ARMA MODEL IDENTIFICATION USING HIGHER ORDER STATISTICS AND FISHER INFORMATION CONCEPTS

Eric LE CARPENTIER and Jean-Luc VUATTOUX
Laboratoire d'Automatique de Nantes, URA C.N.R.S. 823,
Ecole Centrale de Nantes/Université de Nantes,
1 rue de la Noë, 44072 Nantes cedex 03, France
Tel: (33) 40 37 16 46; fax: (33) 40 37 25 22
e-mail: lecarpentier@lan.ec-nantes.fr

ABSTRACT

The problem of estimating the parameters of a non causal ARMA system, driven by an unknown input noise with unknown symmetrical probability density function (PDF) is addressed. A maximum likelihood approach is proposed in this paper. The main idea of our approach is that the assumed PDF of the input noise is the PDF minimizing the Fisher information among PDFs matching the estimated cumulants of $2nd$ and $4th$ order. This minimization problem is hard to solve, so we use an over-parameterized PDF model, which is a gaussian mixture. We obtain two different models for the classes of sub-Gaussian and super-Gaussian PDFs. For this latter class, we get the most robust estimator in Huber's sense, among these generated by this class. A new parameter estimation method is given and its robustness and optimality properties are detailed. The performances of the resulting identification scheme are compared to those of another higher order method.

1 INTRODUCTION

Stochastic modeling using non-Gaussian noise driven ARMA models has attracted considerable attention in the last years. The fact that the output of these systems carries phase information, has introduced numerous methods based on higher order statistics (HOS), in form of moments or cumulants, in Signal Processing and System Identification [3], [1], [9]. The main disadvantage of these methods is that they do not provide any information about the theoretical performances of the estimator and its optimality in the sense of the covariance matrix of the estimated parameters. finite set consequence that

For optimum parameter estimation, exact knowledge of the probability density functions (PDF) is necessary (maximum likelihood (ML) methods). Otherwise, we can assume a certain class of PDFs for the input and obtain the optimality in the minimax sense by using a ML approach with the PDF which minimizes the Fisher information (FI) in this class and provides the most robust (in Huber's sense) [4, 7] parameter estimates.

In connection with higher order statistics, we consider the class of cumulants constrained PDFs and determine the PDF which minimizes the FI under cumulant constraints. This problem was partially solved in [10] but the results are limited to the class of symmetrical sub-gaussian PDFs (negative kurtosis). So this problem is always open for the class of super-Gaussian (SPG) PDFs (positive kurtosis).

In this paper, we propose a new parameter estimation method based on the prediction error method (PEM) using cumulants of second and fourth order and the minimization of the FI.

In order to cover the class of heavy-tailed, asymptotically SPG noise distributions, which represent a very important class in identification problems, we have used a model of PDF, that is appropriate for a non-Gaussian process with heavy tails, which is a mixture of two Gaussian distributions:

$$f_M(u) = p\Phi_1(u) + (1-p)\Phi_2(u), \quad 0 \le p \le 1 \tag{1}$$

where $\Phi_1(u)$ and $\Phi_2(u)$ are Gaussian PDFs.

We have considered only the case of symmetrical PDF constrained by the second and fourth order cumulants, used in practice. In section 2, the proposed parameter estimation scheme is presented. In section 3, the gaussian mixture (GM) PDF parameterized by second and fourth order cumulants is given. The estimation algorithm is presented in section 4. The section 5 analyse the asymptotic behaviour and robustness properties of our method. Simulation results are summarized in section 6. In section 7, a conclusion is given.

2 NON CAUSAL ARMA PARAMETERS ESTIMATION

Let the observed process $\{y_t\}$ be modeled as the output of a discrete stable linear shift-invariant system $H_{\theta_0}(z)$ with unobservable input $\{e_t\}$:

$$y_t = H_{\theta_0}(z)e_t \tag{2}$$

where

$$H_{\theta_0}(z) = \frac{A(z)C(z^{-1})}{B(z)D(z^{-1})} = \frac{(\sum_{i=0}^{n_A} a_i z^{-i})(\sum_{i=0}^{n_C} c_i z)}{(\sum_{i=0}^{n_B} b_i z^{-i})(\sum_{i=0}^{n_D} d_i z)} \tag{3}$$

with $a_0 = b_0 = c_0 = d_0 = 1$ and

$$\theta_0 = [a_1 \dots a_{n_A}\ b_1 \dots b_{n_B}\ c_1 \dots c_{n_C}\ d_1 \dots d_{n_D}]^T \quad (4)$$

We assume that all the roots of $A(z)$ and $B(z)$ are inside the unit-circle (causal minimum phase part) and all the roots of $C(z^{-1})$ and $D(z^{-1})$ are outside the unit-circle (anti-causal maximum phase part). And n_A, n_B, n_C, n_D denote the number of roots of $A(z)$, $B(z)$, $C(z^{-1})$ and $D(z^{-1})$, respectively. The input $\{e_t\}$ is a stationary, zero-mean, identically independent distributed (i.i.d.), non-Gaussian random sequence with unknown symmetrical PDF $f_e(u)$.

