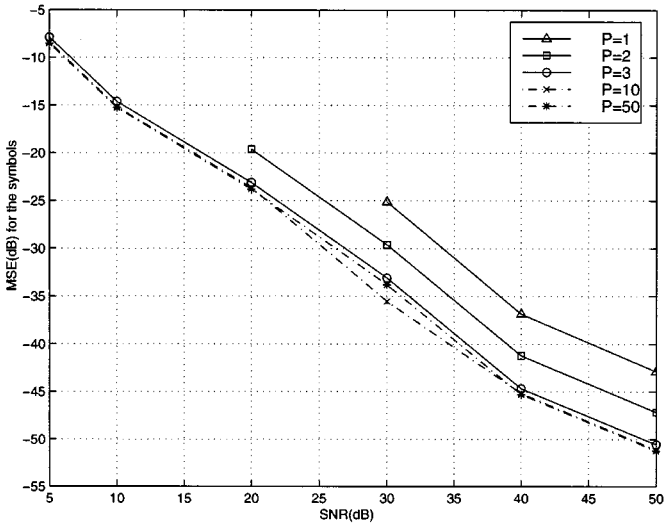

Fig. 8. MSE for the symbols versus SNR obtained with the MLRA for various values of P.

LMS algorithm with decreasing stepsize, provides an explanation for the growing gap between the MSE of the MLRA and of the BGWT and of the MLAA. Actually, the MLRA has no practical interest, but it was an essential step toward the MLAA. Figs. 5 and 6 show that the MLAA is able to improve the estimates with the arrival of new data. Moreover, the simplifications involved in the MLAA lead to small degradations on the performance.

• Figs. 7 and 8 confirm the assumption of Section IV-A dealing with the influence of old symbols on the update of the filters. In this simulation, we compute the MSE on the channels and on the symbols for the MLRA ran for various pairs (P, SNR). We consider the channel described in Table I. The MSE is averaged over 50 Monte Carlo runs (each run corresponds to an independent realization of noise) and is computed at the 1000th iteration. At low SNR and for $P = 1$ or $P = 2$, the MLRA diverges for some realizations, which is why the corresponding lines on the figures are incomplete.

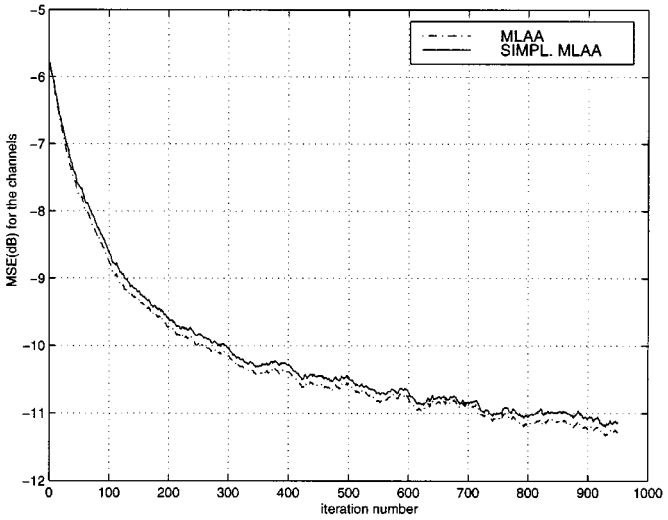

Fig. 9. MSE for the channels versus the iteration number obtained for the MLAA and with the simplified MLAA.

TABLE I
IMPULSE RESPONSE OF CHANNEL $\bar{\mathbf{h}}$

$\bar{\mathbf{h}}_1$	-0.45+0.93i	-0.46-0.74i	0.84-0.97i	-1.19-1.08i
$\bar{\mathbf{h}}_2$	-1.22-0.02i	0.43-1.23i	-0.2-0.30i	0.85+1.36i
$\bar{\mathbf{h}}_3$	-0.16+1.66i	-0.6-0.27i	0.68-0.05i	0.58-0.53i

When divergence occurs, the algorithm can be reinitialized by an appropriate procedure, which has not been done in the simulation. Note also that the reason for divergence is strongly linked to a true growing window procedure, which is known to be very sensitive to this kind of problems. Adaptive versions are less sensitive to this phenomenon, as explained below. For $P > 3$ and for any SNR, the value of the MSE remains very close to the value obtained when $P = 3$. Here, the order of the channel is $M = 3$. Therefore, the update of the filters seems to be influenced only by the symbols in the delay line of the channels.

