PARAMETER ESTIMATION IN NONLINEAR SYSTEMS WITH AUTO AND CROSSCORRELATED NOISE

B. David and G. Bastin *

* *Center for Systems Engineering and Applied Mechanics,*
Université catholique de Louvain,
Av. G. Lemaitre 4, B1348 Louvain-La-Neuve, Belgium.
Phone: +32-10-472382, Fax: +32-10-472180.
Email: david@auto.ucl.ac.be

Abstract: The Gohberg-Heinig explicit formula for the inversion of a block-Toeplitz matrix is used to build an estimator of the inverse of the covariance matrix of a multivariable autoregressive process. This estimator is then conveniently applied to maximum likelihood parameter estimation in nonlinear dynamical systems with measurements corrupted by output additive auto and crosscorrelated noise. The efficiency of the obtained estimation scheme is illustrated via Monte-Carlo simulations. These simulations show that the proposed method improves significantly the statistical properties of the estimator and provides a more accurate confidence region around the estimated parameters, in comparison with classical methods. Furthermore, an appealing computational simplification is obtained due to the particular form taken by the Gohberg-Heinig formula. *Copyright© 2000 IFAC*

Keywords: Parameter Estimation; Nonlinear System; Maximum Likelihood.

1. INTRODUCTION

In (David and Bastin, 1998), a maximum likelihood (ML) parameter estimation method for nonlinear systems was proposed for the case of autocorrelated output noise. It was assumed that all state variables are measured and that there exist no crosscorrelation between the noise sequences corrupting each state measurement. Basically the method consisted of computing a preliminary weighted least-squares (WLS) estimate of the parameter vector, estimating the inverse of the noise covariance matrix from the residuals of this preliminary estimate and using this inverse covariance matrix (ICM) to compute a ML estimate. The originality of the method was in the particular estimate of the ICM. This estimate was derived from the Gohberg-Semencul inversion formula for a Toeplitz matrix (Gohberg and Semencul, 1972) and required to identify an autoregressive (AR) model of the residuals. In (David and Bastin, 1999) the method was successfully applied to a real live application.

In this paper the method is generalized by removing two restrictive assumptions. Firstly, the measured output variables are not necessarily the state variables and secondly, both the auto and the crosscorrelation of the noise sequences are taken into account. This leads to a more complicated block-Toeplitz structure of the covariance matrix. The generalization of the Gohberg-Semencul inversion formula to a block-Toeplitz matrix is therefore required in order to follow the same approach as in (David and Bastin, 1998). This formula is obtained as a particular case of (Gohberg and Heinig, 1974) and will require two multivariable AR models of the residuals to be identified, a causal and an anticausal one.

[1] This paper presents research results of the Belgian Programme on Interuniversity Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. The scientific responsibility rests with its authors.

2. PROBLEM FORMULATION

It is assumed that a phenomenological model of the system under consideration is available to the user. The model is given under the quite general form of a set of deterministic differential nonlinear state-space equations combined with a set of static nonlinear output equations. Both dynamic and static parts may be parametrized, the set of parameters being grouped into a single vector $\theta=[\theta_1,\ldots,\theta_p]$. The model structure is as follows:

$$\begin{cases} \dot{x}(t) = f(x(t),\theta,u(t)) \\ y(t) = g(x(t),\theta,u(t)) \end{cases} \quad (1)$$

where $x(t)=[x_1(t),\ldots,x_n(t)]^T$ is the state vector, $u(t)=[u_1(t),\ldots,u_m(t)]^T$ the input vector and $y(t)=[y_1(t),\ldots,y_q(t)]^T$ the output vector.

The parameter estimation problem is to estimate the parameter values from input and output data obtained from a single experiment carried out on the system in presence of output additive correlated noise. The experiment is performed with a known input signal $u(t)$ and a known initial state $x(t_0)$. The measurements of the output variables are recorded at N evenly distributed time instants $t_1,\ldots,t_N$ and are denoted: $z(j) = [z_1(j),\ldots,z_q(j)]^T$.