Given N consecutive samples of the system output $\{y_t\}$,$t = 1, \dots, N$, we want to estimate the actual parameter θ_0. The prediction error sequence $\{w_t(\theta)\}$ [5, 8] is related to the data through

$$w_t(\theta) = H_\theta^{-1}(z) y_t \quad (5)$$

With PEMs, the estimate $\widehat{\theta}$ of θ_0 is equal to θ which minimizes some criterion depending on the sequence of prediction errors

$$J(\theta) = \frac{1}{N} \sum_{i=1}^{N} \ell(w_t(\theta)) \quad (6)$$

where $\ell(.)$ is a scalar-valued norm, i. e.

$$\widehat{\theta} = \arg\min_\theta J(\theta) \quad (7)$$

If the PDF of the input noise $f_e(u)$ was known, the method which could give the best parameter estimate is the ML method, where the norm is $\ell(w) = -\log[f_e(w)]$. Since the PDF of the input is unknown, there are two possibilities: either to choose a norm giving satisfying results for a broad class of input PDFs (robustification), or to estimate $f_e(u)$ from the available data.

Here we propose an approach based on a GM model (1), noted $f_M^0(u)$, for the PDF of the input noise. This model is parameterized by its second and fourth order cumulants, C_2 and C_4 respectively, which are matching the corresponding cumulants of $f_e(u)$. The criterion to minimize is $J(\theta)$ with the norm $\ell(w) = \ell_0(w) = -\log[f_M^0(w)]$ (ML approach).

3 MODEL FOR THE INPUT PDF

Here we introduce the cumulant matched GM PDF as the model for the input distribution that we consider as symmetric. Despite the fact that this model does not result from any constrained mini- or maximization of PDF measure, it has a lot of very useful characteristics and interesting properties for the class of SPG PDFs.

The two Gaussian mixture's considered is given by (1) with

$$\Phi_i(u) = \frac{1}{\sqrt{2\pi V_i}} \mathrm{e}^{-\frac{1}{2}\frac{(u-m_i)^2}{V_i}} \quad (8)$$

where m_i and V_i are respectively the mean and variance of $\Phi_i(u)$.

So, let C_2 and C_4, respectively the variance and the kurtosis of any symmetrical PDF ($C_2 > 0, C_4 \geq -2C_2^2$). To obtain a centered symmetrical GM (1) with given variance C_2 and kurtosis C_4, we must have:

$$\begin{cases} pm_1 + (1-p)m_2 = 0 & (a) \\ p(V_1 + m_1^2) + (1-p)(V_2 + m_2^2) = C_2 & (b) \\ p(3V_1^2 + 6m_1^2 V_1 + m_1^4) + \\ \qquad (1-p)(3V_2^2 + 6m_2^2 V_2 + m_2^4) - 3C_2^2 = C_4 & (c) \end{cases} \quad (9)$$

We can see that if $p = 0$ or $p = 1$, the PDF $f_M(u)$ (1) is Gaussian with variance C_2 and it is possible only if $C_4 = 0$. So, if $C_4 \neq 0$ then $0 < p < 1$.

We remark immediately from (1) and (8) that there are two possibilities to obtain a symmetrical GM PDF: either $m_1 = m_2 = 0$ or

$$\begin{cases} m_1 = -m_2 = m > 0 \\ V_1 = V_2 = V > 0 \\ p = \frac{1}{2} \end{cases} \quad (10)$$

which corresponds to two different classes of PDFs. It ensues the following results:

Proposition 1: Let C_2 and C_4, to be respectively, the variance and the kurtosis of any PDF such that $-2 \leq C_4/(C_2)^2 \leq 0$ (sub-Gaussian PDFs class). There exists one and only one GM having these cumulants and it is obtained with

$$\begin{cases} m_1 = -m_2 = (\frac{-C_4}{2})^{\frac{1}{4}} \\ V_1 = V_2 = C_2 - \sqrt{\frac{-C_4}{2}} \\ p = \frac{1}{2} \end{cases} \quad (11)$$

This GM is our model f_M^0 for the sub-Gaussian PDFs class.

