33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Recursive Least Squares Estimation of Nonlinear Multiple-Input Systems using Orthonormal Function Expansions

Georgios D. Mitsis, *Senior Member, IEEE* and Marios M. Markou

***Abstract*— We present a computational scheme to obtain adaptive non-linear, multiple-input models of the Volterra-Wiener class, by utilizing function expansions of the Volterra kernels in a recursive least-squares formulation. Function expansions have been proven successful in linear and nonlinear systems identification as they result in a significant reduction of the required free parameters, which is a major limiting factor particularly for nonlinear systems, whereby this number depends exponentially on the nonlinear system order. We illustrate the performance of the proposed scheme by presenting results for a simulated linear two-input system with time-varying characteristics.**

Index Terms: Nonlinear systems identification, Volterra models, Time-Varying Systems, Laguerre functions.

I. INTRODUCTION

The study of many physiological systems has been pursued in the context of Volterra-Wiener models, as these systems are characterized by considerable complexity and often exhibit nonlinear behavior [1]. Typically, physiological systems are affected by a number of physiological factors, some of which may be non-observable; moreover, they are also influenced by physiological mechanisms and rhythms that exert their effects over widely varying time scales (e.g., local mechanisms may act within seconds, whereas diurnal or circadian rhythms have a cycle on the order of several hours). Consequently, these systems may exhibit nonstationary characteristics as well. Therefore, the ability to quantify these nonstationarities accurately is important in the context of physiological systems modeling, particularly in real-time applications. Among many methods proposed for nonlinear systems identification, expansion of the Volterra kernels in terms of an orthonormal basis, such as the Laguerre basis, has been proven successful as it reduces the required number of free parameters considerably [2]. In the present paper, we propose a recursive least-squares scheme to estimate multiple-input Volterra models using Laguerre function expansions. We also evaluate its performance by a simulated example.

II. METHODS

A. Multiple-input Volterra models

The output of a non-linear multiple input single output system can be expressed in terms of a series of functionals [1] that represent higher order convolutions with the input signals. In discrete time we have:

$$\begin{aligned} y(n) &= \sum_{q=0}^{Q} \sum_{i_1,\ldots,i_q}^{I} \sum_{m_1,\ldots,m_q} k_q^{(x_{i_1}\ldots x_{i_q})}(m_1,\ldots,m_q) \\ &\quad x_{i_1}(n-m_1)\ldots x_{i_q}(n-m_q) \\ &= k_0 + \sum_{i=1}^{I}\sum_{m} k_1^{(x_i)}(m)x_i(n-m) \\ &+ \sum_{i_1,i_2=1}^{I}\sum_{m_1,m_2} k_2^{(x_{i_1}x_{i_2})}(m_1,m_2)x_{i_1}(n-m_1)x_{i_2}(n-m_2) \\ &+ \ldots \end{aligned} \tag{1}$$

where I is the number of inputs, x_i are the system inputs and $y(n)$ is the system output. $k_q^{(x_{i_1}\ldots x_{i_q})}$ denotes the q-th order Volterra kernel of the system that corresponds to inputs $(x_{i_1}\ldots x_{i_q})$. The Volterra kernels describe the linear ($q=1$) and nonlinear ($q>1$) dynamic effects of the inputs (and their interactions) on the output. Specifically, for $q=1$ there exist I linear kernels in 1 that correspond to each input, whereas for $q>1$ the terms $k_q^{(x_{i_1}\ldots x_{i_q})}$ correspond to the nonlinear self-kernels of each input for $(x_{i_1}=\ldots=x_{i_q})$ and to the nonlinear cross-kernels between different inputs if some of $(x_{i_1},\ldots,x_{i_q})$ are different. The sum of eq. (1) can be viewed as a generaliztion of the convolution sum, with the Volterra kernels quantifying the effect of past input values (linear kernels), as well as their q-th order products (nonlinear self- and cross-kernels) on the output at present time n. For causal and finite memory systems the sums in 1 are defined for $m_i=0$ to M, where M is the system memory.

B. Orthonormal function expansions of Volterra kernels

Direct estimation of the model of 1 from input-output measurements requires the estimation of a number of parameters that depends exponentially on the nonlinear system order Q, i.e. this number is analogous to M^Q. This may result in a very large number of free parameters if M and/or Q are large. One way to reduce it is to express the Volterra kernels in terms of an orthonormal discrete-time basis as shown below for the first and second-order kernels:

$$\begin{aligned} k_1^{(x_i)}(m) &= \sum_{j_i=0}^{L_{x_i}} c_j^{(x_i)} b_j^{(x_i)}(m) \qquad i=1,2 \\ k_2^{(x_{i_1}x_{i_2})}(m_1,m_2) &= \sum_{j_1=0}^{L_{x_{i_1}}}\sum_{j_2=0}^{L_{x_{i_2}}} c_{j_1 j_2}^{(x_{i_1}x_{i_2})} b_{j_1}^{(x_{i_1})}(m_1) b_{j_2}^{(x_{i_2})}(m_2) \end{aligned} \tag{2}$$

G.D. Mitsis and M.M. Markou are with the Department of Electrical and Computer Engineering, University of Cyprus, Nicosia 1678, Cyprus (e-mails: gmitsis@ucy.ac.cy, mmarkou@ucy.ac.cy)

where $i_1, i_2 = 1, ..., I$ and $b_j^{(x_i)}(m)$ is the j-th order discrete-time basis function corresponding to input i.