For a given input signal and a given initial state, the solution of the state and output equations in (1) is parametrized by θ and denoted $x(t,\theta)$ and $y(t,\theta)$ respectively. The state sensitivity matrix, $\frac{\partial x(t,\theta)}{\partial\theta}$, is obtained by integrating the following matrix differential equation along with the system state equations, see (Walter and Pronzato, 1997):

$$\frac{\partial \dot{x}(t,\theta)}{\partial\theta} = \frac{\partial f}{\partial x^T}\frac{\partial x(t,\theta)}{\partial\theta} + \frac{\partial f}{\partial\theta}. \quad (2)$$

The output sensitivity matrix is derived from the state sensitivity matrix using:

$$G(t,\theta) = \frac{\partial y(t,\theta)}{\partial\theta} = \frac{\partial g}{\partial x^T}\frac{\partial x(t,\theta)}{\partial\theta} + \frac{\partial g}{\partial\theta}. \quad (3)$$

There is always a deviation between the model output evaluated at the sampling instants, $y(t_j,\theta)$, and the measurements $z(j)$. The origin of this deviation may be multiple: modeling error, input or process noise and measurement noise. It is usually called the output error or the residuals and is denoted:

$$w(j,\theta) = z(j) - y(t_j,\theta) \quad j = 1,\ldots,N.$$

In the present framework, the sequence of output error vectors $w(j,\theta)$ is viewed as a realization of a stationary zero mean stochastic process with covariance matrix at lag k defined by:

$$\Sigma_k = E\{w(j)w(j-k)^T\} \quad \in I\!R^{q\times q}. \quad (4)$$

In most applications, the residual sequences corresponding to each output variable are assumed to be independent, i.e. not crosscorrelated, which amounts to considering a diagonal covariance matrix. Often also, each residual sequence is assumed to be uncorrelated along the time, i.e. not autocorrelated, which amounts to considering only Σ_0. In (David and Bastin, 1998), the case of autocorrelated residuals was treated only. Here the more general problem of auto and crosscorrelated residual sequences is addressed.

Let us define the following compact notations:

$$\begin{array}{lll} \mathbf{x}(\theta) & = [x(t_1,\theta)^T & ,\ldots,\ x(t_N,\theta)^T]^T \in I\!R^{nN} \\ \mathbf{y}(\theta) & = [y(t_1,\theta)^T & ,\ldots,\ y(t_N,\theta)^T]^T \in I\!R^{qN} \\ \mathbf{G}(\theta) & = [G(t_1,\theta)^T & ,\ldots,\ G(t_N,\theta)^T]^T \in I\!R^{qN\times p} \\ \mathbf{z} & = [z(1)^T & ,\ldots,\ z(N)^T]^T \in I\!R^{qN} \\ \mathbf{w}(\theta) & = [w(1,\theta)^T & ,\ldots,\ w(N,\theta)^T]^T \in I\!R^{qN} \\ & = \mathbf{z} - \mathbf{y}(\theta). & \end{array}$$

The full covariance matrix, $\mathbf{\Sigma} = E\{\mathbf{w}(\theta)\mathbf{w}(\theta)^T\}$, of the random vector $\mathbf{w}(\theta)$ is then a very large $(qN \times qN)$ block-Toeplitz matrix of the form:

$$\mathbf{\Sigma} = \begin{bmatrix} \Sigma_0 & \Sigma_{-1} & & \Sigma_{1-N} \\ \Sigma_1 & \cdot & \cdot & \\ & \cdot & \cdot & \cdot \\ & & \cdot & \cdot & \Sigma_{-1} \\ \Sigma_{N-1} & & & \Sigma_1 & \Sigma_0 \end{bmatrix}. \quad (5)$$

Assuming furthermore a normal probability density function for $\mathbf{w}(\theta)$, and defining the following scalar cost function:

$$J(\theta) = \mathbf{w}(\theta)^T\mathbf{\Sigma}^{-1}\mathbf{w}(\theta), \quad (6)$$

the ML estimate of θ, for $\mathbf{\Sigma}$ known, is then given by, see e.g. (Seber and Wild, 1989):

$$\hat{\theta}^{ML} = \arg\min_{\theta} J(\theta). \quad (7)$$

The gradient of the cost function and the Cramér-Rao (CR) lower bound on the covariance matrix of the estimator are given respectively by:

$$\frac{\partial J(\theta)}{\partial\theta} = -2\ \mathbf{w}(\theta)^T\mathbf{\Sigma}^{-1}\mathbf{G}(\theta) \quad (8)$$

$$\Sigma_{\hat{\theta}^{ML}} \succeq [\mathbf{G}(\theta)^T\mathbf{\Sigma}^{-1}\mathbf{G}(\theta)]^{-1}. \quad (9)$$

