A FAST GAUSSIAN MAXIMUM-LIKELIHOOD METHOD FOR BLIND MULTICHANNEL ESTIMATION

Elisabeth de Carvalho   and   Dirk T.M. Slock

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

ABSTRACT

We propose a blind Maximum-Likelihood method for FIR multichannel estimation, denoted GML. The GML criterion is derived assuming the input symbols as Gaussian random variables. The performance of GML (computed based on the true symbol distribution) is compared through numerical evaluations to the optimally weighted covariance matching method: both methods are equivalent in a certain asymptotic sense. A fast implementation of the scoring algorithm is proposed to solve GML.

1. PROBLEM FORMULATION

We consider a single-user multichannel model: this model results from the oversampling of the received signal and/or from reception by multiple antennas. Consider a sequence of symbols $a(k)$ received through m channels of length N and coefficients $\boldsymbol{h}(i)$:

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

$\boldsymbol{v}(k)$ is an additive independent white Gaussian circular noise with $r_{\boldsymbol{vv}}(k-i) = \mathrm{E}\,\boldsymbol{v}(k)\boldsymbol{v}(i)^H = \sigma_v^2 I_m \delta_{ki}$. Assume we receive M samples, concatenated in the vector $\boldsymbol{Y}_M(k)$:

$$\boldsymbol{Y}_M(k) = \mathcal{T}_M(h)\,A_{M+N-1}(k) + \boldsymbol{V}_M(k) \tag{2}$$

$\boldsymbol{Y}_M(k) = [\boldsymbol{y}^H(k-M+1) \cdots \boldsymbol{y}^H(k)]^H$, similarly for $\boldsymbol{V}_M(k)$, and $A_M(k) = \left[a^H(k-M-N+2) \cdots a^H(k)\right]^H$, where $(.)^H$ denotes Hermitian transpose. $\mathcal{T}_M(h)$ is a block Toeplitz matrix filled out with the channel coefficients grouped in the vector h. We shall simplify the notation in (2) with $k = M-1$ to:

$$\boldsymbol{Y} = \mathcal{T}(h)\,A + \boldsymbol{V} \tag{3}$$

2. GAUSSIAN MAXIMUM LIKELIHOOD (GML)

Gaussian ML considers the estimation of the parameter $\theta = \left[h^H\ \sigma_v^2\right]^H$. The GML criterion is derived assuming the input symbols as Gaussian i.i.d. random variables of variance σ_a^2. This Gaussian hypothesis is used only to build the GML criterion: a performance analysis shows that GML has a meaning and is, from a performance point of view, the most powerful method among all the blind methods using the second–order statistics of the data.

$$A \sim \mathcal{N}(0, \sigma_a^2 I) \Rightarrow \boldsymbol{Y} \sim \mathcal{N}(0, C_{YY}(\theta)), \tag{4}$$

with $C_{YY}(\theta) = \sigma_a^2 \mathcal{T}(h)\mathcal{T}^H(h) + \sigma_v^2 I$. The Gaussian Maximum Likelihood criterion is then :

$$\min_{\theta=(h,\sigma_v^2)} \left\{ \ln(\det C_{YY}(\theta)) + \boldsymbol{Y}^H C_{YY}^{-1}(\theta)\boldsymbol{Y} \right\} \tag{5}$$

In this paper, we will treat only the single user case. However, one of the reasons for examining GML is its extension to the multiuser case where GML is of particular interest. Apart, from performance advantages, one of the great properties of the GML, like all methods using the second–order statistics of the data, is their robustness to channel length overestimation. It is well known, both for the single user case as for the multiuser case, that deterministic methods fail when the channel length has been overestimated: each channel length for each users has to be tested. In the multiuser case, where the different channels have, in general, different lengths, this drawback nearly condemn deterministic methods to be purely theoretic. On the contrary, the Gaussian approach can be shown not to suffer from this problem (as has been shown for Linear Prediction methods [1]). In multiuser communications, GML has also another advantage: deterministic methods can only identify the channel apart from a triangular dynamical multiplicative factor, whereas Gaussian methods can identify the channel up to a unitary static factor.

In fact, in the multiuser case, these identifiability and robustness properties would be sufficient to tell us that the second–order approaches would be the only viable for blind second–order statistics based methods.