One possible choice for the basis set, which has been widely used, is the discrete-time Laguerre basis [2], as these functions form a complete orthonormal basis in $[0\ \infty)$, making them suitable for causal finite memory systems. In this case the j-th order basis function is given by:

$$b_j^{(x_i)}(m) = \alpha_i^{(m-j)/2}(1-\alpha_i)^{1/2}\sum_{k=0}^{j}(-1)^k\binom{m}{k}\binom{j}{k}\alpha_i^{j-k}(1-\alpha_i)^k \tag{3}$$

This method requires the determination of two parameters for each input: L_{x_i} and α_i where L_{x_i} is the maximum function order required for input i and α_i $(0 < \alpha_i < 1)$ determines the rate of exponential decay of the corresponding Laguerre functions. Larger values of α_i result in slower decay and are thus more suitable for system inputs that exert their effects with large memory and/or slow dynamics. Determination of these two parameters is discussed below.

By combining equations (1) and (2) we can write:

$$\mathbf{y} = \mathbf{V}\mathbf{c} + \epsilon, \tag{4}$$

where $\mathbf{y}$ is the $(N \times 1)$ vector of output observations, $\mathbf{V}$ is a $(N \times d)$ matrix containing the convolution of both inputs with the Laguerre functions $v_j^{(x_i)} = x_i * b_j^{(x_i)}$ (linear models), as well as higher-order products between them $v_{j_1}^{(x_{i1})}v_{j_2}^{(x_{i2})}\ldots v_{j_Q}^{(x_{iQ})}$ (nonlinear models), including self- and cross-terms (whereby $i_1 = i_2 = \cdots = i_Q$ and some of $i_1, i_2, \ldots, i_Q$ are different respectively), and $\mathbf{c}$ is the $(d \times 1)$ vector of the unknown expansion coefficients. For a two-input system, the number of free parameters d is equal to $L + 1$ for $Q = 1$ and $L_1L_2 + (L + L(L+1))/2$ for $Q = 2$,where $L = L_1 + L_2$, due to the symmetry of the second-order self-kernels with respect to $j_1, j_2, \ldots, j_Q$.

The values of $v_j^{(x_i)}$ can be obtained by the following recursive relations [3]:

$$v_j^{(x_i)}(n) = \sqrt{\alpha_i}(v_j^{(x_i)}(n-1) + v_{j-1}^{(x_i)}(n)) - v_{j-1}^{(x_i)}(n-1) \tag{5}$$

$$v_0^{(x_i)}(n) = \sqrt{\alpha_i}v_0^{(x_i)}(n-1) + \sqrt{1-\alpha_i}x_i(n) \tag{6}$$

The least-squares estimate of $\mathbf{c}$ is given by:

$$\hat{\mathbf{c}}_{LS} = (\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T\mathbf{y}. \tag{7}$$

C. Recursive estimation of the Laguerre expansion model

In order to obtain adaptive estimates of the Volterra kernels we formulate eq. (7) using recursive least squares. Initially, we rewrite the least-squares cost function at time point n as [4]:

$$J(n) = \sum_{s=1}^{N}\lambda^{n-s}e^2(s) \tag{8}$$

where $e(s)$ are the residuals at time point s and λ is the forgetting factor $(0 < \lambda \leq 1)$, which determines the weight of the previous time points on the present estimates. A smaller value of λ corresponds to models that adapt more quickly, while larger values of λ yield results that are closer to regular least-squares. The update equations for the coefficient vectors at time point n are then written as:

$$\hat{\mathbf{c}}(n) = \hat{\mathbf{c}}(n-1) + \mathbf{K}(n)\epsilon(n) \tag{9}$$

$$\epsilon(n) = y(n) - \mathbf{v}^T(n)\hat{\mathbf{c}}(n-1) \tag{10}$$

$$\mathbf{P}(n) = \frac{1}{\lambda}\left[\mathbf{P}(n-1) - \frac{\mathbf{P}(n-1)\mathbf{v}(n)\mathbf{v}^T(n)\mathbf{P}(n-1)}{\lambda + \mathbf{v}(n)^T\mathbf{P}(n-1)\mathbf{v}(n)}\right] \tag{11}$$

$$\mathbf{K}(n) = \frac{\mathbf{P}(n-1)\mathbf{v}(n)}{\lambda + \mathbf{v}^T(n)\mathbf{P}(n-1)\mathbf{v}(n)} \tag{12}$$

where $\mathbf{v}(n)$ are estimated from eqs. (5) - (6), $\mathbf{K}(n)$ is a gain matrix that determines the update size for $\hat{\mathbf{c}}(n)$ and $\mathbf{P}(n)$ is the estimate of the coefficient covariance matrix. The initial value for this matrix is typically selected as $\mathbf{P}(0) = \rho\mathbf{I}$. The effect of the value of λ on the results is examined below.

