MAXIMUM-LIKELIHOOD BLIND FIR MULTI-CHANNEL ESTIMATION WITH GAUSSIAN PRIOR FOR THE SYMBOLS

Elisabeth De Carvalho  Dirk T.M. Slock
Institut EURECOM
B.P. 193, 06904 Sophia Antipolis Cedex, FRANCE
{carvalho,slock}@eurecom.fr

ABSTRACT

We present two approaches to stochastic Maximum Likelihood identification of multiple FIR channels, where the input symbols are assumed Gaussian and the channel deterministic. These methods allow semi-blind identification, as they accommodate a priori knowledge in the form of a (short) training sequence and appear to be more relevant in practice than purely blind techniques. The two approaches are parameterized both in terms of channel coefficients and in terms of prediction filter coefficients. Corresponding methods are presented and some are simulated. Furthermore, Cramer-Rao Bounds for semi-blind ML are presented: a significant improvement of the performance for a moderate number of known symbols can be noticed.

1. INTRODUCTION

Consider a sequence of symbols $a(k)$ received through $m$ channels $y(k) = \sum_{i=0}^{N-1} h(i) a(k-i) + \eta(k)$, where $h = H A N$ and $\eta = \eta(a(k-i)) = \mathbb{E}[\eta(k) | a(k-i)] = \sigma^2 I_m \delta_i$. Assume we receive $M$ samples:

$$Y_M = T_N H_A M + V_M$$

where $Y_M = [y_M(k-M+1) \ldots y_M(k)]$ and similarly for $V_M(k)$, and $T_N (H_A)$ is a block Toeplitz matrix with $M$ block rows and $[H N \ O_{m \times (M-1)}]$ as first block row. We shall simplify the notation in (1) with $k = M-1$ to

$$Y = T(A) + V$$

We assume that $m M > M+N-1$ in which case the channel convolution matrix $T(A)$ has more rows than columns. If the $H(z), i = 1, \ldots, m$ have zero in common, then $T(A)$ has full column rank (which we will henceforth assume). For obvious reasons, the column space of $T(A)$ is called the signal subspace and its orthogonal complement the noise subspace.

We propose two stochastic Maximum Likelihood (ML) approaches in which the symbols are assumed Gaussian. These methods can easily accommodate a priori knowledge.

The work of Elisabeth de Carvalho was supported by Laboratoires d’Electronique PHILIPS under contract Gire 29T95. The work of Dirk Slock was supported in part by Laboratoires d’Electronique PHILIPS under contract LEP 95FR8008.
In a second approach, both $h$ and $A$ are being estimated as well as $R_{VV}$, maximum a posteriori (MAP) for $A$ and ML for $h$ and $R_{VV}$. We call this method GMAPML. The quantity to be maximized is:

$$f(Y, A/h, R_{VV}) = \frac{1}{2 \pi n V} \exp\left(-\frac{1}{2} (Y - T^A) R^{-1} (Y - T^A)\right)$$

with:

$$R = \begin{bmatrix} R_{VV} & 0 \\ 0 & R_{AA} \end{bmatrix}, \quad Y' = \begin{bmatrix} Y^H \\ A^o^H \end{bmatrix}^H$$

and $T' = \begin{bmatrix} T^H & I^H \end{bmatrix}^H$. The corresponding log-likelihood function to be minimized is:

$$\ln(\det R) + (Y' - T^A)^H R^{-1} (Y' - T^A)$$

3.2. Relationship between the two Approaches.

The minimization of (6), being separable in $A$ and $h$, is usually done by minimizing w.r.t. $A$ first:

$$A = (T^H R_{VV}^{-1} T + R_{AA}^{-1})^{-1} (T^H R_{VV}^{-1} Y + R_{AA}^{-1} A^o)$$

which is of the form of a MMSE non-causal decision-feedback equalizer (NCDFE), being a MMSE linear estimator in terms of the received data $Y$ and the uncertain preliminary symbols $A^o$ with uncertainty reflected in $R_{AA}$. Note that when $A^o = 0$ (blind case), the estimate of the symbols corresponds to the output of a MMSE linear equalizer. This illustrates the superiority of the GMAPML method over the DML method, for which the symbol estimates are given by the output of a MMSE ZF linear equalizer.

