QUADRATIC SYSTEM IDENTIFICATION BY HEREDITARY APPROACH

Gibran ETCHEVERRY, Wael SULEIMAN, André MONIN
LAAS/CNRS, 7 ave. du Colonel Roche, 31077 Toulouse cedex 4, France
getcheve@laas.fr, suleiman@laas.fr, monin@laas.fr

ABSTRACT

Quadratic systems are the first kind of nonlinear systems in which we are interested in order to study polynomial nonlinearities. They can be approximated by bilinear models with nilpotent structure that approximate certain nonlinearities and generate finite degree Volterra series. The hereditary identification algorithm limited until now to linear systems is extended here for identification of nonlinear systems by implementing a canonical structure to the approximant of degree two (quadratic). A NARX (Nonlinear Autoregressive eXogenous input) multidimensional expression is employed in order to perform identification by hereditary computation.

1. INTRODUCTION

The interest on Volterra series for representing and approximating nonlinear systems has been demonstrated [1]. For instance, we find some applications on reducing amplifiers generated distortion within the transmission set [2-3]. The conditions for Volterra series realization by bilinear systems have been established in [4] and the importance of these systems is based on the fact that they can approximate nonlinear systems behavior with arbitrary precision [5]. The purpose stated in this article allows limiting the Volterra series degree; as a consequence, every kernel in the series can be generated by a bilinear system where the associated algebra is nilpotent under a condition on the kernels to be separable [6].

Identification of linear systems by hereditary computation has been already established in [7]. In order to extend it to quadratic systems, a canonical form which respects the linear dependence from the output to the system parameters has been developed. As an example, a scalar input/output nonlinear difference equation is approximated by a degree two Volterra series and a comparison is made between the hereditary and subspace identification algorithms.

2. CANONICAL FORM

A finite degree Volterra series is expressed as:

$$y_t = \sum_{d=1}^{D}\sum_{\tau_1=1}^{t}\sum_{\tau_2=1}^{\tau_1}\cdots\sum_{\tau_d=1}^{\tau_{d-1}} K_d\left(t,\tau_1,\cdots,\tau_d\right)u_{\tau_1}u_{\tau_2}\cdots u_{\tau_d} \qquad (1)$$

where the input/output are u_t, y_t. Without other assumptions about the kernels K_d, this series cannot be carried out by a finite dimensional system. However, if we force the kernels to be separable, which means to be a sum of finite products of functions of one variable, then the series can be accomplished by a bilinear nilpotent system [8].

Proposition 1: Every d-degree homogenous term on series (1) with a separable kernel can be achieved by the bilinear system:

$$\begin{aligned} x_t^1 &= A_t^1 x_{t-1}^1 + B_t u_t \\ x_t^2 &= A_t^2 x_{t-1}^2 + D_t^2 x_t^1 u_t \\ &\cdots \\ x_t^k &= A_t^d x_{t-1}^d + D_t^d x_t^{d-1} u_t \\ y_t^d &= C_t x_t^d \end{aligned} \qquad (2)$$

where $x_t^k \in \Re^{n_k}$, $\forall k = 1\cdots d$, $A_t^k \in \Re^{n_k \times n_k}$, $\forall k = 1\cdots d$, $D_t^k \in \Re^{n_k \times n_{k-1}}$, $\forall k = 2\cdots d$, $C_t \in \Re^{n_d}$ and $B_t \in \Re^{n_1}$. The dimension of x_t^k gives the separability order on τ_{k-1} and τ_k.

Proof: Every $x_t^k \in \Re^{n_k}$ can be considered as a linear system output with vector input $D_t^k x_t^{k-1} u_t$. Hence, taking as zero the initial values of x_0^k we can state:

$$x_t^1 = \sum_{\tau=1}^{t} \phi^1(t,\tau) B_\tau u_\tau$$

$$x_t^k = \sum_{\tau=1}^{t} \phi^k(t,\tau) D_\tau^k x_\tau^{k-1} u_\tau, \forall k = 2\dots d$$

where $\phi^k(t,\tau)$ designs the transition matrix associated to matrix A_t^k and is defined as $\phi(t,\tau) = A_t^k \phi^k(t-1,\tau)$.

