ADAPTIVE DETERMINISTIC MAXIMUM LIKELIHOOD USING A QUASI-DISCRETE PRIOR

Florence Alberge, Mila Nikolova and Pierre Duhamel

ENST/TSI, 46, rue Barrault, 75634 Paris Cedex 13, France

ABSTRACT
A block algorithm is presented to solve the joint blind channel identification and blind symbol estimation problem. It is based on a Deterministic Maximum Likelihood (DML) method. A partial prior on the symbols is incorporated into the DML criterion in order to improve the estimation accuracy. We propose a test which permits to circumvent the local minima problem and which is pertinent for a large class of criteria. The structure of the block algorithm is well-suited for deriving recursive and adaptive versions. We prove that, in the noiseless case, the obtained recursive algorithm converges only towards the global minimum. Numerical results show that the prior on the symbols improves the accuracy of the estimators and brings robustness to the lack of channel diversity. At the same time, this method introduces fewer local minima than the use of a full prior.

1. INTRODUCTION
Blind identification is an important problem in many areas and especially in wireless communications. Our work addresses Single Input Multiple Outputs (SIMO) systems. The SIMO equalization problem can be solved using second-order statistics only, as long as the sub-channels have no common zeros. In this paper, we focus on Deterministic Maximum Likelihood (DML) methods in order to estimate both the channels and the symbols. Among the major contributions to DML, we can cite The Two Steps Maximum Likelihood (TSM) [1] and the Iterative Quadratic Maximum Likelihood (IQML) [2]. Another DML based approach is the MLB (Maximum Likelihood Block Algorithm) proposed by the authors in [3, 4]. The knowledge that the input symbols belong to a finite alphabet constitutes a strong prior information. Its use can improve the estimation quality. Such algorithms were first proposed by Seshadrin [5] and Ghosh and Weber [6]. The problem with these methods is that their convergence is not guaranteed in general and the incorporation of the finite alphabet property often increases the local minima problem. In this paper, we propose to use only a partial prior on the input. Then, the number of additional local minima is limited and the estimation accuracy is improved as well. This partial prior is introduced into the MLB since the formulation of the latter is well-suited for deriving adaptive algorithms which is very important when working in a fast fading environment. We first recall several properties of the MLB [7, 8], we lay particular stress on a test which permits to circumvent the local minima problem. Next, a block algorithm using a partial prior is derived paying attention that the test remains valid. We also propose a recursive and an adaptive version of it and we prove that the recursive algorithm converges only towards the global minimum. Numerical simulations show the improvements due to the use of the partial prior.

2. PROBLEM FORMULATION
Consider a sequence of unknown binary symbols \( \{ \hat{s}(k) \} \) transmitted through \( L \) unknown channels \( h_i, 1 \leq i \leq L \). Let \( x_i(k) \) denote the output of the \( i^{th} \) channel at time \( k \), and \( X_N(n) = [x_1(n), \ldots, x_L(n - N + 1)]^T \) the LN-length vector obtained by interleaving the output of the various channels. This output is modeled as:

\[
X_N(n) = T_N(h) \tilde{s}_N(n) + B_N(n)
\]

where \( \tilde{h}(k) = [\tilde{h}_1(k), \ldots, \tilde{h}_L(k)]^T \) are the different channels and \( B_N(n) \) stands for the noise vector. The noise sequence on each sensor is assumed to be i.i.d. and the different sequences are mutually uncorrelated. In (1), the operator \( T_N \) transforms a sequence of channels \( h(k) = [h_1(k), \ldots, h_L(k)]^T \) into the following LN x (M+N) generalized Sylvester matrix:

\[
T_N(h) = \begin{pmatrix}
    h(0) & \ldots & h(M) \\
    \vdots & \ddots & \vdots \\
    h(0) & \ldots & h(M)
\end{pmatrix}
\]

Below, \( s_N(n) = [s(n), \ldots, s(n - N - M + 1)]^T \) denotes any vector of \( M + N \) symbols, where \( M \) is the maximum order of a channel. Operator \( \mathcal{U} \) transforms a vector \( s_N(n) \) into a LN x L(M+1) matrix, \( \mathcal{U}(s_N(n)) \), in such a way that [3]:

\[
\mathcal{U}(s_N(n))h = T_N(h)s_N(n), \forall s_N, \forall h
\]

In the following, we assume that both matrices \( T_N(h) \) and \( \mathcal{U}(s_N) \) are full column rank, that \( M \) is known, that \( \{ \hat{s}(k) \} \) has linear complexity \( 2M + 1 \) or greater (persistent excitation) and that the modulation is a PSK (Phase Shift Keying). We also consider that \( \hat{h} \) is not varying.