It turns out that the ICM, $\mathbf{\Sigma}^{-1}$, is required at several levels of the computation of the ML estimate of θ. It appears in the cost function (6) that has to be minimized, in the gradient (8) of the cost function that is needed if a gradient search method is used for the nonlinear minimization problem and finally, in the CR lower bound (9) which is commonly used to build a confidence region around the estimated parameter. This ICM is unknown in practice and has therefore to be estimated. An original solution to this problem is given in the next section.

3. GOHBERG-HEINIG INVERSE

In (Gohberg and Semencul, 1972) an explicit formula was derived for the inversion of a finite Toeplitz matrix. This result was generalized two years later (Gohberg and Heinig, 1974) for a Toeplitz matrix with entries belonging to a noncommutative algebra, the block-Toeplitz matrix being a particular case. Other formulas exist for this inversion problem and several numerical methods to solve it have been developed and are still under investigation. It will be shown in this section that the Gohberg-Heinig formula can be very conveniently used in the ML parameter estimation framework.

Let us postulate a causal and an anticausal multidimensional AR representation of the residuals, defined by the following relations:

$$\begin{cases} w(j) + \sum_{r=1}^{d} A_r^+ w(j-r) = e^+(j) \\ w(j) + \sum_{r=1}^{d} A_r^- w(j+r) = e^-(j) \end{cases} \quad (10)$$

where A_r^+ and A_r^- are q by q matrices containing the coefficients of the causal and anticausal AR models, $e^+(j)$ and $e^-(j)$ are i.i.d. random vectors with covariance matrix equal respectively to $\Sigma_{e^+} = E\{e^+(j)e^+(j)^T\}$ and $\Sigma_{e^-} = E\{e^-(j)e^-(j)^T\}$.

The Yule-Walker equations generalized for the multidimensional case can be explicitly derived by substituting $w(j)$ in (4) by its expression coming from (10), taking into account the stationary assumption. This provides the following sets of relations, for $k = 0, \ldots, N-1$:

$$\begin{cases} \sum_{r=0}^{d} A_r^+ \Sigma_{k-r} = \delta_{0,k} \Sigma_{e^+}, & A_0^+ = I_q \\ \sum_{r=0}^{d} A_r^- \Sigma_{r-k} = \delta_{0,k} \Sigma_{e^-}, & A_0^- = I_q \end{cases} \quad (11)$$

The use of the Gohberg-Heinig formula requires first to find the coefficients that verify the so-called *generating system*. The inverse matrix is then obtained explicitly from these coefficients. In the general case, this system is constituted of four sets of equations. In the particular case of symmetric block-Toeplitz matrix, i.e. with blocks satisfying $\Sigma_{-k} = \Sigma_k^T$, the *generating system* is reduced to two sets of equations and corresponds exactly to (11). Hence, for the inversion of the covariance matrix of an AR process, the coefficients of the causal and anticausal AR models are also the generating coefficients. Therefore (Gohberg and Heinig, 1974, Corollary 1.1) directly gives the following explicit expression for the ICM of $\mathbf{w}(\theta)$:

$$\boldsymbol{\Sigma}^{-1} = \mathbf{U}^T \boldsymbol{\Sigma}_{e^+}^{-1} \mathbf{U} - \mathbf{V}^T \boldsymbol{\Sigma}_{e^-}^{-1} \mathbf{V} \quad (12)$$

where $\mathbf{U}$ and $\mathbf{V}$ are $qN \times qN$ lower triangular block-Toeplitz matrices of the form:

$$\mathbf{U} = \begin{bmatrix} I_q & & & & \\ A_1^+ & \ddots & & & \\ \vdots & & \ddots & & \\ A_d^+ & \ddots & & \ddots & \\ 0 & \ddots & \ddots & & \ddots \\ & \ddots & & \ddots & \\ 0 & 0\, A_d^+ & A_1^+ & I_q & \end{bmatrix}, \mathbf{V} = \begin{bmatrix} 0 & & & & \\ & \ddots & & & \\ 0 & & \ddots & & \\ A_d^- & \ddots & & \ddots & \\ & \ddots & \ddots & & \ddots \\ & & \ddots & \ddots & \\ A_1^- & & A_d^-\, 0 & & 0 \end{bmatrix}$$

