INVERSION OF H-$ARMA$ MODELS

David DECLERCQ, Patrick DUVAUT and Jerome SOUBIELLE
ETIS URA CNRS 2235
6, avenue du Ponceau 95014 Cergy-Pontoise FRANCE
e-mail: declercq@ensea.fr

ABSTRACT

We present in this contribution the problem of nongaussian H-$ARMA$ models inversion. We show that very classical methods of parameters identification based on the likelihood are unefficients in our case and we have chosen a fractionnal distance minimisation approach to estimate the nonlinearity. The $ARMA$ coefficients are identified with maximum likelihood estimators and a comparison study with the cumulant based method has been conducted on synthetic data.

1 INTRODUCTION

Nongaussian H-$ARMA$ models have been introduced in [1] and [2]. The mean of generation of a H-$ARMA$ process is twofold: a first linear $ARMA$ filter that provides a correlated gaussian process and then a nonlinear polynomial transformation with a combination of Hermite polynomials of various degrees. This kind of models is very near to those used in nonlinear digital communications, and especially in satellite communications. In these applications, the channels are considered as linear and the nonlinear effects due to amplification inside the satellite are modeled by a polynomial system [3].
An H-$ARMA(P,p,q)$ model is driven by the equations

$$x[n] = \sum_{i=1}^{p} a_i x[n-i] + \sum_{j=1}^{q} b_j \varepsilon[n-j] + \varepsilon[n] \qquad (1)$$

$$y[n] = \sum_{k=1}^{P} \alpha_k H_k\left(x[n]\right) \qquad (2)$$

where $\varepsilon[n]$ is a standardised white gaussian noise. Those equations define a Markovian representation of the model as $x[n]$ can be rewritten as a state vector.
Previous contributions on this class of models have precised the interesting properties of such processes, especially concerning the ouput cumulants properties. In this paper, we address the problem of inverting H-$ARMA$ models, that is to build estimators of the model parameters. In [2], we gave an intuitive but sub-optimal and computationally intensive solution to this problem. We identify an H-$ARMA$ model sequentially: the nonlinear α_k coefficients first, then the $ARMA$ coefficients (a_i, b_j). We will show in section 2 that methods based on the likelihood are unefficients - almost for the nonlinear part - and that the regularisation of a penalty function can provide acceptable estimates. The second part of the identification procedure deals with the $ARMA$ filter and we consider a likelihood approach to achieve its estimation (section 3). A comparison between the likelihood approach and the cumulants based method described in [2] is made. Finally, we discuss the drawbacks of those kind of inversion approaches and give the first keys to bypass them with the use of MCMC algorithms.

2 NONLINEAR IDENTIFICATION

2.1 Failure of likelihood based methods

The first and natural idea when one has to identify a model, in other words estimate the number of parameters that describe the model and their values, is to make use of the likelihood [4]. Various methods based on this function are available: maximum likelihood, minimizing a Kullback-Leiber distance, bayesian methods ...
For example, when one tries to identify a linear AR process, the likelihood is gaussian

$$L(y[1 \to N], \mathbf{a}, \sigma) \propto \frac{1}{\sigma^N} exp \left(-\frac{1}{2\sigma^2} \sum_{n=1}^{N} \left(y[n] - \sum_{j=1}^{P} a_j y[n-j] \right)^2 \right)$$

given $y[n] = 0 \;\; \forall n < 0$, where σ^2 is the variance of the input gaussian noise and $\mathbf{a}$ the vector of the autoregressive coefficients. $y[1 \to N]$ means that we observe N samples of the output process.
The main reason that make these methods efficient is that a model is *effectively* the most likely given a set of observations, what means that the likelihood function is sufficiently regular and has a global maximum.
In the course of identifying the nonlinear part of H-$ARMA$ models, the likelihood has no explicit expression, but can be numerically calculated. Suppose that we have a white gaussian noise $x[n]$ filtered by a Hermite nonlinearity with only the two polynomials of degree 2 and 3: $y = \alpha_2 H_2(x) + \alpha_3 H_3(x)$. We have generated

$N = 2000$ samples of such a process with $\alpha_2 = 2$ and $\alpha_3 = 1$ and drawn the likelihood $L(y[1 \to N], \alpha_2, \alpha_3)$ on figure 1. Knowing the nonlinear coefficients, we have inverted the polynomial defined with (α_2, α_3) in order to calculate the probability of the transformed process $y[n]$ [5]. We can notice that the likelihood function does not exhibit a single maximum but rather a crest of maxima. It is easy to understand that any method of identification based on that form of likelihood will be *trapped* somewhere in the crest. We have then to find other ways to solve or, at least regularise, this problem (in the sense of making it possible to identify).