3. REVIEW OF THE MLB
The work presented here is based on the MLB [3, 4, 7], briefly recalled below. We consider the minimization of criterion \( J(h, s_N(n)) = \|X(n) - T_N(h)s_N(n)\|^2 \) which is the classical DML criterion. Using (2), \( J \) can also be
rewritten as $J(h, s_N(n)) = \|X_N(n) - U(s_N(n))h\|^2$. The MLBA consists in iterating:

$$\hat{h}_N^{(k)} = \left[U(s_N^{(k-1)})^H U(s_N^{(k-1)}) + \gamma I\right]^{-1} U(s_N^{(k-1)})^H X_N(n)$$  \hspace{1cm} (3)

$$\hat{s}_N^{(k)} = \left[T_N(\hat{h}_N^{(k)})^H T_N(\hat{h}_N^{(k)}) + \gamma I\right]^{-1} T_N(\hat{h}_N^{(k)})^H X_N(n)$$  \hspace{1cm} (4)

where $T_N(h^{(k)})$ and $U(s_N^{(k)})$ are assumed to be full column rank. The following properties hold for the MLBA [7, 3]:

(P1) **Uniqueness of the global minimum:**
In the noiseless case, if $T_N(h)$ is full column rank and if the sequence of emitted symbols is sufficiently exciting then, the global minimum of $J(h, s_N(n))$ is obtained only for the true channels and symbols (see [7] for the proof).

(P2) **Stability of the estimates in a recursive procedure:**
Consider a growing window procedure (i.e. the solution in terms of $h$ and $s_N$ is updated at step $k$ by equations (3-4) and initialized by the result of the step $k - 1$). If all the intermediate matrices $T_{N+k}(h^{(k)})$ and $U(s_N^{(N+k)} + n + k)$ are full column rank and under the same conditions as required for (P1), we proved in [3] that the only stable solution during at least $3M + 1$ iterations (in the sense that the solution $h^{(k)} = h^{(k)}$, $i = 1...3M + 1$) is the global minimum. In other words, the same local minimum cannot be a solution of more than $3M$ consecutive steps.

The accuracy of DML estimators can be improved thanks to the incorporation of some prior information. A full use of such a knowledge, generally introduces many local minima. Instead, we propose to consider only a partial prior on the symbols to limit this phenomenon. In [3], we proposed to introduce a priori information on the symbols, i.e. constraining the estimated symbols to belong to $E^{M+N} = \{ s_N : \| s_N \|^2 \leq 1 \}$. This set is convex and contains all possible symbols. We proved in [3] that this constraint does not add any additional minimum and that (P1) and (P2) hold. On the other hand, the constrained algorithm outperforms the MLBA only in some precise situations. Here, we propose a new criterion which improves the estimation accuracy and which is likely to exhibit fewer local minima than an algorithm using a full prior. The two sufficient conditions above still hold, hence the local minima problem can be circumvented.

4. **Conditional Maximum Likelihood**

The approach proposed here consists in attributing to the symbols a probability distribution function (pdf) reflecting our prior knowledge on the input. When using a uniform prior, the same probability is attributed to each symbol in $E^{M+N}$ and the corresponding pdf is continuous. With a full prior, a discrete pdf is considered which gives the same non-zero probability to the values of the alphabet and zero elsewhere. Instead, we propose to use a continuous pdf which “encourages” the estimated symbols to take values on the boundaries of $E^{M+N}$ as follows:

$$\left\{
\begin{array}{ll}
p(s_k) = 0 & \text{if } \| s_k \| > 1 \\
p(s_k) = \frac{1}{2} e^{\beta s_k^2} & \text{if } \| s_k \| \leq 1
\end{array}
\right.\quad p(s) = \Pi_k p(s_k)$$

where $Z$ is a normalization term, $\beta$ is a positive scalar and $s_k$ is the $k$th component of $s$. The function $p(s_k)$ is plotted in fig. (1) for three values of $\beta$. For $\beta = 0$, we obtain the uniform pdf. When $\beta \to \infty$ and for a BPSK modulation, $p(s_k)$ approaches the discrete pdf corresponding to the full prior. Let $p(X_N(n)|h, s_N)$ denote the likelihood function conditioned on both the channels and the symbols then the conditional likelihood function $p(X_N(n)|h, s_N)$ reads:

$$p(X_N(n)|h, s_N) = p(X_N(n)|h) p(s_N)$$  \hspace{1cm} (5)