Proof: By using the parameterization of the case (10) and the equation (9b), we show that $m = \sqrt{C_2 - V}$ and $C_2 \geq V$. More, with the preceding equality and the equation (9c), we obtain that $C_4 = -2(C_2 - V)^2$. So, this model is valid only in the sub-Gaussian case ($-2 \leq C_4/(C_2)^2 \leq 0$). Then we deduce the Proposition 1 from the relations before.

Remark: If $C_4 = -2C_2^2$, our model corresponds to the Bernoulli distribution.

Proposition 2: Let C_2 and C_4, being respectively the variance and the kurtosis of any PDF such that $C_4 \geq 0$ (SPG PDFs class). Then there exists a set SGM_α of centered GM PDFs having these cumulants, described by

$$\begin{cases} m_1 = m_2 = 0 \\ V_1 = C_2 - \alpha\sqrt{\frac{C_4}{3}} \\ V_2 = C_2 + \frac{1}{\alpha}\sqrt{\frac{C_4}{3}} \\ p = \frac{1}{1+\alpha^2} \end{cases} \quad \forall \alpha \in]0, \sqrt{\frac{3C_2^2}{C_4}}] \quad (12)$$

Proof: If $m_1 = m_2 = 0$, we obtain from (9b) and (9c) that

$$\left\{ \begin{array}{ll} p = & \frac{C_2 - V_2}{V_1 - V_2} \\ C_4 = & -3(C_2 - V_1)(C_2 - V_2) \end{array} \right.$$

To have $0 < p < 1$, it is necessary and sufficient that $C_2 \in]V_1, V_2[$. So, it implies that $C_4 \geq 0$, which corresponds to the SPG PDFs class. We obtain a system of two equations with three unknowns. Then we introduce a degree of liberty with the real number $\alpha \in]0, \sqrt{\frac{3C_2^2}{C_4}}]$ such that we obtain the set of centered GM PDFs SGM_α, parameterized by α and having C_2 and C_4 as cumulants of order 2 and 4, of the Proposition 2.

Now, the problem is how to choose α, i. e. which model to take in SGM_α. So we decide to take the mixture model of SGM_α which minimizes the FI defined as

$$f_M^0 = \arg \min_{f_M \in SGM_\alpha} I_f \tag{13}$$

where I_f is the FI defined as

$$I_f = \int_{-\infty}^{\infty} \frac{(f')^2}{f} \, du \tag{14}$$

With the GM, this integral can be evaluated only numerically. So, we obtain the following results:

Proposition 3: The mixture PDF model of SGM_α minimizing I_f is obtained for $\alpha \longrightarrow 0$ and then $C_2 I_f \longrightarrow 1$.

It seems that the model of Proposition 2, due to the Proposition 3, is an ϵ-approximation of the solution of the FI minimization under constraints of C_2 and C_4 for the class of SPG PDFs since the absolute minimum of $C_2 I_f$ is 1, obtained for the Gaussian PDF. In practice, α is taken small enough (see [4]).

4 ALGORITHM

Each step of the algorithm consists of the three parts:

1) Estimate the cumulants C_2 and C_4 of the prediction error process w_t (5).

2) Calculate the model f_M^0 for the input PDF based on the estimated cumulants of w_t. Following the sign of C_4, we choose between the models (11) or (12) (with $\alpha = (\sqrt{\frac{3C_2^2}{C_4}})/100$ here) for f_M^0.

3) Find the minimum of the criterion (6) (with $\ell(w) = \ell^0(w)$) in the search direction of a quasi-Newton algorithm, calculated with the input model f_M^0. This calculus is not detailed here but it is similar to the one presented in [8].

In the initialization phase of our ML approach, any $4th$-order methods can be used, for example, the W-slice algorithm [9], to avoid convergence to false local minima.

5 ASYMPTOTIC BEHAVIOUR

From the Propositions 2 and 3, f_M^0 (13) is the PDF minimizing the FI under cumulants constraints of the class of SGM_α, but also of the class of SPG PDFs. In Huber's sense [4, 7], if the true PDF f_e belongs to this class, we obtain the most robust estimator (7), of the estimator class generated by the class of SPG PDFs, with the norm $\ell^0(.)$. It is possible to show that the proposed estimator $\widehat{\theta}$ (7) is asymptotically optimal in the minimax sense for the particular class of SPG PDFs.