while $\boldsymbol{\Sigma}_{e^+}^{-1}$ and $\boldsymbol{\Sigma}_{e^-}^{-1}$ are $qN \times qN$ block-diagonal matrices of the form:

$$\boldsymbol{\Sigma}_{e^+}^{-1} = \begin{bmatrix} \Sigma_{e^+}^{-1} & & \\ & \ddots & \\ & & \Sigma_{e^+}^{-1} \end{bmatrix}, \boldsymbol{\Sigma}_{e^-}^{-1} = \begin{bmatrix} \Sigma_{e^-}^{-1} & & \\ & \ddots & \\ & & \Sigma_{e^-}^{-1} \end{bmatrix}.$$

Assuming that the coefficients of the causal and anticausal AR models of the residual sequence are available, (12) provides a straightforward way to obtain the ICM required for computing the ML parameter estimate. In (David and Bastin, 1998), only the causal scalar AR model of each independent output noise sequence was necessary to build the ICM.

Besides the evident advantage that this ICM overcomes matrix inversion, the particular form of (12) provides also an appealing computational simplification. Indeed, in order to compute expressions involving the ICM, like (6), (8) and (9) that are of the form $\mathbf{L}^T \boldsymbol{\Sigma}^{-1} \mathbf{R}$, one has just to apply appropriate filters to the columns of $\mathbf{L}$ and $\mathbf{R}$. Here is a short proof:

Let $L(j)$ and $R(j)$ be any sequences of vector (matrices) [2] for $j = 1, \ldots, N$ and let $\mathbf{L}$ and $\mathbf{R}$ be their corresponding stacked vectors (matrices):

$$\mathbf{L} = [L(1)^T, \ldots, L(N)^T]^T$$
$$\mathbf{R} = [R(1)^T, \ldots, R(N)^T]^T.$$

Let us also define the following multidimensional discrete filters, associated to the AR models (10):

$$\mathcal{A}^+(z^{-1}) = I_q + A_1^+ z^{-1} + \ldots + A_d^+ z^{-d}$$
$$\mathcal{A}^-(z^{-1}) = A_d^- + A_{d-1}^- z^{-1} + \ldots + A_1^- z^{-d+1}$$

and let T_{e^+} and T_{e^-} denote the square root of the innovation covariance matrices: $\Sigma_{e^+} = T_{e^+} T_{e^+}$, $\Sigma_{e^-} = T_{e^-} T_{e^-}$. Then, considering the filtered versions of the $L(j)$ and $R(j)$ and their corresponding stacked matrices:

$$L_U(j) = T_{e^+}^{-1} \mathcal{A}^+(z^{-1}) L(j), \quad \mathbf{L_U} = [L_U(1)^T, \ldots, L_U(N)^T]^T$$
$$R_U(j) = T_{e^+}^{-1} \mathcal{A}^+(z^{-1}) R(j), \quad \mathbf{R_U} = [R_U(1)^T, \ldots, R_U(N)^T]^T$$
$$L_V(j) = T_{e^-}^{-1} \mathcal{A}^-(z^{-1}) L(j), \quad \mathbf{L_V} = [L_V(1)^T, \ldots, L_V(d)^T]^T$$
$$R_V(j) = T_{e^-}^{-1} \mathcal{A}^-(z^{-1}) R(j), \quad \mathbf{R_V} = [R_V(1)^T, \ldots, R_V(d)^T]^T$$

it becomes trivial to see, using the particular form of $\mathbf{U}$ and $\mathbf{V}$, that:

$$\mathbf{L}^T \boldsymbol{\Sigma}^{-1} \mathbf{R} = \mathbf{L_U}^T \mathbf{R_U} - \mathbf{L_V}^T \mathbf{R_V}.$$

[2] $L(j)=w(j)$ in (6,8) and $L(j)=G(t_j)$ in (9), $R(j)=w(j)$ in (6) and $R(j)=G(t_j)$ in (8,9).