$B_N(n)$ is assumed Gaussian then, $p(X_N(n)|h, s_N)$ reads:

$$p(X_N(n)|h, s_N) = C_1 e^{\left(-\frac{1}{2\sigma^2} \|X_N - T_N(h)s_N\|^2\right)}$$  \hspace{1cm} (6)

Inserting (6) into (5), we obtain:

$$\left\{
\begin{array}{ll}
p(X_N, s_N|h) = C_2 e^{\left(-\frac{1}{2\sigma^2} \|X_N - T_N(h)s_N\|^2 + \beta \|s_N\|^2\right)} & \text{if } \|s_N\| \leq 1 \\
p(X_N, s_N|h) = 0 & \text{elsewhere}
\end{array}
\right.$$

The maximization of $p(X_N, s_N|h)$ is equivalent to the maximization of the following criterion over $C^{M+N} \times E^{M+N}$:

$$\mathcal{L}(h, s_N(n)) = \|X(n) - T_N(h)s_N(n)\|^2 - \gamma \|s_N(n)\|^2$$  \hspace{1cm} (7)

$T_N(h)^H T_N(h)$ is assumed full rank, so $\mathcal{L}$ will be convex with respect to each variable separately if $\gamma \geq \lambda_{\text{min}}$ where $\lambda_{\text{min}} = \min(\text{tr}(T_N(h)^H T_N(h)))$. The criterion in (7) is almost similar to the one proposed in [8] for an image reconstruction problem. The minimization of $\mathcal{L}(h, s_N(n))$ is solved in an iterative way as described below:

$$\hat{h}_N^{(k)} = \left[U(s_N^{y(k-1)})^H U(s_N^{y(k-1)}) + \gamma I\right]^{-1} U(s_N^{y(k-1)})^H X_N(n)$$  \hspace{1cm} (8)

$$\hat{s}_N^{(k)} = \arg\min_{s_N \in E^{M+N}} \mathcal{L}(h_N^{(k)})(s_N)$$  \hspace{1cm} (9)

where $\mathcal{L}(h_N^{(k)})(s_N) = \|X(n) - T_N(h_N^{(k)})s_N(n)\|^2 - \gamma \|s_N(n)\|^2$.

Eq. (8-9) form the CMLBA (Conditional MLBA). The optimization problem in (9) is solved by a relaxation method. Let $s_j$ denote the $j$th component of $s_N$. At each iteration of the relaxation method, $s_j$ is computed by:

$$s_j = -p_j C_j A_j C_j^H s_j$$  \hspace{1cm} (10)

$A_j = (T_j^{(k)})^H T_j^{(k)}$, $B_j = (T_j^{(k)})^H \sum_{i \neq j} T_i^{(k)} s_i$, $C_j = (T_j^{(k)})^H X_N(n)$

where $P$ is the projection operator of $C^{M+N}$ onto $E^{M+N}$ and $T_j^{(k)}$ is the $j$th column of $T_N(h^{(k)})$. 

![Figure 1: $p(s_k)$ for $\beta = 2$, $\beta = 0.5$ and $\beta = 0.1$ and a BPSK modulation](image-url)
One can easily verify that if \( \hat{h}(\hat{s}, \hat{s}_N(n)) \) is a global minimizer of \( \mathcal{L}(\hat{h}, \hat{s}_N(n)) \) then it is also a global minimizer of \( \mathcal{J}(\hat{h}, s_N(n)) \). Hence, \( \text{(P1)} \) holds for the CMLBA (i.e. the true channels and symbols are the only global minimizers of \( \mathcal{L} \)). The proof of \( \text{(P2)} \) relies exclusively on the derivative of \( \mathcal{J}(\hat{h}, s_N) \) w.r.t. the filter. (see [3]). Condition \( \mathcal{L} \) satisfies:

\[
\frac{\partial \mathcal{L}}{\partial \hat{h}} = \frac{\partial \mathcal{J}}{\partial \hat{h}} = \mathcal{U}(s_N(n)) \mathcal{H}^{(n)}(\mathcal{U}(s_N(n)) \hat{h} - X_N(n))
\]

Then, \( \text{(P2)} \) holds for the CMLBA. The same properties hold for the unconstrained algorithm and for the CMLBA. Hence, one can expect them to be more widely met, since eq. (10) can be met by other algorithms. Whether this sufficient condition is very frequent or not, and whether weaker requirements would be a necessary condition is still a topic for further study.