• Fig. 9 compares the MLAA and the simplified MLAA (SMLAA—LMS-like algorithm) for $P = 3$ and $\mathrm{SNR} = 10$ dB. The SMLAA is obtained by replacing $\mathbf{R}^{(i)}$ in (29) with the diagonal matrix $(1 - \lambda)/\|\alpha_i\|^2\mathbf{I}_{L(M+1)}$, where α_i is a scale factor at iteration i between the true parameters and their estimates, and λ stands for the forgetting factor. The scale factor α_i is obtained thanks to $\|\alpha_i\| = \sqrt{M+N+i}/\|\hat{\mathbf{s}}_{N+i}^{(i)}(n+i)\|$. The MSE is averaged over 25 realizations. For each realization, the channel, the symbol sequence, and the noise realization change. In our context, the covariance matrix of the estimated symbols is nearly a diagonal matrix. The proposed simplification is justified. This is emphasized in Fig. 9.

VII. CONCLUSION

In this paper, three algorithms are proposed to implement DML methods. The MLBA has several desirable properties, including high-SNR consistency, high-SNR efficiency, and a structure suitable for deriving a recursive algorithm. Moreover, a test that permits us to circumvent the local minima problem is provided. A recursive (MLRA) and an adaptive (MLAA)

algorithm based on least-squares techniques are derived. The MLRA follows from various approximations applied to the BGWT. We can remark that both algorithms are strongly connected with a RLS. Therefore, fast versions could be obtained using techniques similar as for a fast RLS. Furthermore, the update of the filters in the MLRA and the MLAA can be simplified by stochastic gradient techniques. Derivation of the algorithms is straightforward. The MLAA combines several advantages, such as adaptivity, low arithmetic complexity (can be turned into a LMS-like algorithm without loss in the performance), and a DFE-like structure, where the soft decisions limits the error propagation phenomenon, and the local minima can be distinguished from the global minimum (which is unusual with this kind of algorithm). Moreover, the convergence speed of the MLAA can be improved by constraining the symbols into a convex set [31], which will be reported in another paper.

Appendix A
DML Criterion for a Large Amount of Data

In the noisy case, the channel and the symbols are estimated through

$$(\hat{\mathbf{h}}, \hat{\mathbf{s}}_N) = \arg\min_{\mathbf{h},\mathbf{s}_N} \|\mathbf{X}_N(n) - \mathcal{T}_N(\mathbf{h})\mathbf{s}_N\|^2$$

where $\mathbf{X}_N(n) = \mathcal{T}_N(\tilde{\mathbf{h}})\tilde{\mathbf{s}}_N + \mathbf{B}_N(n)$. Both $\tilde{\mathbf{h}}$ and $\tilde{\mathbf{s}}_N$ are deterministic parameters to be determined, whereas $\mathbf{B}_N(n)$ is stochastic. Actually, $\mathbf{B}_N(n)$ is assumed zero-mean Gaussian with covariance $\sigma_b^2\mathbf{I}$. Asymptotically, in the number of data, we have

$$\|\mathbf{P}_h^{\perp}\mathbf{X}_N(n)\|^2 \overset{N\to\infty}{\to} \operatorname{trace}(\mathbf{P}_h^{\perp}E[\mathbf{X}_N(n)\mathbf{X}_N(n)^H])$$