By sequentially using the expression for $x_t^k, \forall k = 1\cdots d$, it is possible to show that:

$$y_t^d = \sum_{\tau_1=1}^{t}\sum_{\tau_2=1}^{\tau_1}\cdots\sum_{\tau_d=1}^{\tau_{d-1}} C_t \phi^d(t,\tau_1) D_{\tau_1}^d \phi^{d-1}(\tau_1,\tau_2)\cdots D_{\tau_{d-1}}^2 \phi^1(\tau_{d-1},\tau_d) B_{\tau_1} u_{\tau_1} u_{\tau_2}\cdots u_{\tau_d}$$

The transition matrices property allows writing the series kernel as:

$$\begin{aligned} K_d(t,\tau_1,\tau_2,\cdots,\tau_d) = {} & C_t \phi^d(t,1)\left(\phi^d(\tau_1,1)\right)^{-1} D_{\tau_1}^d \phi^{d-1}(\tau_1,1)\left(\phi(\tau_2,1)\right)^{-1} \\ & \cdots D_{\tau_{d-1}}^2 \phi^1(\tau_{d-1},1)\left(\phi^1(\tau_d,1)\right)^{-1} B_{\tau_1} \end{aligned}$$

which is decomposed as a sum of function products in $t,\tau_1,\dots,\tau_d$. ■

It is necessary at this stage to assure a canonical representation of the system, this means the parameter number to be minimal avoiding unpredictability on the realization process.

Proposition 2: The bilinear nilpotent system (2) can be written on the canonical form:

$$\begin{aligned}
x_t^1 &= \tilde{A}_t^1 x_{t-1}^1 + \tilde{B}_t u_t \\
x_t^2 &= \tilde{A}_t^2 x_{t-1}^2 + \tilde{D}_t^2 x_t^1 u_t \\
&\cdots \\
x_t^k &= \tilde{A}_t^d x_{t-1}^d + \tilde{D}_t^d x_t^{d-1} u_t \\
y_t^d &= \tilde{C}_t x_t^d
\end{aligned} \tag{3}$$

with

$$\tilde{B}_t^1 = \begin{bmatrix} b_{n_1,t}^1 & \cdots & b_{1,t}^1 \end{bmatrix}^T$$

$$\tilde{D}_t^K = \begin{bmatrix} d_{n_k,n_{k-1},t}^k & \cdots & \cdots & d_{n_k,1,t}^k \\ \vdots & & & \\ d_{2,n_{k-1},t}^k & \cdots & \cdots & d_{2,1,t}^k \\ 0 & \cdots & 0 & 1 \end{bmatrix}, \forall k = 2 \ldots d$$

$$\tilde{A}_t^K = \begin{bmatrix} 0 & \cdots & \cdots & 0 & a_{n_k,t}^k \\ 1 & \ddots & & \vdots & \vdots \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & a_{1,t}^k \end{bmatrix}, \forall k = 1 \ldots d$$

$$\tilde{C} = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}$$

Proof: In the stationary case, without parameters dependence on time, the *d*-dimensional *z*-transform of the homogenous term y_t^d can be expressed as:

$$\begin{aligned}
H(z_1, \ldots, z_d) = &C\left(z_d I - A^d\right)^{-1} D^d \left(z_{d-1} I - A^{d-1}\right)^{-1} D^{d-1} \\
&\ldots D^2 \left(z_1 I - A^1\right)^{-1} B
\end{aligned}$$

where $z_1, \ldots z_d \in \mathbf{C}$ and I as the identity. Using the Cayley-Hamilton theorem for writing the characteristic A^k matrix relation:

$$\left(A^k\right)^{n_k} = \sum_{i=0}^{n_k-1} a_i \left(A^k\right)^i$$

If we consider the matrix $P^d \in \Re^{n_d \times n_d}$ as:

$$P^d = \begin{bmatrix} C \\ CA^d \\ \vdots \\ C\left(A^d\right)^{n_d-1} \end{bmatrix}$$

we have by construction:

$$P^d A^d = \tilde{A}^d P^d$$

and

$$\tilde{C} = C\left(P^d\right)^{-1}$$

We can write then:

$$\begin{aligned}
C\left(z_d I - A^d\right)^{-1} D^d &= C\left(P^d\right)^{-1} P^d \left(z_d I - A^d\right)^{-1} \left(P^d\right)^{-1} P^d D^d \\
&= \tilde{C}\left(z_d I - \tilde{A}^d\right)^{-1} P^d D^d
\end{aligned}$$

As a result we obtain

$$\begin{aligned}
H(z_1, \ldots, z_d) = &\tilde{C}\left(z_d I - \tilde{A}^d\right)^{-1} P^d D^d \left(z_{d-1} I - A^{d-1}\right)^{-1} D^{d-1} \\
&\ldots D^2 \left(z_1 I - A^1\right)^{-1} B
\end{aligned}$$

Now the matrix decomposition of $\overline{D}^d = P^d D^d$ in rows gives:

$$\overline{D}^d = \begin{bmatrix} D_1^d \\ D_2^d \\ \vdots \\ D_{n_n}^d \end{bmatrix}$$

If we define the matrix $P^{d-1} \in \Re^{n_{d-1} \times n_{d-1}}$ at present as:

$$P^{d-1} = \begin{bmatrix} D_1^d \\ D_1^d A^{d-1} \\ \vdots \\ D_1^d \left(A^{d-1}\right)^{n_{d-1}-1} \end{bmatrix}$$

by construction we have the relation:

$$P^{d-1} A^{d-1} = \tilde{A}^{d-1} P^{d-1}$$

and

$$\begin{aligned}
\overline{D}^d \left(z_{d-1} I - A^{d-1}\right)^{-1} &= \overline{D}^d \left(P^{d-1}\right)^{-1} P^{d-1} \left(z_{d-1} I - A^{d-1}\right)^{-1} \left(P^{d-1}\right)^{-1} P^{d-1} D^{d-1} \\
&= \tilde{D}^d \left(z_{d-1} I - \tilde{A}^{d-1}\right)^{-1} P^{d-1} D^{d-1}
\end{aligned}$$

where the matrix $\tilde{D}^d = \overline{D}^d \left(P^{d-1}\right)^{-1}$ is given by:

$$\tilde{D}^d = \begin{bmatrix} \tilde{C} \\ D_2^d \left(P^{d-1}\right)^{-1} \\ \vdots \\ D_{n_d}^d \left(P^{d-1}\right)^{-1} \end{bmatrix}$$

Finally we have

$$\begin{aligned}
H(z_1, \ldots, z_d) = &\tilde{C}\left(z_d I - \tilde{A}^d\right)^{-1} \tilde{D}^d \left(z_{d-1} I - \tilde{A}^{d-1}\right)^{-1} \overline{D}^{d-1} \\
&\ldots D^2 \left(z_1 I - A^1\right)^{-1} \tilde{B}
\end{aligned}$$

with $\overline{D}^{d-1} = P^{d-1} D^{d-1}$ and $\tilde{B} = P^1 B$. Iterating in this way, we find the given form. ■

At this stage, in order to ensure the linear dependence from the output to the system parameters, we make appear the delayed output from every homogenous subsystem. In the quadratic case we have:

Proposition 3: The quadratic bilinear nilpotent system output can be obtained by a multidimensional NARX as:

$$\begin{aligned}
y_t^1 &= \sum_{i=1}^{n_1} a_{i,t}^1 y_{t-i}^1 + \sum_{i=1}^{n_1} b_{i,t}^1 u_{t-i+1} \\
y_t^2 &= \sum_{i=1}^{n_2} a_{i,t}^2 y_{t-i}^2 + y_t^1 u_t + \sum_{i=2}^{n_2} \sum_{j=1}^{n_1} d_{i,j,t}^2 y_{t-i-j+2}^1 u_{t-i+1}
\end{aligned} \tag{4}$$

Proof: We can have the canonical form in proposition 2 for the quadratic case as:

$$\begin{aligned} x_t^1 &= \widetilde{A}_t^1 x_{t-1}^1 + \widetilde{B}_t u_t, \quad && x_t^1 \in \Re^{n_1} \\ y_t^1 &= \widetilde{C} x_t^1, && y_t^1 \in \Re \\ x_t^2 &= \widetilde{A}_t^2 x_{t-1}^2 + \widetilde{D}_t^2 x_t^1 u_t \\ y_t^2 &= \widetilde{C} x_t^2, && y_t^2 \in \Re \end{aligned}$$

Then taking matrices for d=2 and representing the linear state as a linear output delayed:

$$x_t^1 = \left[y_{t-n_1+1}^1 \cdots y_{t-1}^1 \ y_t^1 \right]^T$$

we obtain equation (4). ■

3. HEREDITARY IDENTIFICATION

3.1. Hereditary algorithm

The parameters estimation by hereditary approach is carried out by considering the identified system as time variant. Thus, at an instant t the predictor in (4) is given by:

$$\begin{aligned} \hat{y}_\tau^{1,t} &= \sum_{i=1}^{n_1} a_{i,t}^1 \hat{y}_{\tau-i}^{1,t-i} + \sum_{i=1}^{n_1} b_{i,t}^1 u_{\tau-i+1} \\ \hat{y}_\tau^{2,t} &= \sum_{i=1}^{n_2} a_{i,t}^2 \hat{y}_{\tau-i}^{2,t-i} + \hat{y}_\tau^{1,t} u_\tau + \sum_{i=2}^{n_2} \sum_{j=1}^{n_1} d_{i,j,t}^2 \hat{y}_{\tau-i-j+2}^{1,t-i-j+2} u_{\tau-i+1} \end{aligned} \tag{5}$$

where $\hat{y}_\tau^{1,t-1}$ and $\hat{y}_\tau^{2,t-1}$, $\forall \tau = 1,\ldots,t-1$ were obtained by minimizing the criterion :

$$J_{t-1} = \mathrm{E}^{t-1}[(y_\bullet - \hat{y}_\bullet^{2,t-1})^2] = \frac{1}{t-1} \sum_{\tau=1}^{t-1} \left(y_\tau - \hat{y}_\tau^{2,t-1} \right)^2$$

The development of the linear part gives us the linear dependence already mentioned:

$$\hat{y}_\tau^{2,t} = \sum_{i=1}^{n_2} a_{i,t}^2 \hat{y}_{\tau-i}^{2,t-i} + \sum_{i=1}^{n_1} a_{i,t}^1 \hat{y}_{\tau-i}^{1,t-i} u_\tau + \sum_{i=1}^{n_1} b_{i,t}^1 u_{\tau-i+1} u_\tau + \sum_{i=2}^{n_2} \sum_{j=1}^{n_1} d_{i,j,t}^2 \hat{y}_{\tau-i-j+2}^{1,t-i-j+2} u_{\tau-i+1}$$