Substituting (7) in (6) results in the minimization of:

$$\ln(\det R) + (Y - T A^o)^H R_{VV}^{-1} (Y - T A^o)$$

If we suppose $R$ known, GMAPML appears to be a simplified version of GML from which the term $\ln(\det R_{VV})$ has been removed.

3.3. Semi-Blind Estimation

We specialize now this general formulation to the case of semi-blind equalization in which we use a partial training sequence and a Gaussian prior for the remaining unknown symbols. Let $A = P \begin{bmatrix} A_1^o & A_2^o \end{bmatrix}^H$ where $A_1^o$ are the training symbols and $P$ is a permutation matrix to account for the fact that the training sequence does not necessarily occur at the beginning. Hence we have:

$$A^o = P \begin{bmatrix} A_1^o \\ 0 \end{bmatrix}, \quad R_{AA} = \sigma_a^2 P \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} P^H$$

where we shall let $\epsilon \gg 0$. Also, as mentioned earlier, we take $R_{VV} = \sigma_v^2 I$.

3.4. Cramer-Rao Bounds

GML estimates reach asymptotically their Cramer-Rao Bound (CRB). In GMAPML, however, the input symbols estimates cannot be consistent because their number is asymptotically infinite. Hence, joint estimates of $A$ and $h$ cannot be consistent and do not reach their CRB.

A comparison in terms of CRB between both methods is not meaningful. GML appears however to give better performance. Whereas DML cannot estimate the module of the channel, the introduction of the Gaussian assumption allows this estimation in GML but not in GMAPML. For a channel of length $N = 1$, GMAPML gives an estimate of the normalized channel in the same direction as GML, but with a infinite module. We will now consider the complex CRBs for GML and GMAPML.

3.4.1. GML

The Fisher Information Matrix (FIM) for deterministic complex parameters $\theta$ is defined as:

$$J(\theta) = -E_{Y,\theta} \frac{\partial \ln f(Y/\theta)}{\partial \theta^*}$$

$J(\theta)^{-1}$ is the complex CRB: $C_{\theta \theta} \geq J(\theta)^{-1}$, where $C_{\theta \theta} = E_{\theta} \theta^* \theta^*, \quad \theta = \theta - \theta^*, \quad E\theta = 0$. The parameters are $\theta = [h^H \sigma_v^2]^H$.

$$J(\theta)(i,j) = (A^o^H R_{VV}^{-1} A)^{-1} + \text{trace}\left(R_{VV}^{-1} \frac{\partial R_{VV}}{\partial \theta_i} R_{VV}^{-1} \frac{\partial R_{VV}}{\partial \theta_j}\right),$$

$\mathcal{T}(H) A^o = A^o h$ is a structured matrix filled with the elements of $A^o$.

$$\frac{\partial R_{VV}}{\partial \theta_i} = \mathcal{T}(H) R_{AA} \mathcal{T}(H) \left(\frac{\partial H^*}{\partial \theta_i}\right) \quad \frac{\partial R_{VV}}{\partial \theta_i} = \sigma_v^2 I.$$ 

Note that these expressions apply to the general case of sections 3.1 and 3.2. The real FIM corresponding to the estimation of the real and imaginary part of the channel is singular: indeed the phase of the channel cannot be estimated. The complex FIM is not singular: the complex CRB gives a lower bound for a channel estimate for which the correct phase factor is known.

3.4.2. GMAPML

The estimate parameters are $\theta = [A^H h^H \sigma_a^2]^H$.

$$J(\theta) = -E_{Y,\theta} \frac{\partial \ln f(Y/A, \sigma_a^2)}{\partial \theta^*}$$

Applied to our problem where $A$ is stochastic and $h$, $\sigma_a^2$ deterministic, the FIM is:

$$J(\theta) = -E_{Y,\theta} \frac{\partial \ln f(Y/A, \sigma_a^2)}{\partial \theta^*}$$

The CRB for $h$ is:

$$\sigma_v^2 \left[M \sigma_a^2 I - A^o^H T \left(T_2^H T_2 + \frac{\sigma_a^2}{\sigma_v^2} I\right) \left(T_2^H A^o\right)^{-1}\right]^{-1},$$

where $T A = T_1 A_1^o + T_2 A_2$. In the blind case, the CRB is $\frac{\sigma_v^2}{M \sigma_a^2} I$. It corresponds asymptotically to the CRB when all input symbols are known. This CRB for GMAPML appears to be much too optimistic.