The property \( \text{(P2)} \) proves that the global minimum is the only stable point for \( K = 3M + 1 \) consecutive steps. This justifies the derivation of recursive algorithms which have the additional advantage to require a lower arithmetic complexity than the growing window procedure.

5. RECURSIVE ALGORITHM

The recursive algorithm is such that, given the least squares estimates of the symbols and of the filters at iteration \( k - 1 \), the estimates of these vectors are updated at iteration \( k \) upon the arrival of a new symbol. It makes use of the following simplifications:

(S1) The iterative minimization w.r.t. the joint variable in the growing window procedure is replaced by a minimization w.r.t. each variable separately. Hence, at step \( k \), we calculate:

\[
\hat{s}_{N+k}^{(k)}(n+k) = \arg \min_{s_{N+k}^{(k+1)}} \mathcal{L}(\hat{h}^{(k-1)}, s_{N+k})
\]

\[
\hat{h}^{(k)} = \arg \min_{\hat{h}} \mathcal{L}(\hat{h}, \hat{s}_{N+k}^{(k)}(n+k))
\]

We can remark that these equations coincide with the first iteration of step \( k \) of the growing window procedure.

(S2) At iteration \( k \), eq. (11) updates \( M + N + k \) symbols. Hence, the computational complexity involved in eq. (11) increases with \( k \). Here, we propose to compute, at step \( k \), the new emitted symbol while updating only the previous \( Q \) (independent of \( k \)) symbols in the delay line where \( Q \) is a parameter to be determined. Implicitly, the previous symbols (farther than \( Q \)) are supposed not to be affected by the incoming measurement. Then, eq. (11) reads:

\[
\hat{s}_{Q+1-M}^{(k)}(n+k) = \arg \min_{s_{Q+1}} \mathcal{L}(\hat{h}^{(k-1)}, F(Z))
\]

where \( F(Z) = [Z^T \hat{s}_0^{(k-1)}(n+k-Q-1)]^T \).

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

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

5.1. Minimization with respect to the symbols

The optimization problem in (13) is solved by a relaxation method. Let \( s_j \) denote the \( j \)th component of \( s \). Then, \( \hat{s}_{Q+1-M}^{(k)}(n+k) \) is the stationary point obtained through the following iterative algorithm:

\[
s_j = -P \left( \frac{B_j - C_j}{A_j^j} \right) \quad j = 1, \ldots, Q + 1
\]

\[
A_j = \left(T_j^{(k-1)} \right)^H T_j^{(k-1)} \quad C_j = \left(T_j^{(k-1)} \right)^H X_{Q+1}(n)
\]

\[
B_j = \left( T_j^{(k-1)} \right)^H \sum_{l=1}^{Q+1+M} \left(T_l^{(k-1)} \right) X_{Q+1}(n)
\]

where \( T_j^{(k-1)} \) is the \( j \)th column of \( T_{Q+1}(\hat{h}^{(k-1)}) \).

5.2. Minimization with respect to the filter

Since \( \mathcal{L} \) satisfies (10), the update of the filter can be performed in the same way for \( \mathcal{L} \) and \( \mathcal{J} \). A recursive update algorithm for the channels corresponding to \( \mathcal{J} \) has already been presented in [4]. The corresponding equation is:

\[
\hat{h}^{(i)} = \hat{h}^{(i-1)} + \left( R^{(i)} \right)^{-1} \left\{ \mathcal{U}(\hat{s}_Q^{(i)}(n+i)) [X_{Q+1}(n+i) \right.
\]

\[
- \mathcal{U}(\hat{s}_Q^{(i)}(n+i)) \hat{h}^{(i-1)} - \mathcal{U}(\hat{s}_Q^{(i-1)}(n+i-1)) X_{Q+1}(n+i-1) \right\}
\]

\[
- \mathcal{U}(\hat{s}_Q^{(i-1)}(n+i-1)) \hat{h}^{(i-1)} \right\}
\]

where \( R^{(i)} = \mathcal{U}(\hat{s}_Q^{(i)}(n+i)) \mathcal{H}(\hat{s}_Q^{(i)}(n+i)) \). \( R^{(i)} \) is computed from \( R^{(i-1)} \) using:

\[
R^{(i)} = R^{(i-1)} + \mathcal{U}(\hat{s}_Q^{(i)}(n+i)) \mathcal{H}(\hat{s}_Q^{(i)}(n+i)) \]

\[
- \mathcal{U}(\hat{s}_Q^{(i-1)}(n+i-1)) \mathcal{H}(\hat{s}_Q^{(i-1)}(n+i-1))
\]

The recursive update of \( R^{(i-1)} \) is then obtained by applying twice the matrix inversion lemma. Eq. (14) and (15) form the CMLRA (CML Recursive Algorithm).