The parameter and regression vectors at time t are defined respectively like:

$$\begin{aligned} \theta_t^T &= \left(a_{i,t}^2 \ \ a_{i,t}^1 \ \ b_{i,t}^1 \ \ d_{i,j,t}^2 \right) \\ \varphi_\tau^{t-1^T} &= \left(\hat{y}_{\tau-i}^{2,t-i}, \ \hat{y}_{\tau-j}^{1,t-j} u_{\tau-k}, \ u_{\tau-l} u_\tau \right) \end{aligned} \tag{6}$$

where indexes i,j,k,l depend from the linear and quadratic systems dimensions. Hence, the predictor (5) can be shortly written:

$$\hat{y}_\tau^{2,t} = \theta_t^T \varphi_\tau^{t-1} \tag{7}$$

The optimization criterion J_t with respect to θ_t gives us the orthogonal relation:

$$\frac{\partial J_t}{\partial \theta_t} = 0 \ \Rightarrow \ \mathrm{E}^t\left[\left(y_\bullet - \hat{y}_\bullet^{2,t} \right) \varphi_\bullet^{t-1} \right] = 0 \tag{8}$$

In fact, the regression vector depends only on the past parameters values and not on θ_t giving in this way a convex minimization problem. Replacing (7) in (8) we obtain:

$$\mathrm{E}^t[\varphi_\bullet^{t-1} \varphi_\bullet^{t-1^T}] \times \theta_t \triangleq \mathrm{E}^t[y_\bullet \varphi_\bullet^{t-1}] \tag{9}$$

Algorithm

1. At time t-1 we have the previous data from optimized trajectories $\hat{y}_\tau^{2,t-i}, \hat{y}_\tau^{1,t-j}, u_\tau, \forall \tau = 1 \ldots t-1$.
2. At instant t the new data u_t, y_t let us actualize the matrix and vector correlations in (9), respectively:
$$\mathrm{E}^t[\varphi_\bullet^{t-1} \varphi_\bullet^{t-1^T}] = \frac{1}{t} \sum_{\tau=1}^{t} \varphi_\tau^{t-1} \varphi_\tau^{t-1^T} \ \& \ \mathrm{E}^t[y_\bullet \varphi_\bullet^{t-1}] = \frac{1}{t} \sum_{\tau=1}^{t} y_\bullet \varphi_\tau^{t-1}$$
3. We obtain the new system parameters values at time t by inversion of the linear system:
$$\theta_t \triangleq \mathrm{E}^t[\varphi_\bullet^{t-1} \varphi_\bullet^{t-1^T}]^{-1} \times \mathrm{E}^t[y_\bullet \varphi_\bullet^{t-1}]$$
4. (Hereditary part). Once the parameters computed, we obtain the new predictors trajectories $\hat{y}_\tau^{1,t}$, $\hat{y}_\tau^{2,t}$ which stand on trajectories $\hat{y}_\tau^{1,t-i}$ and $\hat{y}_\tau^{2,t-j}$ computed at last step by using the system $\forall \tau = 1 \ldots t$:
$$\begin{aligned} \hat{y}_\tau^{1,t} &= \sum_{i=1}^{n_1} a_{i,t}^1 \hat{y}_{\tau-i}^{1,t-i} + \sum_{i=1}^{n_1} b_{i,t}^1 u_{\tau-i+1} \\ \hat{y}_\tau^{2,t} &= \sum_{i=1}^{n_2} a_{i,t}^2 \hat{y}_{\tau-i}^{2,t-i} + \hat{y}_\tau^{1,t} u_\tau + \sum_{i=2}^{n_2} \sum_{j=1}^{n_1} d_{i,j,t}^2 \hat{y}_{\tau-i-j+2}^{1,t-i-j+2} u_{\tau-i+1} \end{aligned}$$
5. Return at 2.

3.2. Application example: nonlinear identification

A finite dimension input/output nonlinear system can be defined:

$$y_t = f\left(y_{t-1}, \ldots, y_{t-n_y}, u_t, \ldots, u_{t-n_u} \right) \tag{10}$$

where $f(\cdot)$ is a given nonlinear function.