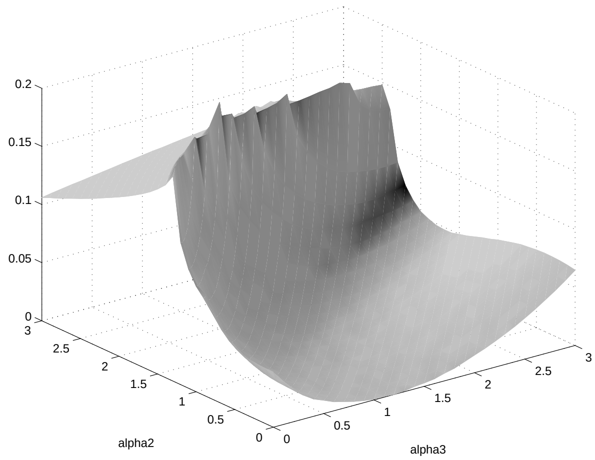

Figure 1: numerical likelihood of an H-$ARMA$ process

2.2 Regularisation with fractionnal distance

A lot of distances are used in signal processing - between processes, probability density functions, spectra ... we can refer to [6] for an extensive presentation. We have chosen, after a few attempts with classical distances such as $\mathcal{L}_1$ or $\mathcal{L}_2$, a *fractionnal distance* between an histogram of the data and a numerical evaluation of the posterior distribution knowing the nonlinear coefficients. The first step is to obtain a normalised histogram $\hat{f}$ of the data ($N = 2000$) and to keep only the relevant values (those which have more than 5 counts for example). This ensures that we will try to match the estimated distribution only on a restricted and valid domain. The distance taken is defined as

$$d_\beta(\hat{f}, f) = \left(\sum_i \left(\hat{f}(i) - f(i|\alpha_2, \alpha_3) \right)^\beta \right)^{1/\beta} \qquad \beta < 1 \qquad (3)$$

d_β is then a function of (α_2, α_3) and the value of β has been empirically fixed to 0.1. We have drawn on figure 2 minus the distance d_β in order to exhibit a maximum instead of a minimum. We can remark that, with the same process used in the above section, this kind of distance has a single maximum with a regular behaviour in its neighbourhood. Using this fractionnal distance as a penalty function, we have the solution:

$$(\hat{\alpha}_2, \hat{\alpha}_3) = \min_{\alpha_2, \alpha_3} d_\beta(\hat{f}, f_{|\alpha_2, \alpha_3}) \qquad (4)$$

The global minimization is achieved using a sequential grid search with decreasing step length. We use an exhaustive search in the parameter space because the classical means to increase the convergence rate such as a *gradient descent* are sometimes trapped in local minima. A combined algorithm with simulated annealing and gradient descent has been tested. The same minimum as in the grid search strategy has been reached, but in a more extensive time computation. The results are reported in table 1.

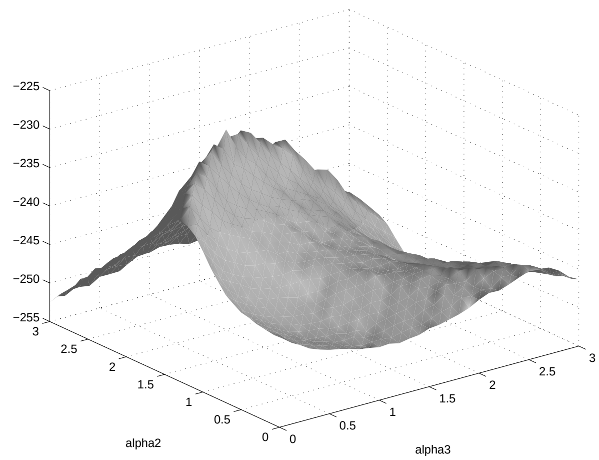

Figure 2: regularisation of the inversion problem with a fractionnal distance

3 ARMA IDENTIFICATION

As previously told, we make the identification of the nonlinear H-$ARMA$ models with a sequential approach. We have obtained the nonlinear coefficients α_k and with this prior information, we face the problem of estimating the $ARMA$ order and coefficients. In [2], we have presented a method based on the covariance and the bicovariance functions of the output process $y[n]$. This method provides acceptable results when we have enough sample to yield accurate estimates for the covariance an the bicovariance of $y[n]$. Moreover, we have shown in [1] that for H-$ARMA$ process, the bicovariance is well estimated when almost 3000 samples are considered. To tackle with this limitation, we propose in this section to build a MLE (Maximum Likelihood Estimator) and compare the estimation performances when the sample size become smaller $N < 500$.
Given the feasible AR coefficients $\mathbf{a}$ and MA coefficients $\mathbf{b}$, and knowing the Hermite coefficients α_k, the likelihood function is $L(y[1 \to N], \mathbf{a}, \mathbf{b}|\alpha_k)$ and cannot, as in the previous section, be analytically expressed. We must then evaluate it numerically. We describe below

the mean to obtain the likelihood for an AR filter of order p and an Hermite filter $\mathbf{H}(x)=\sum_{k=1}^{P}\alpha_k H_k(x)$.