In practice, to compute expressions involving the ICM, like (6), (8) and (9), one has just to apply appropriate filters, $\mathcal{A}^+(z^{-1})$ and $\mathcal{A}^-(z^{-1})$, to the columns or part of the columns of $\mathbf{L}$ and $\mathbf{R}$ and normalize with T_{e+} and T_{e-}. Hence, only the coefficients of the AR filters, A_r^+, A_r^-, Σ_{e+} and Σ_{e-} have to be carried along. The large ICM does not need to be formed explicitly. This is a major advantage that leads us to propose the following two-step ML estimation algorithm.

4. ESTIMATION ALGORITHM

The idea is to obtain a rough WLS estimate of θ in a first step and identify the causal and anticausal AR models on the residuals of this preliminary estimate. The anticausal model is simply obtained by presenting the residual vector in the reversed order to the AR model identification procedure. Then, these AR models are used to estimate the ICM needed to compute the ML estimate in a second step. The orders of the AR models can be fixed a priori or selected automatically using any appropriate order selection criterion, the minimum description length for instance. The WLS preliminary estimate is defined by:

$$\hat{\theta}^{WLS} = \arg\min_{\theta} \sum_{j=1}^{N} w(j,\theta)^T \zeta w(j,\theta)$$

where ζ is a q by q diagonal weighting matrix used to normalize the residuals in order to balance the individual contribution of each output variable. The weights can be determined from a preliminary data analysis. For instance, one could take them inversely proportional to the variances of the output measurements.

The proposed algorithm for ML parameter estimation in dynamical systems with auto and crosscorrelated output noise is then as follows:

(1) - obtain ζ from data analysis,
- compute a preliminary WLS estimate,

(2) - compute the residuals $\mathbf{w}(\hat{\theta}^{WLS})$,
- estimate the AR models of $\mathbf{w}(\hat{\theta}^{WLS})$,
- compute the ML estimate.

5. MONTE-CARLO SIMULATIONS

The purpose of this section is to illustrate, from Monte-Carlo simulations on a particular example, the statistical properties of the ML estimator obtained with the proposed two-step algorithm.

A three-parameter estimation problem on a nonlinear fourth order dynamical system with two output measurements is treated. The model is as follows:

$$\begin{cases} \dot{x}_1 = u - \theta_1 x_1 \\ \dot{x}_2 = -2x_2 + h_1(x_3, x_4, \theta_2) \\ \dot{x}_3 = x_1 - x_3 - h_2(x_2, x_3, \theta_3) \\ \dot{x}_4 = -x_4 + h_2(x_2, x_3, \theta_3) \\ \\ y_1 = x_4 \qquad\qquad h_1(x_3, x_4, \theta_2) = \dfrac{x_4}{1 + x_3/\theta_2} \\ y_2 = h_1(x_3, x_4, \theta_2) \quad h_2(x_2, x_3, \theta_3) = 4x_3 e^{-\theta_3 x_2} \end{cases}$$

The output trajectories of this model are computed for a *true* parameter $\theta^* = [8, 5, 10]$ and zero initial states. The experiment lasts 20 seconds. The input $u(t)$ is a piecewise constant signal and the output trajectories are sampled with a frequency of $10Hz$ to form the $\mathbf{y}(\theta^*)$ vector containing N=200 data samples.

Two different types of noise generator are used to produce the output error sequences. The first type is a first order bidimensional AR stochastic process with i.i.d. normal random innovation vectors. The objective of this first structure is to illustrate the behavior of the algorithm when the stochastic part of the system belongs to the model set scanned by the algorithm. The structure of the process is:

$$\begin{cases} w^{AR}(j) + \begin{bmatrix} -0.65 & -1.80 \\ -0.07 & -0.5 \end{bmatrix} w^{AR}(j-1) = e^{AR}(j) \\ \Sigma_{eAR} = 10^{-5} \begin{bmatrix} 90 & 3 \\ 3 & 2.5 \end{bmatrix} \end{cases}$$

The objective of the second type of noise generator is to evaluate the robustness of the algorithm when the stochastic part of the system is not in the model set. For that case, a bidimensional finite impulse response (FIR) filter of the following form is used:

$$w^{FIR}(j) = \begin{bmatrix} 0.5 & 0.25 \\ 0.05 & 0.1 \end{bmatrix} \begin{bmatrix} B_1(z^{-1})\, e_1^{FIR}(j) \\ B_2(z^{-1})\, e_2^{FIR}(j) \end{bmatrix}$$

where $B_1(z^{-1})$ and $B_2(z^{-1})$ are tenth order scalar FIR lowpass discrete filters with normalized cutoff frequencies equal to 0.1 and 0.08 respectively and $e_{1,2}^{FIR}(j)$ are i.i.d. normal random variables with unit variance.