3. COMPARISON WITH COVARIANCE MATCHING METHOD

We compare, through simulations, GML to the Optimally weighted Covariance Matching (OCM) method [2] which has been said to be the most powerful method based on the second–order moments of the data. We will see that this result is true when the length of the correlation sequence considered is infinite. The GML criterion can also be written as:

$$\min_{\theta=(h,\sigma_v^2)} \left\{ \ln(\det C_{YY}(\theta)) + \mathrm{tr}\left\{ C_{YY}^{-1}(\theta)\boldsymbol{Y}\boldsymbol{Y}^H \right\} \right\} \Leftrightarrow$$
$$\min_{\theta=(h,\sigma_v^2)} \left\{ \ln(\det C_{YY}(\theta)) + \mathrm{tr}\left\{ C_{YY}^{-1}(\theta)\widehat{C}_{YY}(\theta) \right\} \right\}$$
(6)

where $\widehat{C}_{YY}(\theta) = \boldsymbol{Y}\boldsymbol{Y}^H$. The fact that $\boldsymbol{Y}\boldsymbol{Y}^H$ can be considered in the criterion as an estimate of $C_{YY}(\theta)$ is justified because, asymptotically in the number of data, GML criterion behaves like its expected value:

$$\min_{\theta=(h,\sigma_v^2)} \left\{ \ln(\det C_{YY}(\theta)) + \mathrm{tr}\left\{ C_{YY}^{-1}(\theta)E\left(\boldsymbol{Y}\boldsymbol{Y}^H\right) \right\} \right\}$$
(7)

Equation (6) looks then like a criterion matching $C_{YY}(\theta)$ to one estimate of it. And then can be seen as a form of covariance matching.

The CM method proceeds to a weighted least-squares fit between the model of the covariance matrix of the received signal:

$$R_L(\theta) = \sigma_a^2 \mathcal{T}_L(h)\mathcal{T}_L^H(h) + \sigma_v^2 I \tag{8}$$

and its sample estimate:

$$\widehat{R}_L = \frac{1}{M}\sum_{k=0}^{M} \boldsymbol{Y}_L(k)\boldsymbol{Y}_L^H(k) \tag{9}$$

where L is the length of the covariance matrix chosen. Note that this formulation of the CM method is in fact asymptotically exact only: indeed, you have $M - i$ samples to estimate the moment of order i and not M as indicated. We prefer this formulation which allows to get closed form expressions of the optimal weighting matrix, and then of the performance.

Let $r(\theta)$ containing the non-redundant elements of $R_L(\theta)$, *i.e.* the first block column and block row, and similarly for $\hat{r}$. The CM criterion writes as:

$$\min_{h,\sigma_v^2} (\hat{r} - r(\theta))^H \mathcal{W} (\hat{r} - r(\theta)) \tag{10}$$

where $\mathcal{W}$ is a weighting matrix. The optimal weighting matrix is:

$$\mathcal{W}^o = \left(\mathrm{E}\left[\hat{r} - r(\theta^o)\right]\left[\hat{r} - r(\theta^o)\right]^H \right)^{-1} \tag{11}$$

θ^o is the true parameter value. When replacing θ^o by a consistent estimate, the performance remains the same.
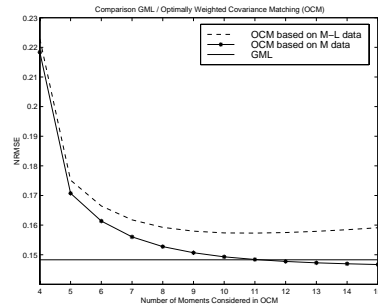

Figure 1: Comparison between GML and the optimally weighted covariance matching method w.r.t. the number of correlation coefficients considered in the OCM method.

Which elements should be considered in $\tilde{r}(\theta) = \hat{r} - r(\theta)$? The authors of [2, 4] consider only the (non-redundant) *non zero* coefficients and claim that they are sufficient to get the optimal performance. This is not true however as stated in [3]. The optimal performance are obtained when the number of correlation coefficients involved tends to ∞. The asymptotical ($M \to \infty$, $L \to \infty$, $L \ll M$) OCM corresponds then to the best method exploiting the second order moments of the data: we will see, by simulations, that this optimal performance corresponds also to the performance of GML. Note that the non weighted CM method only require the N non-null correlation coefficients. The performance of OCM is given by:

$$\mathrm{E}(\theta_R - \hat{\theta}_R)(\theta_R - \hat{\theta}_R)^H = \left(\frac{\partial r^H(\theta)}{\partial \theta_R}\right)\mathcal{W}^{o-1}\left(\frac{\partial r^H(\theta)}{\partial \theta_R}\right)^H \tag{12}$$

