IDENTIFICATION OF BILINEAR SYSTEMS USING BAYESIAN INFERENCE

Souad MEDDEB, Jean Yves TOURNERET and Francis CASTANIE

ENSEEIHT /GAPSE
2, rue Camichel BP7122 31071 Toulouse Cedex 7
e-mail: meddeb@len7.enseeiht.fr

ABSTRACT

A large class of non-linear phenomena can be described using bilinear systems. Such systems are very attractive since they usually require few parameters, to approximate most non-linearities (compared to other systems). This paper addresses the problems of bilinear system identification using Bayesian inference. The Gibbs sampler is used to estimate the bilinear system parameters, from measurements of the system input and output signals.

1. INTRODUCTION

Bilinear system belongs to the class of recursive polynomial systems. Its ability to represent many nonlinear systems efficiently and with a relatively small number of parameters is owing to its feedback structure [10]. The bilinear system has been used successfully in many signal processing applications [8]. For instance, a particular attention has been devoted to the modelling of seismological data, related to earthquakes and underground explosions using bilinear systems. The problem of estimating bilinear system parameters using measurements of the system input output signals has received much attention in the literature [2][4][5][6]. This paper studies a new approach for bilinear system identification based on Bayesian inference. The Gibbs sampler is used to simulate the a posteriori probability density function (pdf) of the parameters conditionally to the observation vector. The simulated a posteriori pdf can then be used to derive the Maximum A Posteriori (MAP) or Mean A Posteriori estimators. The paper is organized as follows. Section II presents the problem. Section III studies the Bayesian inference to solve the bilinear system identification problem. One of the most popular Monte Carlo Markov Chain algorithm denoted the Gibbs sampler is then studied. Simulation results and conclusions are presented in section IV and V respectively.

2. THE BILINEAR SYSTEM

Consider a discrete time invariant bilinear system defined by

$$y(t) = \alpha + \sum_{i=1}^{q_1} a(i)y(t-i) + \sum_{i=0}^{q_2} b(i)x(t-i) + \sum_{i=1}^{q_3}\sum_{j=1}^{q_3} c(i,j)x(t-i)y(t-j) + \varepsilon(t) \quad (1)$$

where $x(t)$ and $y(t)$, $t = 0, ..., T$ are the observed input and noisy output signals. $\varepsilon(t)$ is a zero-mean signal representing observation error whose variance σ^2 is unknown. The input signal $x(t)$ and the error $\varepsilon(t)$ are assumed to be statistically independent. The parameters $a(k)$ $(k = 0, ..., q_1)$, $b(k)$ $(k = 0, ..., q_2)$ are the linear kernels whereas $c(k_1, k_2)$ $(k_1, k_2 = 1, .., q_3)$ is the bilinear kernel. The bilinear kernel can be assumed symmetric without loss of generality, i.e. $c(k_1, k_2) = c(k_2, k_1)$. Eq.(1) can then be rewritten as follows

$$y(t) = \alpha + \sum_{i=1}^{q_1} a(i)y(t-i) + \sum_{i=0}^{q_2} b(i)x(t-i) + \sum_{i=1}^{q_3}\sum_{j=i}^{q_3} c(i,j)x(t-i)y(t-j) + \varepsilon(t) \quad (2)$$

Fig.1 shows the basic configuration of the bilinear system described by eq.(1).

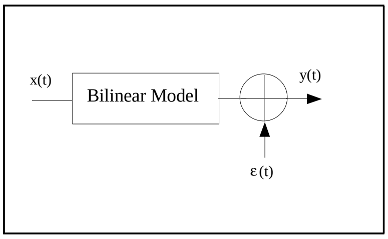

Fig.1: A schematic of the bilinear system

3. PARAMETER ESTIMATION

The bilinear system includes as a special case the standard ARMA model ($q_3 = 0$). Many ARMA parameter estimators have been generalized to the bilinear system identification problem (see [2] for a review of many existing method). However, many estimators face optimization problems. This paper proposes to use the Bayesian formalism and the Gibbs sampler to solve these optimization problems. The bilinear model in (2) is expressed as

$$y(t) = \alpha + Y(t-1)^T A + X(t)^T B + Z^T(t-1)C + \varepsilon(t) \quad (3)$$

where the parameter vectors are

$$\begin{aligned} A &= (a(1), ..., a(q_1))^T \\ B &= (b(0), ..., b(q_2))^T \\ C &= (c(1,1), \ldots, c(1,q_3), c(2,2), ..., c(q_3,q_3))^T \\ \Theta &= \left(\alpha,\ A^T,\ B^T,\ C^T\right)^T \end{aligned} \tag{4}$$

and

$$\begin{aligned} Y(t-1) &= (y(t-1), ..., y(t-q_1))^T \\ X(t-1) &= (x(t-1), ..., x(t-q_2), x(t-1-q_2))^T \\ Z(t-1) &= (x(t-1)y(t-1), \ldots, 2x(t-1)y(t-q_3), \\ &\quad x(t-2)y(t-2), ..., x(t-q_3)y(t-q_3))^T \\ E(t) &= \left(1,\ Y^T(t-1),\ X^T(t),\ Z^T(t-1)\right)^T \end{aligned} \tag{5}$$