3.4.3. Simulations

We present some simulations to illustrate the effect of the number of known symbols on the CRB for GML. We plot the trace of the CRB normalized w.r.t. the channel module and the number of channel coefficients. The SNR, defined as $\frac{\|\mathcal{H}\|_{\Sigma_a}^2}{M \sigma_a^2}$ is of 10dB. $M=50$. The channel is:

$$\begin{bmatrix} 1.0000 & 0.8000 \\ -1.5000 & 1.4000 \end{bmatrix}.$$ 

We notice the significant improvement of semi-blind equalization as more and more symbols are known, especially with few known symbols. Furthermore, the estimation of $\sigma_a^2$, which is also done in GML, does not influence much the performance of channel estimation itself.
The principle of this method has been applied to sinusoids in noise estimation [7] as well as to channel-based DML in equalization [8]. In this last method, the gradient of the cost function $C(h)$ to be minimized may be written as $B(h)$, where $B(h)$ is positive semi-definite. In a first step, $B(h)$ is considered as constant: we have the equivalent to a quadratic problem, and the solution is the minimal eigenvector of $B(h)$, this solution is used to reevaluate $B(h)$ and other iterations may be done.

For blind GML, the decomposition is of the form $B(h)h$, for semi-blind GML, it is of the form $B(h)h + D(h)$, where $B(h)$ is now positive definite. For the semi-blind GML, both $B(h)$ and $D(h)$ are supposed constant in the first step of the procedure, $h$ is the solution of a linear system. The difficulty is to find the right decomposition.

For a channel of length 1, we derived the PQML method, the general case is still under study. In fig.1, we show the normalized errors on $h$, at each step of PQML, for 20 known symbols and SNR=10dB. The channel considered is $\begin{bmatrix} 1 & -1.5 \end{bmatrix}^T$.

### 3.5.2. Approximate semi-blind DML

The following one-shot method can be used in its own right or as initialization step for PQGML. For conciseness, we shall explain the case of $m=2$ channels. Several popular methods are based on the observation that for the noise-free signals, we have $H_2(z)y_1(k) - H_1(z)y_2(k) = 0$ or in matrix form $T(H_2)Y_2 = T(H_1)Y_1 = Y_2h = 0$ where superscript $(2)$ denotes the data used for the blind part.

In the presence of noise, we can solve this equation in a least-squares sense for $h$. Let $Y^{(1)}$ denote the data corresponding to the training symbols $A^c$. If $Y^{(1)}$ and $Y^{(2)}$ do not overlap, then an optimally weighted semi-blind least-squares criterion is

$$\min_h \left\| \begin{bmatrix} Y^{(1)} \\ 0 \end{bmatrix} - A^c Y^{(2)} \right\|^2_{W^{-1}}$$

where in $W = T(H_1)T^H(H_1) + T(H_2)T^H(H_2)$ the two terms can be estimated consistently from $\frac{1}{N} \left( \hat{R}_{YY} - \lambda_{\text{min}}(\hat{R}_{YY})I \right)$.

### 4. PREDICTION BASED STOCHASTIC ML

The GMAPML minimization criterion can be parametrized in terms of linear prediction quantities, which has for advantage to be robust to channel order overestimation. The extended noise subspace (corresponding to $T^\perp$) can be linearly parametrized. Let $T^\perp(G)$, parameterized by $G$, be of full column rank equal to the noise subspace dimension such that $T^\perp H^* T^\perp = 0$.

The signal subspace is parameterized linearly by $H$. Now let $P(z) = \sum_{l=0}^{\infty} p(l) z^{-l}$ with $p(0) = 1_m$ be the MMSE multivariate prediction error filter of order $L$ for the noise-free received signal $y(k)$. If $L \geq 2L = \left\lceil \frac{N}{2\lambda_{\text{min}}} \right\rceil$, then it can be shown [9] that

$$P(z)H(z) = h(0).$$

From (18) it is clear that $H(z)$ and $P(z)h(0)$ are equivalent parameterizations.