where $\theta_R = [\mathrm{Re}^H(\theta) \quad \mathrm{Im}^H(\theta) \quad \sigma_v^2]^H$ and $\hat{\theta}_R$ is the corresponding CM estimate. The expression of $\mathcal{W}^o$ is not given here for lack of space. In figure 3, we show the performance of GML and OCM for the channel only (σ_v^2 is assumed known), *i.e.* $\|h_R - \hat{h}_R\|/\|h_R\|$, when the sample matrix is based on M and $M - L$ data samples (the true performance of OCM is in fact between the two curves, the burst length being of 100, we have not reach completely asymptotic conditions, which explains why the curves are distinct). The channel is of length 4. In this figure, it can be noticed that the performance of OCM get better as more and more correlation coefficients are included. A quasi steady–state is rapidly attained, but considering only the N first moments is definitively not enough.

4. METHOD OF SCORING: FAST IMPLEMENTATION

We propose to solve the GML criterion by the method of scoring. The method of scoring consists in an approximation of the Newton-Raphson algorithm which finds an esti-

mate $\theta^{(i)}$ at iteration i from $\theta^{(i-1)}$, the estimate at iteration $i-1$, as:

$$\theta^{(i)} = \theta^{(i-1)} - \left[\frac{\partial}{\partial \theta^*} \left(\frac{\partial c(\theta)}{\partial \theta^*} \right)^H \Bigg|_{\theta^{(i-1)}} \right]^{-1} \frac{\partial c(\theta)}{\partial \theta^*} \Bigg|_{\theta^{(i-1)}} \tag{13}$$

where $c(\theta)$ is the cost function and θ contains the parameters to estimate. The method of scoring approximates the Hessian by its expected value, which is here the Gaussian Fisher Information Matrix (FIM). This approximation is justified by the law of large numbers as the number of data is generally large.

4.1. Approximated Scoring Method

We detail here only the case where the channel is complex. To simplify things, we consider furthermore that the noise variance is known. The problem parametrized in $h_R = \left[\text{Re}^H(h) \ \text{Im}^H(h) \right]^H$ can be equivalently parametrized in $h_C = \left[h^H \ h^{*H} \right]^H$. Let

$$J_{\phi\xi} = -\text{E}\left(\frac{\partial}{\partial \phi^*} \left(\frac{\partial c(h)}{\partial \xi^*} \right)^H \right) \tag{14}$$

The FIM for h_C is then:

$$J_{h_C h_C} = \begin{bmatrix} J_{hh} & J_{hh^*} \\ J_{hh^*}^* & J_{hh}^* \end{bmatrix} \tag{15}$$

The coefficient (i, j) of J_{hh} and J_{hh^*} is:

$$J_{hh}(i, j) = \text{tr}\left\{ C_{YY}^{-1} \frac{\partial C_{YY}}{\partial h_i^*} C_{YY}^{-1} \left(\frac{\partial C_{YY}}{\partial h_j^*} \right)^H \right\} \tag{16}$$

$$J_{hh^*}(i, j) = \text{tr}\left\{ C_{YY}^{-1} \frac{\partial C_{YY}}{\partial h_i^*} C_{YY}^{-1} \frac{\partial C_{YY}}{\partial h_j^*} \right\} \tag{17}$$

Our fast implementation of the method of scoring is based on a frequential asymptotical (in the number of data) approximation of the FIM and of the gradient of the cost function. Let's consider first the term J_{hh}: it can be asymptotically approximated as [5]:

$$J_{hh}(i, j) = \frac{M}{4\pi} \int_{-\pi}^{\pi} \text{tr}\left\{ S_{yy}^{-1} \frac{\partial S_{yy}}{\partial h_i^*} S_{yy}^{-1} \left(\frac{\partial S_{yy}}{\partial h_j^*} \right)^H \right\} d\omega \tag{18}$$

where $S_{yy} = S_{yy}(\omega) = \mathbf{h}(e^{-i\omega})\mathbf{h}^H(e^{i\omega}) + \sigma_v^2 I$ is the spectral density of the received signal. From this expression, we see that $J_{hh}(i, j)$ can be approximated as a block Toeplitz matrix (which is also symmetric). The block $(1, j_b)$ of its first line is the coefficient of order $1 - j_b$ of the filter:

$$\frac{\mathbf{h}^\dagger(z)\mathbf{h}(z) \left[\left(\mathbf{h}^\dagger(z)\mathbf{h}(z) + \sigma_v^2 \right) I - \mathbf{h}(z)\mathbf{h}^\dagger(z) \right]}{\sigma_v^2 \left(\mathbf{h}^\dagger(z)\mathbf{h}(z) + \sigma_v^2 \right)^2} \tag{19}$$