Eq.(2) can be written in matrix form as

$$y(t) = \Theta^T E(t) + \varepsilon(t) \tag{6}$$

4. BAYESIAN FORMULATION

The Bayesian formalism assumes that some prior knowledge about the vector of unknown parameters Θ is available. It assumes that Θ is a random vector with a given prior probability density function (pdf). Bayesian approach, when applicable, can improve the estimation accuracy. Indeed, as specified in [9], the resultant estimator is optimal "on the average" or with respect to the assumed prior pdf of Θ. The problem of eliciting prior pdf on the parameter space of a signal hypothesis which is studied in [1], yields intractable computation for our problem. Instead, some usual criteria proposed in the literature for Bayesian signal identification are preferred [3]. More precisely, this paper assumes that the prior information I is expressed as follows:

$$\begin{aligned} A &\sim N(m_a, \Sigma_a), \quad B \sim N(m_b, \Sigma_b), \\ C &\sim N(m_c, \Sigma_c), \quad \alpha \sim N(m_\alpha, \sigma_\alpha^2) \\ \sigma^2 &\sim \mathcal{IG}(\frac{\nu_0}{2}, \frac{\gamma_0}{2}). \end{aligned} \tag{7}$$

Parameters $m_a, \Sigma_a, m_b, \Sigma_b, m_c, \Sigma_c, \nu_0$, and γ_0 are assumed to be known. However, the Bayesian algorithm is insensitive ti initialization of these parameters. The Bayesian approach is based on computing the a posteriori pdf of the parameter vector (Θ, σ^2) conditioned to the observation vector $y = (y(1), .., y(N))^T$ and the prior information I, abbreviated as $p(\Theta, \sigma^2/y, I)$.

Assume that the errors ε, $t = 1, ..., N$ are i.i.d Gaussian with variance σ^2. The pdf of $\varepsilon = (\varepsilon(1), \varepsilon(2), ..., \varepsilon(N))^T$ is

$$p(\varepsilon) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\varepsilon(t)^2\right\} \tag{8}$$

The Jacobian matrix of the transformation from ε to y is $\det(\frac{\partial y}{\partial \varepsilon}) = 1$. Consequently, using standard computation, the pdf of y conditioned to ε, Θ and σ^2 is

$$p(y/e, \Theta, \sigma^2, I) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y(t) - \Theta^T E\ (t))^2\right\} \tag{9}$$

By applying Bayes' theorem, the joint posterior pdf of (Θ, σ^2) is

$$p(\Theta, \sigma^2/y, I) = \frac{p(\Theta, \sigma^2/I)p(y/\Theta, \sigma^2, I)}{p(y/I)} \tag{10}$$

where $p(\Theta, \sigma^2/I)$ is the a priori pdf of the parameters given the prior information I, $p(y/\Theta, \sigma^2, I)$ is the direct pdf or likelihood function of the data, and $p(y/I)$ is the probability of the data given I, a normalization constant here. Thus,

$$p(\Theta, \sigma^2/y, I) \ \propto\ p(\Theta, \sigma^2/I)p(y/\Theta, \sigma^2, I) \tag{11}$$

where $\propto$ means proportional to. The knowledge of the a posteriori pdf in eq.(11) allows to compute two types of estimators:

- Mean estimators

$$E[f(\Theta, \sigma^2)/y] = \int \cdots \int f(\Theta, \sigma^2)p(\Theta, \sigma^2/y, I)\ d\Theta d\sigma^2$$

These estimators include the conditional expectation (f is the identity operator) conditional covariance and the marginal pdf.

- Maximum a posteriori (MAP) estimator.

Closed form expressions for the integral or maxima are often difficult to obtain. One solution of the numerical problem is the use of Monte Carlo Markov Chain [7]. This paper studies one of the most popular MCMC methods, the Gibbs sampler.