We note that the input/output bilinear systems are a particular case of this nonlinear system. Its equation is given by [9] :

$$y_t = \sum_{i=1}^{n_y} a_i y_{t-i} + \sum_{j=0}^{n_u} b_j u_{t-j} + \sum_{i=1}^{n_y} \sum_{j=0}^{n_u} c_{ij} y_{t-i} u_{t-j} \tag{11}$$

Observe that this system is not part of the bilinear nilpotent systems presented here, which produce a finite degree Volterra series. Choosing the system orders in (11) as $n_y = n_u = 1$, we obtain as a particular model:

$$y_t = a_1 y_{t-1} + b_0 u_t + b_1 u_{t-1} + c_{10} y_{t-1} u_t + c_{11} y_{t-1} u_{t-1} \tag{12}$$

Approximating this system by a truncated degree two Volterra series:

$$\hat{y}_t = \sum_{\tau_1=1}^{t} K_1(t, \tau_1) u_{\tau 1} + \sum_{\tau_1=1}^{t} \sum_{\tau_2=1}^{\tau_1} K_2(t, \tau_1, \tau_2) u_{\tau_1} u_{\tau_2} \tag{13}$$

where kernel $K_1(t, \tau_1)$ is approximated by a canonical linear system in addition to kernel $K_2(t, \tau_1, \tau_2)$ approximated by (5).

We have employed for identification as exciting input a uniform white noise [10]. A white Gaussian noise has been added to the output in order to obtain a SNR$\cong$5dB $y_{bruitée}^t = y_t + v_t$. We have chosen the predictor separability orders in (5) to be equal $n_1 = n_2 = n$.

The algorithm quality has been assessed in order to validate after model identification by measuring the criterion:

$$R = \left(1 - \frac{\sum_{t=1}^{N} |y_t - \hat{y}_t|^2}{\sum_{t=1}^{N} |y_t|^2}\right) \times 100$$

where R is the output part y_t which is rightly explained by $\hat{y}_t$ in (12). It is called *multiple correlation coefficient* (squared) and is often given in percent [11].

The particular bilinear model (12)has been approximated initially without additive noise in order to measure our algorithm efficiency (see Fig. 1). By using (13) with dimension n=3, we were able to achieve a coefficient of R=89.25%.

Comparison between hereditary and subspace algorithms is shown in Fig. 2. The model used for subspace identification was also bilinear nilpotent without defined structure for the system matrices [12]. We can see from the plot that the subspace approach fits less well the nonlinear system output mainly at high and low peaks in comparison to the hereditary approach.

Table 1.

Nonlinear model identification			
SNR 5dB	Identification and Validation samples	R (%)	Dimension Lin+Quadr
Hereditary	50 / 50 (%)	87.15	3+3
Subspace		78.75	4+4

Table 1 shows the employed dimension and approximation result R for both identification techniques. It is observed that the hereditary algorithm performs better of about 10% with smaller dimension compared to the subspace algorithm for the polynomial approach.

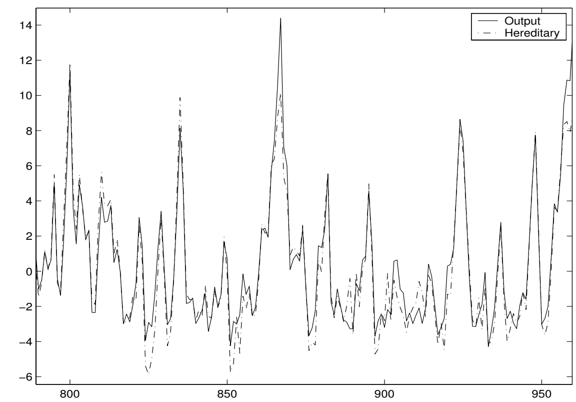

Fig.1: Noise free output approximation of (12) by (13). Identification/validation (50% of samples each task).