where $E[x]$ is the mathematical expectation w.r.t. the random variable x. Moreover, $\operatorname{trace}(E[(\mathbf{X}_N(n) - \mathcal{T}_N(\mathbf{h})\mathbf{s}_N)(\mathbf{X}_N(n) - \mathcal{T}_N(\mathbf{h})\mathbf{s}_N)^H])$ can be read $\operatorname{trace}(\mathcal{T}_N(\tilde{\mathbf{h}})\tilde{\mathbf{s}}_N - \mathcal{T}_N(\mathbf{h})\mathbf{s}_N)(\mathcal{T}_N(\tilde{\mathbf{h}})\tilde{\mathbf{s}}_N - \mathcal{T}_N(\mathbf{h})\mathbf{s}_N)^H + \sigma_b^2\mathbf{I})$. The term $\operatorname{trace}(\sigma_b^2\mathbf{I})$ is independent of $\mathbf{h}$ and $\mathbf{s}_N$, and $\operatorname{trace}(\mathcal{T}_N(\tilde{\mathbf{h}})\tilde{\mathbf{s}}_N - \mathcal{T}_N(\mathbf{h})\mathbf{s}_N)(\mathcal{T}_N(\tilde{\mathbf{h}})\tilde{\mathbf{s}}_N - \mathcal{T}_N(\mathbf{h})\mathbf{s}_N)^H)$ is the noiseless DML criterion. Then, asymptotically (in the number of data), the DML criterion is equivalent to the noiseless DML criterion.

Appendix B
Recursive Computation of the Filter

At iteration i, $\hat{\mathbf{h}}^{(i)}$ verifies the stationary point condition

$$\mathcal{U}\left(\hat{\mathbf{s}}_{N+i}^{(i)}(n+i)\right)^H \mathcal{U}\left(\hat{\mathbf{s}}_{N+i}^{(i)}(n+i)\right)\hat{\mathbf{h}}^{(i)} = \mathcal{U}\left(\hat{\mathbf{s}}_{N+i}^{(i)}(n+i)\right)^H \mathbf{X}_{N+i}(n+i). \quad (34)$$

We split the covariance matrix of the symbols into two terms in (35), shown at the bottom of the page. The first term will be updated at the next iteration, whereas the second will remain unchanged. Replace $\hat{\mathbf{h}}^{(i)}$ by $\hat{\mathbf{h}}^{(i-1)}$ in (34) to obtain a recursive solution for $\hat{\mathbf{h}}^{(i)}$ and, thanks to (35), we have

$$\begin{aligned}
&\mathcal{U}\left(\hat{\mathbf{s}}_{N+i}^{(i)}(n+i)\right)^H \mathcal{U}\left(\hat{\mathbf{s}}_{N+i}^{(i)}(n+i)\right)\hat{\mathbf{h}}^{(i-1)} \\
&\quad = \mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)^H \mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)\hat{\mathbf{h}}^{(i-1)} \\
&\quad + \left[\mathcal{U}\left(\hat{\mathbf{s}}_{N+i-1}^{(i-1)}(n+i-1)\right)^H \mathcal{U}\left(\hat{\mathbf{s}}_{N+i-1}^{(i-1)}(n+i-1)\right)\right. \\
&\quad \left. -\mathcal{U}\left(\hat{\mathbf{s}}_{P}^{(i-1)}(n+i-1)\right)^H \mathcal{U}\left(\hat{\mathbf{s}}_{P}^{(i-1)}(n+i-1)\right)\right]\hat{\mathbf{h}}^{(i-1)}.
\end{aligned} \quad (36)$$

At iteration $i-1$, the stationary point condition is

$$\mathcal{U}\left(\hat{\mathbf{s}}_{N+i-1}^{(i-1)}(n+i-1)\right)^H \mathcal{U}\left(\hat{\mathbf{s}}_{N+i-1}^{(i-1)}(n+i-1)\right)\hat{\mathbf{h}}^{(i-1)} = \mathcal{U}\left(\hat{\mathbf{s}}_{N+i-1}^{(i-1)}(n+i-1)\right)^H \mathbf{X}_{N+i-1}(n+i-1). \quad (37)$$

The product $\mathcal{U}(\hat{\mathbf{s}}_{N+i}^{(i)}(n+i))^H\mathbf{X}_{N+i}(n+i)$ can be expressed as

$$\begin{aligned}
&\mathcal{U}\left(\hat{\mathbf{s}}_{N+i}^{(i)}(n+i)\right)^H \mathbf{X}_{N+i}(n+i) \\
&\quad = \mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)^H \mathbf{X}_{P+1}(n+i) \\
&\qquad + \mathcal{U}\left(\hat{\mathbf{s}}_{N+i-1}^{(i-1)}(n+i-1)\right)^H \mathbf{X}_{N+i-1}(n+i-1) \\
&\qquad - \mathcal{U}\left(\hat{\mathbf{s}}_{P}^{(i-1)}(n+i-1)\right)^H \mathbf{X}_{P}(n+i-1).
\end{aligned} \quad (38)$$