The error covariance matrix of $\sqrt{N}(\widehat{\theta} - \theta_0)$ is noted $V(\ell, f)$, where f would be the actual PDF and ℓ would be the norm used for the identification. Under some assumptions (see [8]), the estimate (7) is consistent and the following expressions hold

$$\left\{ \begin{array}{lcll} \sqrt{N}(\widehat{\theta} - \theta_0) & \sim & \mathcal{N}(0, V(\ell^0, f_e)) & (a) \\ V = V(\ell^0, f_e) & \leq & V(\ell^0, f_M^0) = V^* & (b) \end{array} \right. \tag{15}$$

Thus, for $f_e = f_M^0$, the asymptotic covariance $V(\ell^0, f_e)$ of the proposed estimate (7) reaches the lower possible boundary V^*, which depends on the FI of f_M^0 and on θ_0. Its calculus is detailed in [8]. For other f_e belongs to the SPG class, the asymptotic covariance does not exceed V^*. If f_e is not in the SPG class of PDFs, only the relation (15a) holds, which is the case for the class of sub-Gaussian PDFs here. In all the cases, $V(\ell^0, f_e)$ is obtained theoretically with the results of [8] and [5].

6 SIMULATION RESULTS

In this section, we compare the performance of our ML approach to a method [1, 2] (noted LS+max$|K|$) that deals exactly the same problem we consider here. Then we will check the asymptotic efficiency of our approach.

We made many simulations with a non causal ARMA model [6] driven by different symmetrical input noises belonging to sub- or super-gaussian class of PDFs: laplacian (I)(SPG), uniform (II)(sub-gaussian) or Bernoulli-Gaussian (III)(SPG) with a probability of spike equals to 0.1.

So, the following non causal ARMA model [6] is considered:

$$H(z) = \frac{1 + 0.5z^{-1}}{(1 + 0.75z^{-1}).(1 - 0.8z)}$$

with two poles at -0.75 and 1.25 and one zero at -0.5.

First, 100 independant Monte-Carlo runs (MCR) were performed for each simulation. The signal's length used is $N = 2000$ samples. We compared this results to the method LS+max$|K|$, where the spectrally equivalent minimum phase system is primarily identified using least squares method (LS). Then, among all the spectrally equivalent systems, we choose the model which maximizes the absolute value of the estimated normalized kurtosis of the innovation process. Our ML approach is initialized by a $4th$-order method (W-slice [9]), due to the presence of local minima.

In Table 1, the mean and the standard deviation (Std) of the parameter estimates are summarized for

Table 1: ARMA parameter estimates (N=2000, 100 MCR).

Input	True model	ML Mean	ML Std	LS+max$\|K\|$ Mean	LS+max$\|K\|$ Std
I	a_1=-0.50	-0.5003	0.0379	-0.4987	0.0422
	b_1= 0.75	0.7491	0.0158	0.7496	0.0157
	d_1=-0.80	-0.7997	0.0277	-0.8008	0.0305
II	a_1=-0.50	-0.4956	0.0449	-0.4944	0.0488
	b_1= 0.75	0.7491	0.0201	0.7508	0.0184
	d_1=-0.80	-0.7974	0.0291	-0.7984	0.0326

Table 2: The Empirical Bias (N=2000, 2500 MCR).

True Value	Initial Value	Input noise I Bias	Input noise III Bias
a_1=-0.50	-0.38	0.0015	0.0001
b_1= 0.75	0.85	-0.0008	-0.0001
d_1=-0.80	-0.71	0.0011	0.0008

the input noises I and II belonging to two different classes. The presented results show the good behaviour of our method compare to the LS+max$|K|$ method with smaller bias and Std.

Secondly, to demonstrate the asymptotic efficiency and the optimality in the minimax sense of our ML approach for the SPG class of PDFs, we made many simulations with the same non causal ARMA model, but driven by the input noises I and III, the two being of the SPG class with one (III) belonging to the set SGM_α and the other (I) not.