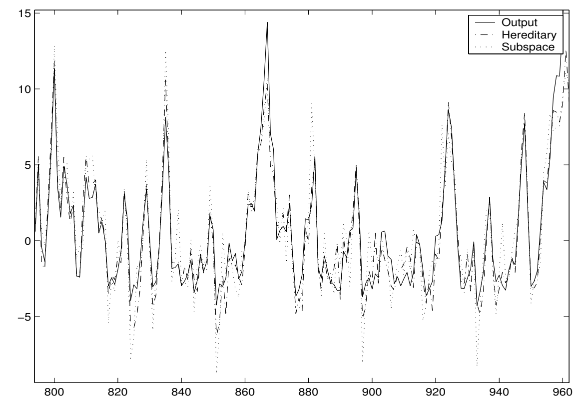

Fig.2: Noisy output approximation of (12) by (13) and comparison. Identification/validation (50% of samples each task).

4. CONCLUSION

The main contributions of this work are:

- The use of the bilinear nilpotent model in order to approach a nonlinear system by finite degree Volterra series realization.
- The canonical derivation of a structure for better conditioning of the problem by reducing the number of parameters estimated.
- The implementation of the hereditary algorithm as a rugged tool without employing nonlinear optimization techniques for system identification as in other cases [13].

5. ACKNOWLEDGMENTS

The fist author would like to thank the mexican council of science and technology (CONACyT) for his support.

6. REFERENCES

[1] C. Lesiak and A.J. Krener, "The Existence and Uniqueness of Volterra Series for Nonlinear Systems," *IEEE Trans. Automat., Contr.,* vol. AC-23, no. 6, pp.1090-1095, Dec. 1978.

[2] J.A. Cherry and W.M. Snelgrove, "On the Characterization and Reduction of Distortion in Bandpass Filters," *IEEE Trans. Circuits Sys.-I ,* vol. 45, no. 5, pp. 523-537 May 1998.

[3] N. Le Gallou, E. Ngoya, et al., "An Improved Behavioral Modeling Technique for High Power Amplifiers with Memory," *IEEE MTT-S Digest*, vol. 2, pp. 983-986, May 2001.

[4] P. D'Alessandro, A. Isidori, A. Ruberti, "Realization and Structure Theory of Bilinear Dynamical Systems," *SIAM J. Control*, vol. 12, no. 3, pp. 517-535, Aug. 1974.

[5] A.J. Krener, "Linearization and Bilinearization of Control Systems," *Proc. 1974 Allerton Conf. on Circuit and System Theory*, pp. 834-843, 1974.

[6] R.W. Brockett, "Volterra Series and Geometric Control Theory," *Automatica*, vol. 12, pp. 167-176, 1976.

[7] A. Monin and G. Salut, "ARMA Lattice Identification: A New Hereditary Algorithm," *IEEE Trans. on Signal Processing*, vol. 44, no. 2, pp. 360-370, Feb. 1996.

[8] P.E. Crouch, "Dynamical Realizations of Finite Volterra Series," *SIAM J. Control*, vol. 19, no. 2, pp. 177-202, March 1981.

[9] H. Ki Baik and V.J. Mathews, "Adaptive Lattice Bilinear Filters," *IEEE Trans. Signal Processing*, vol. 41, no. 6, pp. 2033-2046, June 1993.

[10] I.J. Leonartis and S.A. Billings, "Experimental Design and Identifiability for Non-linear Systems," *Int. J. Systems Sci.*, vol. 18, no. 1, pp. 189-202, 1987.

[11] L. Ljung, "System Identification: Theory for the User," 2nd Edition, Prentice-Hall, 1999.

[12] W. Suleiman, G. Etcheverry, A. Monin, "An Approximate Identification of Nonlinear System: Nilpotent Approach," *14th IFAC Symposium on System Identification, SYSID 2006*, submitted.

[13] S. Chen and S.A. Billings, "Recursive Prediction Error Parameter Estimator for Non-linear Models," *Int. J. Control*, vol. 49, no.2, pp. 569-594, 1989.