$$\begin{aligned}
&\mathcal{U}\left(\hat{\mathbf{s}}_{N+i-1}^{(i-1)}(n+i-1)\right)^H \mathcal{U}\left(\hat{\mathbf{s}}_{N+i-1}^{(i-1)}(n+i-1)\right) \\
&\quad = \mathcal{U}\left(\hat{\mathbf{s}}_{P}^{(i-1)}(n+i-1)\right)^H \mathcal{U}\left(\hat{\mathbf{s}}_{P}^{(i-1)}(n+i-1)\right) \\
&\qquad + \underbrace{\mathcal{U}\left(\hat{\mathbf{s}}_{N+i-P-1}^{(i-1)}(n+i-P-1)\right)^H \mathcal{U}\left(\hat{\mathbf{s}}_{N+i-P-1}^{(i-1)}(n+i-P-1)\right)}_{=\mathcal{U}\left(\hat{\mathbf{s}}_{N+i-P-1}^{(i)}(n+i-P-1)\right)^H \mathcal{U}\left(\hat{\mathbf{s}}_{N+i-P-1}^{(i)}(n+i-P-1)\right)}
\end{aligned} \quad (35)$$

Replacing the previous equations in (36) leads to

$$\begin{aligned}
&\mathcal{U}\left(\hat{\mathbf{s}}_{N+i}^{(i)}(n+i)\right)^H \mathcal{U}\left(\hat{\mathbf{s}}_{N+i}^{(i)}(n+i)\right)\left[\hat{\mathbf{h}}^{(i-1)}\right. \\
&\quad + \left[\mathcal{U}\left(\hat{\mathbf{s}}_{N+i}^{(i)}(n+i)\right)^H \mathcal{U}\left(\hat{\mathbf{s}}_{N+i}^{(i)}(n+i)\right)\right]^{(-1)} \\
&\quad \times \left\{\mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)^H \left(\mathbf{X}_{P+1}(n+i)\right.\right. \\
&\quad -\mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)\hat{\mathbf{h}}^{(i-1)}\Big) \\
&\quad -\mathcal{U}\left(\hat{\mathbf{s}}_{P}^{(i-1)}(n+i-1)\right)^H \left(\mathbf{X}_P(n+i-1)\right. \\
&\quad \left.\left.\left.-\mathcal{U}\left(\hat{\mathbf{s}}_{P}^{(i-1)}(n+i-1)\right).\hat{\mathbf{h}}^{(i-1)}\right)\right\}\right] \\
&= \mathcal{U}\left(\hat{\mathbf{s}}_{N+i}^{(i)}(n+i)\right)^H \mathbf{X}_{N+i}(n+i).
\end{aligned}$$

Finally, comparing (39) with (34) leads to the update equation for the filters

$$\begin{aligned}
\hat{\mathbf{h}}^{(i)} = &\hat{\mathbf{h}}^{(i-1)} + \left[\mathbf{R}^{(i)}\right]^{(-1)} \left[\mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)^H \left(\mathbf{X}_{P+1}(n+i)\right.\right. \\
&\left.-\mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)\hat{\mathbf{h}}^{(i-1)}\right) - \mathcal{U}\left(\hat{\mathbf{s}}_{P}^{(i-1)}(n+i-1)\right)^H \\
&\left.\times \left(\mathbf{X}_P(n+i-1) - \mathcal{U}\left(\hat{\mathbf{s}}_{P}^{(i-1)}(n+i-1)\right)\hat{\mathbf{h}}^{(i-1)}\right)\right]
\end{aligned}$$

where $\mathbf{R}^{(i)} = \mathcal{U}(\hat{\mathbf{s}}_{N+i}^{(i)}(n+i))^H \mathcal{U}(\hat{\mathbf{s}}_{N+i}^{(i)}(n+i))$. $\mathbf{R}^{(i)}$ is also updated recursively:

$$\begin{aligned}
\mathbf{R}^{(i)} = \mathbf{R}^{(i-1)} &+ \mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)^H \mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right) \\
&-\mathcal{U}\left(\hat{\mathbf{s}}_{P}^{(i-1)}(n+i-1)\right)^H \mathcal{U}\left(\hat{\mathbf{s}}_{P}^{(i-1)}(n+i-1)\right).
\end{aligned}$$