The structure of the two processes have been scaled in order to produce highly auto and crosscorrelated noise sequences with a similar signal-to-noise ratio of 20dB on each measured output trajectory.

A bunch of independent realizations of the AR and FIR stochastic processes have been generated and added to the model output trajectory vector $y(t_j, \theta^*)$ to form two benchmarks of 5000 sets of measurements each. The input signal, the model output trajectories and one example of the measured output trajectories under both cases of stochastic process are illustrated in Fig.1.

For each set of measurements, an ML estimate of the parameter vector has been computed with

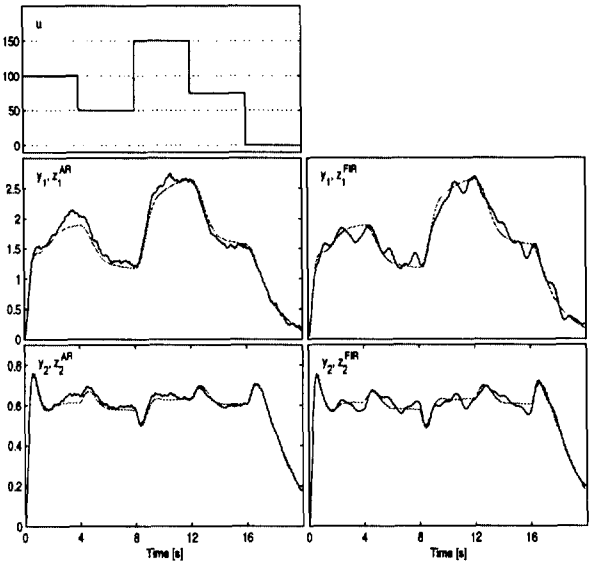

Fig. 1. Input signal u, output model trajectories y_i in dotted line and output measurements z_i in plain line.

the algorithm. The WLS estimate obtained in the first step has been recorded as well for comparison purpose. For a more appropriate comparison, the parameters estimated by a modified two-step algorithm is also computed. In that case, the full ICM estimate is replaced by the classical sample covariance estimate. The refined WLS estimate obtained with this modified algorithm is denoted RWLS .

The nonlinear minimization problem required to find the parameter estimates has been solved using the Nelder-Mead simplex method. The orders of the AR models were fixed a priori to d=2 for the first benchmark and d=10 for the second one.

The huge number of simulations will allow us to represent accurately several statistical properties of the estimators such as their empirical distribution and their standard deviation. Since the *true* parameter vector and the *true* noise structure are known, the CR lower bound on the covariance matrix of the parameter estimates can be computed exactly using (9). This bound provides the *minimum possible* standard deviation that can be achieved and will serve as a comparison basis.

The standard deviation of the 5000 independent WLS, RWLS and ML estimates of θ obtained under each case of stochastic process are given in Table 1, along with the CR bounds and quality indicators. These indicators measure the improvement achieved by the ML estimate with respect to the WLS and RWLS ones, relatively to the maximum possible improvement. These performance indicators range between 0% (no improvement) and 100% (maximum possible improvement) and are defined by:

$$\delta^{WLS}=100\frac{\sigma WLS-\sigma ML}{\sigma WLS-\sigma CR}, \quad \delta^{RWLS}=100\frac{\sigma RWLS-\sigma ML}{\sigma RWLS-\sigma CR}.$$

Table 1. Standard deviations and indicators.