For the noise subspace parameterization, note that from (18) we get

$$T_{M-L}(P)H(z) = \left[ T_{M-L}(P) - I \otimes h(0) \right] T(\hat{H}) = 0 \quad (19)$$

However

$$\text{Range} \left( \left[ T_{M-L}(P) - I \otimes h(0) \right] \right)^\perp \subset \left( \text{Range} \left( T^\perp \right) \right)^\perp$$

(20)

which is the noise subspace is not completely spanned so that replacing $T^\perp H$ by $\left[ T_{M-L}(P) - I \otimes h(0) \right]$ in (17) (with the role of $G$ played by $p(1:L), h(0)$) leads to estimates that are only asymptotically (in $M$) equal to the ML estimates.

We shall specialize now this general formulation to the case of semi-blind equalization. (17) now becomes:

$$\min_{p(1:L), h(0)} \left\| T(P)Y - (I \otimes A^c) \otimes h(0) \right\|_{\mathbb{R}^N}$$

(21)

where

$$\mathcal{R} = \sigma^2 \left( T(P)T^H(P) + (I \otimes R \otimes I^H) \right) h(0)$$

This equation can be seen as the (maximized w.r.t. $A$) GMAPML log likelihood of $T(P)Y \sim \mathcal{N}((I \otimes A^c) \otimes h(0), \mathcal{R})$. $T(P)Y$ represents a reduction of data w.r.t. $Y$ which is however asymptotically negligible.

We can then define 2 approaches to ML based on $P$ and $h(0)$: the GMAPML approach described in (21) and the GML approach which minimizes:

$$\min_{\text{det}{\mathcal{R}}} \left\| T(P)Y - (I \otimes A^c) \otimes h(0) \right\|_{\mathbb{R}^N}$$

(22)

Like for the channel based approaches, the GML method appears better than the GMAPML.
IQML is not guaranteed to converge. However, if it is initialized by a consistent estimate for $P$ and $h(0)$, it converges sufficiently at the first iteration.

We simulated this algorithm for $M=100$, $SNR=7dB$ and $10$ dB, and the following channel:

$P(1:1,10,\ldots,0) \begin{bmatrix} 1.0000 & 0.8000 \\ -1.5000 & 1.4000 \end{bmatrix}$

The number of known symbols is successively: 20, 10, 5 and 0. Two initializations were used: initialization by training sequence when 20 symbols are known and initialization by a blind algorithm based on linear prediction [6] tested for all cases. For 10 and 5 known symbols, we either kept the $h(0)$ of the initialization or reevaluated it at each iteration. When no symbols are known, we kept the $h(0)$ of the initialization. We get back the channel from $P$ and $h(0)$ by solving equation (18) in the least squares sense.

We notice that even with 5 known symbols, the estimation of $h(0)$ gets improved by the algorithm. It appears to be preferable to do only one iteration of the IQML. Furthermore, the performance is good when 20 symbols are known.

The estimate of $h(0)$ can be refined by a PQLM method applied to the GML cost function.

After the submission of this paper, we became aware of [10] in which a semiblind GML method and corresponding Cramer-Rao Bound have also been presented. The modeling of the training-sequence information in [10] is incorrect through (instead of the training sequence, the information considered in [10] is the training sequence times an unknown zero-mean unit-variance normal variable).

REFERENCES


MAXIMUM-LIKELIHOOD BLIND FIR MULTI-CHANNEL ESTIMATION WITH GAUSSIAN PRIOR FOR THE SYMBOLS

Elisabeth De Carvalho and Dirk T.M. Stock

Institut EURECOM
B.P. 193, 06904 Sophia Antipolis Cedex, FRANCE
{carvalho,stock}@eurecom.fr

We present two approaches to stochastic Maximum Likelihood identification of multiple FIR channels, where the input symbols are assumed Gaussian and the channel deterministic. These methods allow semi-blind identification, as they accommodate a priori knowledge in the form of a (short) training sequence and appears to be more relevant in practice than purely blind techniques. The two approaches are parameterized both in terms of channel coefficients and in terms of prediction filter coefficients. Corresponding methods are presented and some are simulated. Furthermore, Cramer-Rao Bounds for semi-blind ML are presented: a significant improvement of the performance for a moderate number of known symbols can be noticed.