Then, applying twice the Woodbury's identity, the recursive equations for the update of $\mathbf{R}^{(i)}$ is

$$\begin{aligned}
\mathbf{A}^{-1} = &\left[\mathbf{R}^{(i-1)}\right]^{-1} - \left[\mathbf{R}^{(i-1)}\right]^{-1} \mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)^H \\
&\times \left[\mathbf{I}_{L(P+1)} + \mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)\left[\mathbf{R}^{(i-1)}\right]^{-1}\right. \\
&\left.\times \mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)^H\right]^{-1} \\
&\times \mathcal{U}\left(\hat{\mathbf{s}}_{P+1}^{(i)}(n+i)\right)\left[\mathbf{R}^{(i-1)}\right]^{-1} \\
&\times [\mathbf{R}^{(i)}]^{-1} = \mathbf{A}^{-1} \\
&+ \mathbf{A}^{-1}\mathcal{U}\left(\hat{\mathbf{s}}_{P}^{(i-1)}(n+i-1)\right)^H \\
&\times \left[\mathbf{I}_{L.P} - \mathcal{U}\left(\hat{\mathbf{s}}_{P}^{(i-1)}(n+i-1)\right)\right. \\
&\left.\times \mathbf{A}^{-1}\mathcal{U}\left(\hat{\mathbf{s}}_{P}^{(i-1)}(n+i-1)\right)^H\right]^{-1} \\
&\times \mathcal{U}\left(\hat{\mathbf{s}}_{P}^{(i-1)}(n+i-1)\right)\mathbf{A}^{-1}
\end{aligned}$$

where $\mathbf{A}$ is an intermediary matrix.

ACKNOWLEDGMENT

The authors wish to thank the anonymous reviewers for insightful comments and helpful suggestions.

REFERENCES

[1] E. de Carvalho and D. Slock, "Semi-blind methods for FIR multichannel estimation," in *Signal Processing Advances in Wireless Communications*, Y. Hua, G. B. Giannakis, P. Stoica, and L. Tong, Eds. Englewood Cliffs, NJ: Prentice-Hall, 2000.

[2] D. Hatzinakos and C. Nikias, "Estimation of multipath channel response in frequency selective channels," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 12–19, Jan. 1989.

[3] J. K. Tugnait, "Blind equalization and estimation of digital communication FIR channels using cumulant matching," *IEEE Trans. Commun.*, vol. 43, pp. 1240–1245, Feb. 1995.

[4] B. Porat and B. Friedlander, "Blind equalization of digital communication channels using higher-order moments," *IEEE Trans. Signal Prossessing*, vol. 39, pp. 522–526, Feb. 1991.

[5] G. B. Giannakis and J. M. Mendel, "Identification of nonminimum phase systems using higher-order-satistics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 360–377, Mar. 1989.

[6] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channels," presented at the Proc. 25th Asilomar Conf, Signals, Syst., Comput., Nov. 1991.

[7] ——, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. Inform. Theory*, vol. 40, pp. 340–349, Mar. 1994.

[8] Y. Li and Z. Ding, "Blind channel identification based on second-order cyclostationary statistics," in *Proc. ICASSP*, vol. 4, Apr. 1993, pp. 81–84.

[9] K. Abed-Meraim, E. Moulines, and P. Loubaton, "Prediction error method for second-order blind identification," *IEEE Trans. Signal Processing*, vol. 45, pp. 694–705, Mar. 1997.

[10] Y. Hua, H. Yang, and W. Qiu, "Source correlation compensation for blind channel identification based on second-order statistics," *IEEE Signal Processing Lett.*, vol. 1, pp. 119–120, Aug. 1994.

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

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

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

[14] G. Harikuma and Y. Bresler, "Analysis and comparative evaluation of techniques for multichannel blind deconvolution," *Proc. Eighth IEEE Signal Process. Workshop Statist. Signal Array Processing*, pp. 332–335, 1996.