> Let $y[1\to N]$ be the observed data,
>
> $\forall i \in [p+1, N]$
>
> $\mathbf{y}_i = [y[i]\dots y[i-p]]$
>
> $\mathbf{X_i} = [\mathbf{x}[i]\dots \mathbf{x}[i-p]] = \mathbf{H}^{-1}(\mathbf{y}_i)$
>
> where $\mathbf{x}[i]$ is a vector that collects the $P_1 \le P$ solutions of $\mathbf{H}^{-1}(y[i])$
>
> $\Delta = \{X_i^{\mathbf{k}} = [x_{k_i}[i]\dots x_{k_{i-p}}[i-p]]\}$
>
> where $x_{k_i}[i]$ is one of the components in $\mathbf{x}[i]$. Δ collects all the possible vectors extracted from $\mathbf{X}_i$.
>
> $L_i(\mathbf{y}_i,\mathbf{a}) = \sum_{\mathbf{k}\in\Delta} f(X_i^{\mathbf{k}},\boldsymbol{\mu}_p,\boldsymbol{\Gamma}_p) * J(\mathbf{H}^{-1})[X_i^{\mathbf{k}}]$
>
> where $f(X_i^{\mathbf{k}},\boldsymbol{\mu}_p,\boldsymbol{\Gamma}_p)$ is the p-dimensional gaussian density with mean $\boldsymbol{\mu}_p$ and covariance $\boldsymbol{\Gamma}_p$ build with the AR coefficients $\mathbf{a}$. $J(\mathbf{H}^{-1})[X_i^{\mathbf{k}}]$ is the jacobien of the transformation $\mathbf{H}^{-1}$ at $X_i^{\mathbf{k}}$.
>
> and then
>
> $L(y[1\to N],\mathbf{a}) = \left(\prod_i L_i(\mathbf{y}_i,\mathbf{a})\right)^{1/i}$

The maximisation of the function $L(y[1\to N],\mathbf{a})$ over $\mathbf{a}$ gives the MLE estimator of the AR coefficients.
Let us briefly present the cumulant's method: it is based on the theoretical expressions of the covariance and the bicovariance of the H-$ARMA$ process $y[n]$ in terms of the covariance of the linear process $x[n]$. Those expressions are determined knowing the estimates of α_k and we make use of them to build estimators of the autocovariance sequence of $x[n]$. A Levinson type algorithm gives the $ARMA$ coefficients. We can refer to [2] for more details.
We have gathered in table 1 the results on a synthetic sample drawn from a H-$ARMA$ model with hermite coefficients ($\alpha_2 = 2, \alpha_3 = 1$), AR poles $\mathbf{p} = (0.9, 0.7)$ and no MA zeros. We study the influence of the number of observation samples on the estimation accuracy for the two methods.

N	100	500	3000
$\boldsymbol{\alpha}$	(2.38,0.86)	(1.97,0.96)	(2.02,-1)
$\mathbf{p}$ (cumulants)	(0.99,0.62)	(0.95,0.45)	(0.898,0.66)
$\mathbf{p}$ (MLE)	(0.94,0.63)	(0.93,0.64)	(0.9,0.69)

Table 1: estimations of an H-$ARMA$ model

A first sight at this table shows that the accuracy of the estimators increases with the number of samples. The nonlinear parameters are well estimated for very large samples (notice that α_3 is determined up to a sign) but badly for $N = 100$. This was expected and this poor behaviour comes mainly from the estimation of an histogram with only 100 points. We also see that the MLE esimators of the poles of the linear filter are always better than the cumulant's method ones. Actually, the cumulants based method seems not to be valid, even for large sample sizes, while the MLE give good estimates for samples sizes greater than $N = 1000$. We stress the fact that the choosen nonlinear coefficients lead to a *hard* nonlinearity and therefore to a bad behaviour for the estimators. Other *cooler* nonlinearities (with differents coefficients) have provided more accurate estimates.