	σWLS	$\sigma RWLS$	σML	σCR	δWLS	$\delta RWLS$
$\hat{\theta}_1^{AR}$	0.535	0.197	0.136	0.108	93.6%	69.2%
$\hat{\theta}_2^{AR}$	0.400	0.159	0.103	0.085	94.3%	75.8%
$\hat{\theta}_3^{AR}$	0.286	0.152	0.093	0.069	89.1%	72.6%
$\hat{\theta}_1^{FIR}$	0.566	0.351	0.136	0.087	89.7%	81.3%
$\hat{\theta}_2^{FIR}$	0.426	0.265	0.101	0.064	89.9%	81.7%
$\hat{\theta}_3^{FIR}$	0.288	0.190	0.093	0.069	89.1%	80.2%

The empirical distributions of the estimated parameters are illustrated in Fig.2, superimposed on normal distributions corresponding to the CR lower bound. In the AR case, i.e. the stochastic part of the system is in the model set, the ML estimate should asymptotically converge to the CR bound. It stands already very close to this bound for N=200 only. That proves the coherence of the method. The improvement with respect to the single step WLS estimate is impressive. In the FIR case a very high improvement with respect to the WLS estimate is also observed. That shows the ability of the proposed ICM estimate to be used also when the actual noise has a different structure. This enlarge considerably the domain of applicability of the ML estimate presented in this paper and let us believe that it could be used in the practical situation where the noise structure is even unknown. The benefit of the ML estimate in comparison with the RWLS is also quite valuable. This proves that for the considered example at least, the full ICM estimate clearly outperforms the classical sample covariance estimate.

The ICM estimate can still be advantageously applied to the construction of a confidence region around the estimated parameter vector. It is evident that the same experiment is rarely repeated 5000 times in order to have an idea of the parameter dispersion. Often, only one experiment is carried out and a confidence region around the identified parameter has to be evaluated. The covariance matrix of the parameter vector, $\Sigma_{\hat{\theta}}$, is commonly estimated using its CR bound (9) computed for $\hat{\theta}$ and $\hat{\mathbf{\Sigma}}^{-1}$. The large matrix $\hat{\mathbf{\Sigma}}^{-1}$ being estimated by inverting the sample covariance in the WLS case and with the full ICM estimate in the ML case. Assuming a normal distribution for $\hat{\theta}$, an approximate $100(1-\alpha)\%$ confidence region for the parameter vector $\hat{\theta}$ is given by, see e.g. (Seber and Wild, 1989):

$$\{\theta : (\theta-\hat{\theta})\hat{\Sigma}_{\hat{\theta}}^{-1}(\theta-\hat{\theta})^T \leq p\mathcal{F}_{p,N-p}^{\alpha}\} \quad (13)$$

where $\mathcal{F}_{i,j}^{\alpha}$ is the upper α critical value of the $\mathcal{F}_{i,j}$ distribution. The set described by (13) is the inner space of an ellipsoid centered on $\hat{\theta}$.

Fig.3 represents two projections of the 3D plot of the 5000 WLS and ML parameter estimates taken from the FIR set. Superimposed are the projections of the 99% confidence ellipsoids corresponding to one estimate chosen arbitrarily. It clearly