The signal's length used is always $N = 2000$ samples. 2500 independant MCR were performed for each simulation, to estimate the covariance matrix $\hat{C} = \text{cov}(\widehat{\theta} - \theta_0)$ and to compare it to $N^{-1}V$ and $N^{-1}V^*$ (15) caculated theoretically with the results of [8] and [5] (see Table 3). It's obvious that the matrix V can not be estimated for real data since the true PDF f_e is unknown. Table 2 summarize the empirical bias for the two different input noises. Its second column gives the initial values used in the implementation of the algorithm to guarantee its global convergence. In Table 3, results seems to be worse for the noise I, perhaps because of the finite number of MCR or the small number of data points. But we see that $\hat{C}$ verify the relations (15), that is $\hat{C}$ tends to the matrix $N^{-1}V$ and is lower than $N^{-1}V^*$, for the two noises. And the bias squared is negligible compared to the diagonal of $\hat{C}$ which is the mean square error of the parameters. It seems that the most robust estimator, in Huber's sense, is obtained.

7 CONCLUSION

A possible way to obtain a robust parameter estimates in case of symmetrical non-Gaussian input is presented. The innovation of the proposed PDF model is that it is a GM parameterized by its second and fourth order cumulants and valid for the classes of sub-Gaussian or SPG PDFs. In the latter case, we take the model minimizing

Table 3: Performance of the ML approach (N=2000, 2500 MCR).

	Input noise I	Input noise III
$N^{-1}V$	$10^{-4}\cdot\begin{bmatrix}15.1 & 2.61 & 8.21\\ 2.61 & 2.71 & 0.93\\ 8.21 & 0.93 & 6.34\end{bmatrix}$	$10^{-4}\cdot\begin{bmatrix}3.81 & 1.00 & 1.26\\ 1.00 & 1.64 & 0.14\\ 1.26 & 0.14 & 1.55\end{bmatrix}$
$\hat{C}$	$10^{-4}\cdot\begin{bmatrix}15.9 & 2.71 & 8.86\\ 2.71 & 2.72 & 1.02\\ 8.86 & 1.02 & 6.87\end{bmatrix}$	$10^{-4}\cdot\begin{bmatrix}4.70 & 1.08 & 1.80\\ 1.08 & 1.75 & 0.18\\ 1.80 & 0.18 & 1.92\end{bmatrix}$
$N^{-1}V^*$	$10^{-4}\cdot\begin{bmatrix}17.0 & 2.84 & 9.44\\ 2.84 & 2.78 & 1.07\\ 9.44 & 1.07 & 7.15\end{bmatrix}$	$10^{-4}\cdot\begin{bmatrix}16.9 & 2.83 & 9.41\\ 2.83 & 2.78 & 1.06\\ 9.41 & 1.06 & 7.13\end{bmatrix}$

the FI in the set SGM_α (Propostion 2). The obtained estimator is the most robust estimator in Huber's sense and it is asymptotically optimal in the minimax sense for the SPG class.

Simulation results seem to confirm the good behaviour and robustness of our method compared to other methods based on higher order statistics, and the asymptotic efficiency of our approach.

References

[1] M. Boumahdi and J.-L. Lacoume. Blind identification of non-minimum phase filters by maximising the kurtosis. *Proc. of EUSIPCO 94, Edinburgh, Scotland, UK*, pages 195–198, September 1994.

[2] D. Donoho. *On minimum entropy deconvolution*, pages 565–608. In Applied Time Series Analysis, II. D.F. Findley, New York, Academic Press edition, 1981.

[3] G. Giannakis and J. Mendel. Identification of non-minimum phase systems using higher-order statistics. *IEEE Trans. on A.S.S.P.*, 37(3):360–377, March 1989.

[4] P. Huber. Robust estimation of a location parameter. *Ann. Math. Stat.*, 35:73–104, 1964.

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

[6] A. Petropulu. Noncausal nonminimum phase ARMA modeling of non-Gaussian processes. *IEEE Trans. on S.P.*, 43(8):1946–1954, August 1995.

[7] B. Poljak and J. Tsypkin. Robust identification. *Automatica*, 16:53–63, 1980.

[8] O. Shalvi and E. Weinstein. Maximum likelihood and lower bounds in system identification with non-Gaussian inputs. *IEEE Trans. Inform. Theory*, 40(2):328–339, March 1994.

[9] J.-L. Vuattoux, D. Zazula, and E. Le Carpentier. Causal ARMA parameter estimation using a linear combination of third-order cumulant slices. *Proc. of ATHOS, Begur, Girona, Spain*, pages 464–468, June 1995.

[10] V. Živojnović. Higher-order statistics and Huber's robustness. *Proc. of ATHOS, Begur, Girona, Spain*, pages 236–240, June 1995.