4 MCMC APPROACH TO H-$ARMA$ IDENTIFICATION

The two major drawbacks to the estimation procedures described in the previous sections are *(i)* the fact that we estimate separately the nonlinear and the linear coefficients ; it is well known that sequential methods lead to sub-optimal estimators for the parameters and it would be preferable to use a *global* methods *(ii)* these methods are valid when a large number of observation samples are concerned.
With the help of the Markovian description of the model (1),(2), we can solve the problem of parameter estimation by a *fully* Bayesian approach [7]. The key of such a method is the famous *data augmentation* procedure which allows to express analytically the conditionnal posterior densities of the coefficients with respect to the augmented samples [8]. We can therewith make use of a Monte Carlo Markov Chain algorithm to sample from these posterior distributions and obtain maximum *a posteriori* estimators. This approach allows to jointly estimate both the linear and the nonlinear parameters and to overcome the problem of small observation sample sizes. It has moreover the advantage of being nonsensitive to the initialisation of the algorithm, which is actually a major drawback in MLE methods.
Fisrt, we have to modify equation (2) by introducing a *regularisation* white gaussian noise, which can be interpreted as an estimation deviate:

$$y[n] = \sum_{k=1}^{P}\alpha_k H_k\left(x[n]\right) + \eta[n] \tag{5}$$

The Bayesian expansion of the fully posterior density follows

$$\begin{aligned} &p(x[1\to N],\boldsymbol{\alpha},\mathbf{a},\sigma_\varepsilon^2,\sigma_\eta^2|y[1\to N]) \\ &\quad\propto p(y[1\to N]|x[1\to N],\boldsymbol{\alpha},\sigma_\eta^2)\; p(x[1\to N]|\mathbf{a},\sigma_\varepsilon^2) \\ &\quad\quad p(\boldsymbol{\alpha},\sigma_\eta^2)\; p(\mathbf{a},\sigma_\varepsilon^2) \end{aligned} \tag{6}$$

where the likelihood of the augmented model $p(y[1\to N]|x[1\to N],\boldsymbol{\alpha},\sigma_\eta^2)$ is gaussian due to the gaussianity of $\eta[n]$. The conjugate priors are *gaussian* for $\mathbf{a}$ and $\boldsymbol{\alpha}$ and *inverse gamma* for σ_η^2 and σ_ε^2: these parameters are sampled directly with Gibbs steps. The

simulation of the hidden state variable $x[1 \to N]$ is more difficult and a Metropolis-Hastings step is required.
The implementation of this algorithm for the inversion of H-$ARMA$ models is under investigation and we are facing different problems such as the very low rate of convergence and the multimodality of the posterior density when hard nonlinearities are considered. The MCMC identification of this family of nonlinear processes will be detailled in a future contribution, and will be compared with the methods described in this paper and other classical Bayesian approaches (EM,...).

5 CONCLUSION

We have presented in this paper the inversion of H-$ARMA$ nongaussian-nonlinear models. We achieve this identification with two separated steps, the nonlinear Hermite coefficients first and then the $ARMA$ part. The former problem is ill-conditionned in the sense that the likelihood function exibits a crest of maxima and a regularisation of the problem with the help of a fractionnal distance has been considered. We achieve the minimisation of that distance with a combined simulated annealing-gradient descent algorithm. For the $ARMA$ coefficient, we have considered maximum likelihood estimators instead of the cumulant based method presented in [2]. This method allows to obtain more accurate estimations for every sample sizes. We have discussed the drawbacks of those methods which are mainly the sequential aspect of the approach and its lack of robustness when the number of observations becomes small. We introduce briefly a MCMC based method, which seems to be convenient for the identification of nonlinear processes. We will develop this approach in a future contribution.

References

[1] D. Declercq and P. Duvaut. Propriétés des processus *h-arma*. In *GRETSI-97*, volume 1, pages 151–154, Grenoble, 1997.

[2] D. Declercq and P. Duvaut. Nonlinear *h-arma* models. *submitted to ICASSP-98*, 1998.

[3] V. Castellani S. Benedetto, E.Biglieri. *Digital Transmission Theory.* Prentice-Hall, 1987.

[4] D. Guegan. *Séries Chronologiques non Linéaires à Temps Discret.* Stat. Math. Prob., Paris, 1994.

[5] B. Levine. *Fondements Théoriques de la Radiotechnique Statistique - Tome II.* Mir, Moscou, 1973.

[6] M. Basseville. Distance measures for signal processing and pattern recognition. *Signal Processing*, 18:349–369, 1989.

[7] J.M. Bernardo and A.F.M. Smith. *Bayesian Theory.* John Wiley & Sons, 1995.

[8] M.A. Tanner. *Tools for Statistical Inference : Methods for the Exploration of Posterior Distributions and Likelihood Functions.* Spinger-Verlag, New York, 1993.