4.1. Gibbs sampler

The Gibbs sampler is a special case of the Metropolis-Hasting algorithm which is sometimes referred as algorithm of Metropolis-Hasting once at a time [7]. The Gibbs sampler proceeds as follows:

1. Initialization ($k = 1$): specify arbitrary starting values. A random parameter vector $(\Theta^0, \sigma^{2(0)}) = (\alpha^0, A^0, B^0, C^0, \sigma^{2(0)})$ is drawn according to $p(I)$.

2. Iteration k: draw $(\Theta^k, \sigma^{2(k)}) = (\alpha^k, A^k, B^k, C^k, \sigma^{2(k)})$ according to $p(\Theta, \sigma^2/y, I)$, i.e. draw:

- α^k according to $p(\alpha/y, A^{k-1}, B^{k-1}, C^{k-1}, \sigma^{2(k-1)})$

$$\alpha^k \sim N(m_{\alpha^k}, \sigma_\alpha^{2(k)})$$

- A^k according to $p(A/y, \alpha^k, B^{k-1}, C^{k-1}, \sigma^{2(k-1)})$

$$A^k \sim N(m_{a^k}, \Sigma_{a^k})$$

- B^k according to $p(B/y, \alpha^k, A^k, C^{k-1}, \sigma^{2(k-1)})$

$$B^k \sim N(m_{b^k}, \Sigma_{b^k})$$

- C^k according to $p(C/y, \alpha^k, A^k, B^k, \sigma^{2(k-1)})$

$$C^k \sim N(m_{c^k}, \Sigma_{c^k})$$

- $\sigma^{2(k)}$ according to $p(\sigma^2/y, \alpha^k, A^k, B^k, C^k)$

$$\sigma^{2(k)} \sim \mathcal{IG}(\frac{\nu_0 + \nu}{2}, \frac{\gamma_0 + \gamma}{2})$$

The Gibbs sampler yields a random vector at each step (the expressions of $m_{\alpha^k}, \sigma_\alpha^{2(k)}, m_{a^k}, \Sigma_{a^k}, m_{b^k}, \Sigma_{b^k}, m_{c^k}, \Sigma_{c^k}, \nu$

and γ are given in appendix). It is well known that the Markov Chain associated to the Gibbs sampler algorithm is irreducible and aperiodic. These two properties ensure the algorithm convergence. After a sufficiently long (so-called) burn-in, Markov Chain elements are used to approximate the marginal mean a posteriori estimates.

5. SIMULATION

The proposed bilinear system identification procedure is applied to a simulated input-output system following the exemple in [5]. The time-invariant bilinear system is given by

$$\begin{aligned} Y(t) &= 1.5X(t) + 1.2X(t-1) - 0.2X(t-2) \\ &\quad +0.7X(t-1)Y(t-1) - 0.1X(t-2)Y(t-2) \\ &\quad +\varepsilon(t) \end{aligned} \tag{12}$$

where $A = 0$, $\alpha = 0$, $B = [1.5\ 1.2\ {-0.2}]$, $C = [0.7\ 0\ {-0.1}]$. Note that $\Theta = [B\ C]^T$. Zero initial conditions were assumed. The input signal is a white Gaussian sequence with $T = 300$ and $SNR = 30dB$. In sampling process, the first 1000 draws are ignored (so called burn-in draws) and the next 4000 ($k = 5000$) are collected to approximate the posterior pdf $p(\theta_i/\Theta - \{\theta_i\})$. Fig's (2) and (3) show the a posteriori pdfs for the linear and bilinear parameters. Clearly, these pdfs are Gaussian whose means are the true parameters. Fig's (4) and (5) show the mean-square errors (MSE's) of linear and bilinear parameters averaged over 50 realizations as a function of SNR(dB).

The model order overestimation problem is then addressed. Table 1 shows the mean and standard deviation of linear and bilinear parameters estimates, in a particular example ($SNR = 20dB$). The true orders are $P = Q = 2$. The overestimated orders are $P = Q = 3$. Table 1 shows the insensitivity of the proposed algorithm to model order overestimation.

6. CONCLUSIONS

Time invariant bilinear systems were identified using Bayesian inference, from input, output measurements. The Gibbs sampler was used to simulate the parameter a posteriori probability density function. The marginal a posteriori mean estimators were then derived from the simulated a posteriori pdf. The proposed algorithm was show to be insensitive to model order overestimation. It can be generalized to other non-linear models including Volterra models, radial basis function models,...