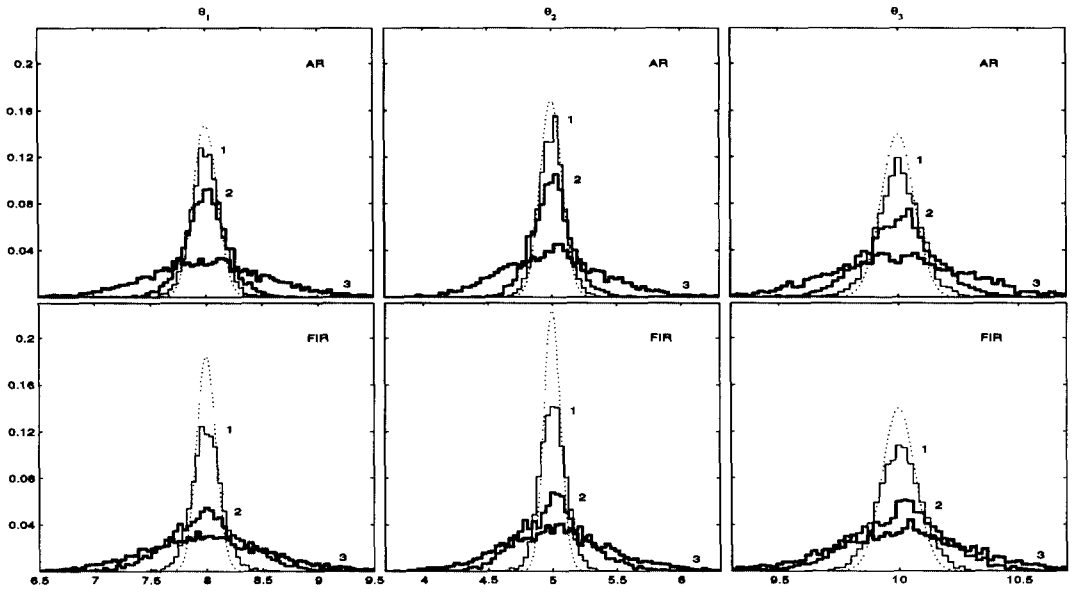

Fig. 2. The histograms are represented with a decreasing degree of line thickness: 3 is for WLS, 2 for RWLS and 1 for ML. The dotted lines represent the normal distributions corresponding to CR bounds.

appears that the confidence region estimated in ML case is far better than the one estimated in the WLS case. It includes the *true* parameter value and its shape better fits the actual parameters dispersion.

The computation of $\hat{\Sigma}_{\hat{\theta}}$ has been done systematically for each estimated parameter vector. If $\hat{\theta}$ had really been normally distributed and the estimate of each confidence region had been exact, 99% of those regions would have included the *true* parameter. Since those conditions are far to be met, it turns out that among the confidence regions computed with the WLS method, only 6% include the *true* parameter in the AR case and 21% in the FIR case. These percentages become 93% and 83% respectively with the ML method.

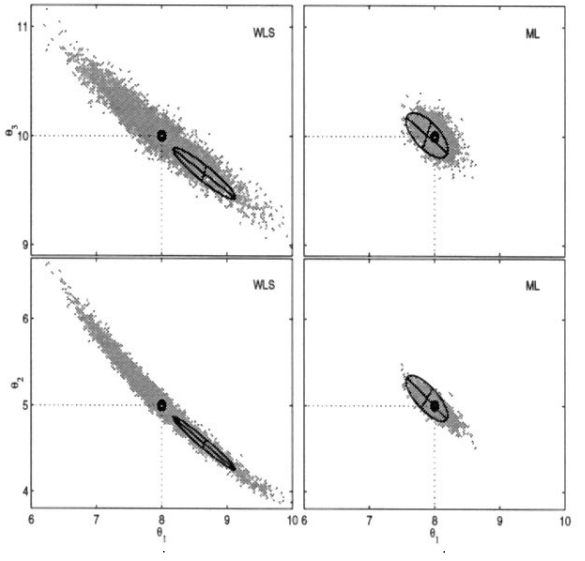

Fig. 3. Projections of the 5000 WLS and ML parameter estimates. The projections of the confidence ellipsoids are drawn for one estimate chosen arbitrarily. The bold circles point the true value of θ.

6. CONCLUSION

The maximum likelihood parameter estimation method developed in (David and Bastin, 1998) has been generalized in this paper to the case of auto and crosscorrelated output noise. The efficiency of the estimation method has been emphasized in comparison with classical one from Monte-Carlo simulations. A significant gain is achieved in the statistical properties of the estimator, namely a reduction of its standard deviation. In addition, the proposed ICM estimate provides an efficient way to compute accurately the confidence region around the estimated parameter.

The quite general form of the nonlinear deterministic part of the model (1) and the ability of the proposed ICM estimate to be successfully applied also when the noise structure is not in the model set suggest a wide range of application of the proposed method.

7. REFERENCES

David, B. and G. Bastin (1998). An estimator of the inverse covariance matrix and its application to ML parameter estimation in dynamical systems. Submitted.

David, B. and G. Bastin (1999). A maximum likelihood parameter estimation method for nonlinear dynamical systems. In: *38th IEEE Conf. on Decision and Control.* Phoenix, Arizona USA. pp. 612–617.

Gohberg, I. C. and A. A. Semencul (1972). On inversion of finite-section Toeplitz matrices and their continuous analogues. *Matem. Issled.* **7**, 201–224. (in Russian).

Gohberg, I. C. and G. Heinig (1974). Inversion of finite Toeplitz matrices with entries being elements from a noncommutative algebra. *Rev. Roum. Math. Pures et Appl.*. **XIX**(5), 623–663. (in Russian).

Seber, G.A.F. and C.J. Wild (1989). *Nonlinear Regression.* John Wiley & Sons.

Walter, E. and L. Pronzato (1997). *Identification of Parametric Models from Experimental Data.* Springer.