The following convergence result holds for the CMLRA:

Theorem: If the assumptions of \( \text{(P2)} \) are met and if the CMLRA converges, then it converges towards the global minimum.

The proof relies on the derivative of the criterion with respect to the filter and is similar to the proof in [4] established for the recursive version of the MLBA. Then, this theorem also holds for an algorithm using a full prior. However, the theorem above is of practical interest as long as there are few local minima (since our result requires that the algorithm converges) which is more likely to occur with a partial prior.

5.3. Initialization

The CMLRA needs to start from a reliable initialization point. Here, it is initialized with \( \mathcal{H}(\hat{s}^{(0)}(n)), \hat{s}^{(0)}(n) \) defined as the stationary point of the block algorithm in (8-9). When \( N \) is small, calculating \( \mathcal{H}(\hat{s}^{(0)}(n)), \hat{s}^{(0)}(n) \) involves few computations.

5.4. Adaptive algorithms

The CMLRA is easily transformed into an adaptive algorithm by considering a criterion involving an exponential forgetting factor.
6. SIMULATIONS

We first compare the performances of the CMLBA with the EM (Expectation-Maximization) algorithm [9] in a digital wireless communication situation at 900MHz. The EM algorithm provides good quality estimation when it converges towards the global minimum. But, since it uses a full prior on the symbols, its implementation is often complicated by the existence of numerous local minima. The CMLBA uses only a partial prior which is likely to introduce fewer local minima than a full prior. In fig (2), the BER (Bit Error Rate) against the SNR is plotted for both algorithms. We also plot the graph obtained with the MLBA. We consider a BPSK modulation and a semi-blind setup. The results are averaged over 1000 runs where \( \hat{h}, \hat{s}_N \) and the noise vector changes at each iteration. The BER obtained for the CMLBA is very close to that of the EM algorithm which underlines the efficiency of the proposed criterion.

Now, we investigate the influence of the prior on the estimation accuracy. In fig. (2), we compare the CMLBA with the MLBA [4] and with the MLBA-UP (MLBA using a Uniform Prior – \( \beta = 0 \)). A BPSK modulation is considered, \( h_1 = [1 -2 \text{ ]}, h_2 = [1 -2\cos(\theta) \text{ ]} \) and SNR = 45dB. We plot the MSE on the channels against \( \delta \) (diversity indicator). The MSE is averaged over 100 independent noise realizations. In situations of sufficient diversity the MLBA and the MLBA-UP have similar performances. Whereas, in situations of poor channel diversity, the MLBA-UP outperforms the MLBA thanks to the additional constraint. In all situations, the CMLBA outperforms the others algorithms.

To get valuable insights into the improvements brought by the use of the prior, we plot the histograms of the estimated symbols obtained for the MLBA-UP and the CMLBA. The length of the symbol vector is 600, \( h_1 = [0.5 + 0.3126i - 0.5173 + 2.69i - 0.56 + 0.2887i - 0.7534 - 1.4224], h_2 = [0.93 + 0.2468i - 0.2485 - 1.4358i - 0.1498 + 0.1486i - 1.2584 - 1.693i] \) and the SNR is fixed to 10dB. The MLBA-UP corresponds to eq. (8-9) with \( \gamma = 0 \) and is equivalent to the minimization of \( J \) with constraint: \( s_N \in \mathcal{E}^{M+N} \). Hence, it has small influence on the repartition of the symbols in \( \mathcal{E}^{M+N} \). At the opposite, the CMLBA prioritizes the extreme values of the set. Fig. (4) show that the estimated symbols are concentrated near the bounds of the interval when using the CMLBA.

7. CONCLUSION

In this paper, a block algorithm based on DML methods is derived. It includes a partial prior information on the symbols which improves the estimation accuracy without introducing as much local minima as a full prior. We also proposed a test to circumvent the local minima problem which is pertinent for a large class of criteria. Since the CMLBA is not adapted for working with large data sequences, a recursive version of the latter is derived. We proved in the noiseless case that, when the recursive algorithm converges then it converges towards the global minimum. The adaptivity property is obtained thanks to a weighting factor. Numerical results show the improvements inherent to the use of the partial prior.

8. REFERENCES