7. APPENDIX

This appendix determines the means and the covariance matrices of parameter posterior pdfs for the k^{th} Gibbs sampler iteration. Assume that parameters priors are independent such that

$$p(\Theta, \sigma^2) = p(\alpha)p(A)p(B)p(C)p(\sigma^2)$$

Bayes' theorem (see (11)) yields:

$$p(A/y, I, \alpha, B, C, \sigma^2) \propto p(y/\alpha, A, B, C, \sigma^2)p(A) \tag{13}$$

The same results are obtained for parameters α, B and C. This paper uses Gaussian priors for the parameters (7). Consequently, the conditional pdf in (13) can be computed explicitly. A straightforward computation shows that these pdf's are Gaussian pdf's, whose means and covariance matrices are defined as follows:

Constant α

- $\frac{1}{\sigma_\alpha^{2(k)}} = \frac{N}{\sigma^2} + \frac{1}{\sigma_\alpha^2}$
- $m_{\alpha^k} = \sigma_\alpha^{2(k)}(\frac{1}{\sigma^2}\sum_{j=1}^N (y(t) - Y(t-1)^T A^k - X^T(t)B^k - Z^T(t-1)C^k) + \frac{m_\alpha^T}{\sigma_\alpha^2})$

Linear parameters A, B

- $\Sigma_{a^k}^{-1} = \frac{1}{\sigma^2}\sum_{j=1}^N Y(j-1)Y^T(j-1) + \Sigma_a^{-1}$
- $m_{a^k} = \Sigma_{a^k}(\frac{1}{\sigma^2}\sum_{j=1}^N (y(t) - \alpha - X(t)^T B^{(k-1)} - Z^T(t-1)C^{(k-1)})Y^T(j-1) + m_a^T\Sigma_a^{-1})$
- $\Sigma_{b^k}^{-1} = \frac{1}{\sigma^2}\sum_{j=1}^N X(j)X^T(j) + \Sigma_b^{-1}$
- $m_{b^k} = \Sigma_{b^k}(\frac{1}{\sigma^2}\sum_{j=1}^N (y(t) - \alpha - Y(t-1)^T A^k - Z^T(t-1)C^{(k-1)})X^T(j) + m_b^T\Sigma_b^{-1})$

Bilinear parameters C

- $\Sigma_{c^k}^{-1} = \frac{1}{\sigma^2}\sum_{j=1}^N Z(j-1)Z^T(j-1) + \Sigma_c^{-1}$
- $m_{c^k} = \Sigma_{c^k}(\frac{1}{\sigma^2}\sum_{j=1}^N (y(t) - \alpha^{(k-1)} - Y(t-1)^T A^k - X^T(t)B^k)Z^T(j-1) + m_c^T\Sigma_c^{-1})$

The noise variance conditional pdf can be determined as previously (using Bayes' theorem):

$$p(\sigma^2/y, I, \alpha, A, B, C) \propto p(y/\alpha, A, B, C, \sigma^2)p(\sigma^2)$$

An interesting choice for the noise variance prior is a conjugate prior. This kind of prior ensures computational tractability for many Bayesian problems. This paper uses the inverse gamma prior $\mathcal{IG}(\frac{\nu_0}{2}, \frac{\gamma_0}{2})$ for σ^2, which yields:

$$\sigma^{2(k)} \sim \mathcal{IG}(\frac{\nu_0 + \nu}{2}, \frac{\gamma_0 + \gamma}{2})$$

with

- $\nu = N$
- $\gamma = \sum_{j=1}^{N} (y(t) - Y(t-1)^T A^k - X^T(t)B^k - Z^T(t-1)C^k)^2$.

8. REFERENCES

[1] A. Quinn, "Novel Parameter Prior for Bayesian Signal Identification," Proc. IEEE ICASSP, pp. 3909-3912, 1997.

[2] F. Fnaiech and L. Ljung, "Recursive Identification of Bilinear Systems," Int. J. Control, Vol. 45, No. 2, pp. 453-470, 1987.

[3] G. E. P Bov and G. C Tiao, Bayesian Inference in Statistical Analysis, Addison Wesley, 1973.

[4] Junghsi Lee and V. John Mathews, "Output-error LMS Bilinear Filters with Stability Monitoring," Proc. IEEE ICASSP, pp. 3553-3556, 1997.

[5] Jonathon C. Ralston, Boualem Boashash, "Identification of bilinear systems using bandlimited regression," Proc IEEE ICASSP, pp. 3925-3928, 1997.

[6] M. B. Priestley, Non-linear and non-stationary time series analysis, Academic Press, 1991.

[7] Joseph J. K. O Ruanaidh, William J. Fitzgerald, Numerical Bayesian Methods applied to Signal Processing, Springer-Verlag New York, 1996.

[8] R. R. mohler and W. J. Kolodziej, "An over view of bilinear system theory and applications," IEEE Transaction on Systems, Man, and Cybernetics, Vol. SMC-10, pp. 683-688, 1982.