The numerator implies only FIR filtering operations. Using the Gohberg-Semencul formula: $\left(\mathbf{h}^\dagger(z)\mathbf{h}(z) + \sigma_v^2 \right)^{-2} = p(z)p^\dagger(z)/\tilde{\sigma}^2$, where $p(z)$ is the linear prediction filter associated to $\left(\mathbf{h}^\dagger(z)\mathbf{h}(z) + \sigma_v^2 \right)^2$, and $\tilde{\sigma}^2$, the prediction error. The coefficients of the filter (19) are then computed by filtering the numerator by a troncated version of $p(z)$ and $p^\dagger(z)$ (a troncation αN, where α is 3 or 4 is in general sufficient). So computing the elements of J_{hh} is of order N.

The same kind of treatement holds for J_{hh^*} which can be approximated as a block Hankel matrix. The block $(1, j_b)$ of its first line is the coefficient of order $j_b - 1$ of the filter:

$$\frac{\mathbf{h}(z)\mathbf{h}^T(z)}{\left(\mathbf{h}^\dagger(z)\mathbf{h}(z) + \sigma_v^2 \right)^2} \tag{20}$$

The block $(i_b, 1)$ of the last column is its coefficient of order $N - 2 + i_b$.

Let D_h be the gradient w.r.t. h^* of the cost function.

$$D_h(i) = \text{tr}\left\{ C_{YY}^{-1} \frac{\partial C_{YY}}{\partial h_i^*} \right\} - \mathbf{Y}^H C_{YY}^{-1} \frac{\partial C_{YY}}{\partial h_i^*} C_{YY}^{-1} \mathbf{Y} \tag{21}$$

Using the band property of C_{YY}, a fast computation of the output of $C_{YY}^{-1}\mathbf{Y}$ is of order NM. $\frac{\partial C_{YY}}{\partial h_i^*} = \mathcal{T}(h)\mathcal{T}^H\left(\frac{\partial h}{\partial h_i} \right)$ and both terms being banded, the computation of the second term is of order MN. Using a frequential approximation, the block i_b of the first term of D_h can be approximated as the element $i_b - 1$ of the filter:

$$\frac{\mathbf{h}(z)}{\mathbf{h}^\dagger(z)\mathbf{h}(z) + \sigma_v^2} \tag{22}$$

D_{h^*} can be computed using $D_{h^*} = (D_h)^*$. At each step of the algorithm, equation (13) is solved using the Toeplitz and Hankel property of J_{hh} and J_{hh^*}, which gives a complexity of order N^2.

4.2. Regularization of the FIM

GML can estimate the channel up to a phase factor only, which results in a singularity of the FIM, spanned by $h_s = \left[-\text{Im}^H(h^o) \ \text{Re}^H(h^o) \right]^H$ (where h^o is the true channel value). The scoring algorithm as described in equation (13) cannot be directly applied. In order to regularize the estimation problem a constraint has to be considered to adjust the phase of the channel. The constraint considered here is $h_s^H h_R = 0$ (and $h^{o\,H} h > 0$ to determine the right sign).

The initialization is given by the Schur algorithm developed in [6]: we chose this method because this is a method based on the second–order moments of the data which can be used as initialization in the multi–user case also. The previous constraint is used to estimate the phase factor of the channel estimate given by the Schur method. This method gives a consistent estimate of the channel.
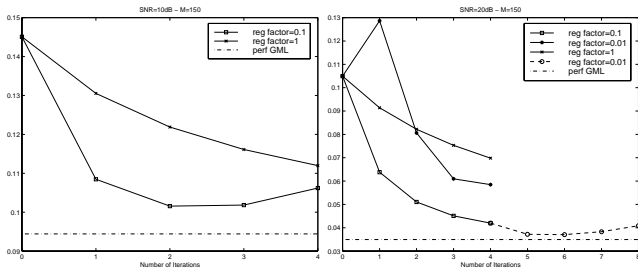
Figure 2: Approximated scoring algorithm for 10dB and 20dB for different regularization factors.

Figure 3: Comparison between the non–approximated and the approximated scoring algorithm for 10dB and 20dB.