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

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

[17] P. Stoica, J. Li, and T. Sderstrm, "On the iconsistency of IQML," *Signal Process.*, vol. 56, pp. 185–190, July 1997.

[18] J. Ayadi, E. de Carvalho, and D. T. M. Slock, "Blind and semi-blind maximum likelihood methods for FIR multichannel identification," presented at the ICASSP, WA, May 1998.

[19] D. Gesbert, P. Duhamel, and S. Mayrargue, "Blind least squares criteria for joint data/channel estimation," presented at the IEEE DSP Workshop, Sept. 1996.

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

[21] J. Laurila, K. Kopsa, and E. Bonek, "Semi-blind signal estimation for smart antennas using subspace tracking," in *Proc. SPAWC*, Annapolis, MD, 1999, pp. 271–274.

[22] E. Pité and P. Duhamel, "Bilinear methods for blind channel equalization: (No) local minimum issue," in *Proc. ICASSP*, May 1998.

[23] R. R. Bitmead, S. Y. Kung, B. D. O. Anderson, and T. Kailath, "Greatest common division via generalized Sylvester and Bezout matrices," *IEEE Trans. Automat. Contr.*, vol. AC-23, pp. 1043–1047, Dec. 1978.

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

[25] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
[26] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 516–526, Feb. 1995.
[27] G. Xu and L. Tong, "A deterministic approach to blind channel identification of multi-channel FIR systems," presented at the ICASSP, Apr. 1994.
[28] M. Gurelli and C. Nikias, "EVAM: An eigenvector-based algorithm for multi-channel blind deconvolution of input colored signals," *IEEE Trans. Signal Processing*, vol. 43, pp. 134–149, Jan. 1995.
[29] J. Proakis, *Digital Communications*. New York: Mc Graw-Hill, 1989.
[30] E. de Carvalho and D. T. M. Slock, "Asymptotic performance of ML methods for semi-blind channel estimation," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1997.
[31] F. Alberge, P. Duhamel, and M. Nikolova, "Blind identification/equalization using deterministic maximum likelihood and a partial information on the input," in *Proc. SPAWC*, Annapolis, MD, 1999.

Florence Alberge was born in Albi, France, in 1971. She received the B.E. degree from Ecole Nationale Supérieure de l'Electronique et de ses Applications, Paris, France, and the M.Sc. degree from the University of Cergy, Cergy, France, in 1996. She received the Agregation de Physique degree in 1998 and the Ph.D. degree in 1999 from the Ecole Nationale Supérieure des Télécommunications (ENST), Paris.

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

Pierre Duhamel (F'98) was born in France in 1953. He received the Dipl.-Ing. degree in electrical engineering from the National Institute for Applied Sciences (INSA), Rennes, France, in 1975 and the Dr.-Ing. degree in 1978 and the Doctorat ès sciences degree in 1986, both from Orsay University, Orsay, France.

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 CCD 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 September 2000, he was a Professor with the National School of Engineering in Telecommunications (ENST), 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, CNRS, Gif sur Yvette, France, where he is developing studies in signal processing for communications (including equalization, iterative decoding, and multicarrier systems) and signal/image processing for multimedia applications, including source coding, joint source/channel coding, watermarking, and audio processing.

Dr. Duhamel was chairman of the IEEE DSP committee from 1996 to 1998, was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1989 to 1991, and was Associate Editor for the IEEE SIGNAL PROCESSING LETTERS. He was a Guest Editor for the special issue of the IEEE TRANSACTIONS ON SIGNAL PROCESSING on wavelets. He is now a member of the Signal Processing for Communications committee. He was an IEEE Distiguished Lecturer for 1999 and was co-general chair of the 2001 International Workshop on Multimedia Signal Processing, Cannes, France. The paper on subspace-based methods for blind equalization, which he co-authored, received the "Best Paper Award" from the IEEE TRANSACTIONS ON SIGNAL PROCESSING in 1998. He was awarded the "Grand Prix France Telecom" by the French Science Academy in 2000.

Mila Nikolova received the Ph.D. degree in signal processing from the Université Paris-Sud, Paris, France, in 1995.

Currently, she is a Researcher with the National Center for Scientific Research (CNRS), Paris. Her research interests are in inverse problems and reconstruction of images and signals.