[9] Steven M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, PTR Prentice-Hall, 1993.

[10] V. john Mathews, "Adaptive Polynomial Filters," IEEE SP Magazine, pp. 10-26, July 1991.

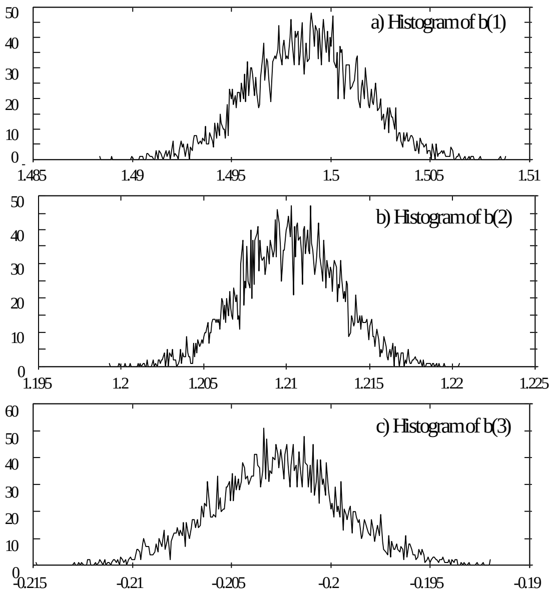

Fig.2: Estimation of $p(B/y, I, \Theta - \{B\}, \sigma^2)$.

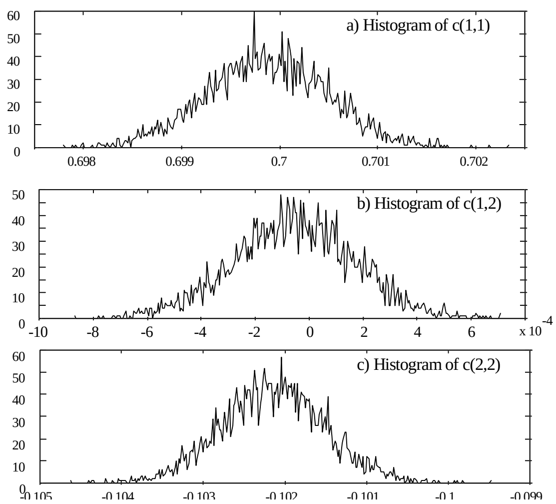

Fig.3: Estimation of $p(C/y, I, B, \sigma^2)$.

Parameter	Mean	Standard Deviation
$c(1,1) = 0.7$	0.694	$0.29e-3$
$c(1,2) = 0$	0.0039	$0.11e-3$
$c(2,2) = -0.1$	-0.108	$0.22e-3$
$\mathbf{c(1,3) = 0}$	0.0027	$0.12e-3$
$\mathbf{c(2,3) = 0}$	0.0000	$0.08e-3$
$\mathbf{c(3,3) = 0}$	-0.006	$0.03e-3$
$b(1) = 1.5$	1.5040	$0.11e-2$
$b(2) = 1.2$	1.1991	$0.11e-2$
$b(3) = -0.2$	-0.1978	$0.13e-2$

Table 1: Mean and strandard deviation of linear and bilinear parameter estimates

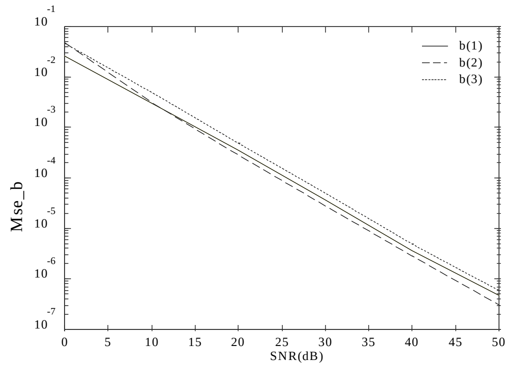

Fig.4: MSE of Linear Kernel Estimates

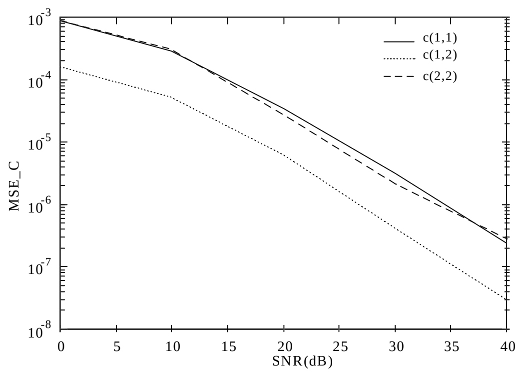

Fig.5: MSE of Bilinear Kernel estimates