In the scoring algorithm, we take the Moore-Penrose pseudo–inverse of the FIM: it corresponds to the constraint $\hat{h}_R^{(i-1)\,H}\hat{h}_R^{(i)} = 0$, where $\hat{h}^{(i)}$ is the channel estimate at iteration i. When the algorithm converges correctly, this constraint is equivalent to the previously mentioned constraint $h_s^H h_R = 0$.

The approximated FIM is nonsingular: it has an eigenvalue (negative or positive) closed to 0. The inverse of the approximated FIM could then be directly taken in the scoring algorithm: this solution makes the algorithm diverge however, as the step in the direction of the associated eigenvector is too large. The best solution here is to regularize the FIM by a certain λI. In our different tries, it appears that the λ should be quite large: we tested $\lambda = \lambda_{max}, 0.1\lambda_{max}, 0.01\lambda_{max}$ (and also $\lambda = 0$, *i.e.* no regularization), where λ_{max} is the maximal eigenvalue of the approximated FIM. The best solution was found to be $0.1\lambda_{max}$: $0.01\lambda_{max}$ and 0 made the algorithm diverge; λ_{max} results in a slow convergence with a steady state worse than with $0.1\lambda_{max}$. When the channel is real, the FIM is regular: in this case no regularization of the approximated FIM is necessary in the scoring algorithm.

4.3. Simulations

The simulation in figure 2 illustrates this fact: we plot the averaged normalized errors ($\|h - h_e\|/\|h\|$) over 50 noise and input symbol realizations for a randomly chosen channel ($N = 4$, $m = 2$). At 10dB, only the regularization factors $0.1\lambda_{max}$ and λ_{max} work: the first choice gives performance closed to the theoretical performance of GML. At 20dB, the regularization factor $0.01\lambda_{max}$ works more or less, but the best regularization is still $0.1\lambda_{max}$. We also tested the scoring algorithm with regularization $0.01\lambda_{max}$, but initialized by the algorithm with regularization $0.1\lambda_{max}$: a small improvement can be noticed only. At last, in figure 3, we compare the non-approximated to the approximated scoring algorithm (with regularization factor $0.1\lambda_{max}$). Performance of the approximated scoring is quite closed to the true scoring especially at 20dB.
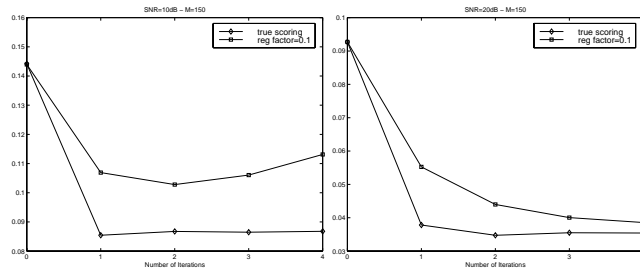
5. CONCLUSION

We have developped a fast implementation of the scoring algorithm to solve GML. GML was compared to the optimally weighted covariance matching method, which was shown, through simulations, to have the same performance asymptotically (in the number of data but also in the number of moments considered). The fast GML should be next generalized to the multi–user case as it provides a low–computational solution with potentially the best performance among all the methods exploiting the second–order moments of the data.

6. REFERENCES

[1] Alexei Gorokhov. *Séparation autodidacte des mélanges convolutifs: méthodes du second ordre*. PhD thesis, Ecole Nationale Supérieure des Télécommunications, 1997.

[2] G.B. Giannakis and S.D. Halford. Asymptotically Optimal Blind Fractionally-Spaced Channel Estimation and Performance Analysis. *IEEE Transactions on Signal Processing*, 45(7):1815–1830, July 1997.

[3] M. Kristensson. *On Parameter Estimation in Wireless Communications, Sensor Array Processing and Spectral Analysis*. PhD thesis, Royal Institute of Technology (KTH), Sweden, 1998.

[4] H.H. Zeng and L. Tong. Blind Channel Estimation Using the Second–Order Statistics: Asymptotic Performance and Limitations. *IEEE Transactions on Signal Processing*, 45(8):2060–2071, Aug. 1997.

[5] P. Whittle. The Analysis of Multiple Stationary Time Series. *J. Royal Statistic. Soc.*, 45:125–139, 1953.

[6] Luc Deneire and Dirk Slock. A Schur Method for Multiuser Multichannel Blind Identification. In *Proc. ICASSP 98 Conf.*, Phoenix, USA, March 1999.