D. Model order selection

In order to select the model complexity (i.e., the value of L_{x_i} for each input), we propose a heuristic scheme that is based on the Bayesian information criterion (BIC). The BIC determines the optimal model structure as the one that minimizes:

$$BIC = \left(2 * \log\left[\frac{\sum_1^N(e(n))^2}{N-d-1}\right]\right) + (d * \log(N)) \tag{13}$$

where $e(n)$ are the residuals between the output measurements and the model output prediction and d,N were defined above as the number of model free parameters and number of observations respectively. Since the recursive estimation above (eqs. (9)-(12)) requires the use of a fixed structure for $\mathbf{V}$ the values of α_i and L_{x_i} may be determined as follows. First, the entire data sets is segmented using sliding windows with a length that is adequate to capture the slower system dynamics. For each of the sliding windows regular least squares may be used to select the optimal value of α_i as the value that minimizes the normalized mean square error of the model prediction for a range of L_{x_i} values. Consequently, for these optimal α_i values, the values of L_{x_i} may be selected as those that minimize the BIC criterion. For the recursive scheme, the median model orders that result from this procedure as the most representative for the entire data set are then selected.

E. Simulated system

In order to evaluate the method, we created a linear, two-input simulated system of the same class, i.e. its output is given by 4, with time-varying coefficients. Its structural parameters were as follows: $L_1 = 7$, $L_2 = 5$, $\alpha_1 =$

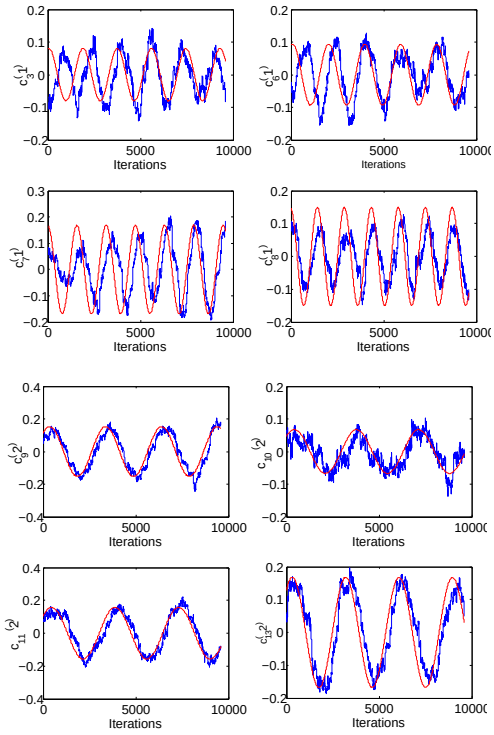

Fig. 1. Representative estimated coefficients (blue) versus true coefficients (red) for $\lambda = 0.993$ for both inputs (left panel: input 1, right panel: input 2). Note that the estimated coefficients track the true coefficients well.

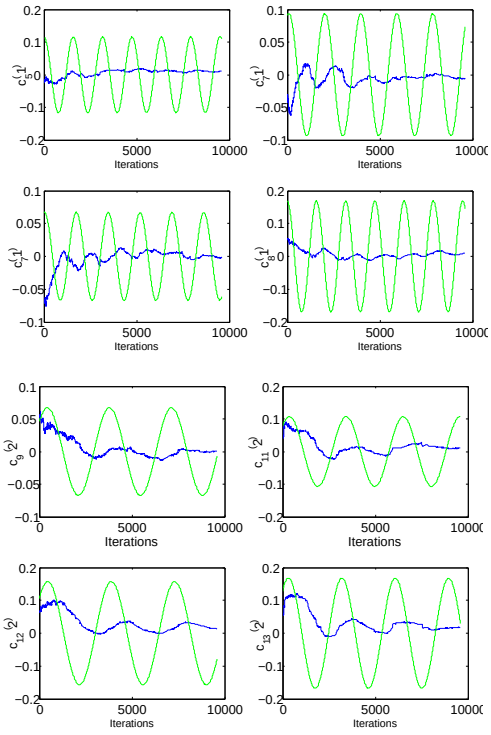

Fig. 2. Representative estimated coefficients (blue) versus true coefficients (red) for $\lambda = 1$ for both inputs (left panel: input 1, right panel: input 2). Note that the estimated coefficients converge to small values close to the mean value of the true, time-varying coefficients

0.55, and $\alpha_2 = 0.75$. The true coefficients were assumed to vary in a sinusoidal manner with different frequencies for each input. The inputs were zero mean independent Gaussian White Noise signals with a length of 10000 points, while independent zero mean white noise ϵ was added to the output for a signal-to-noise ratio (SNR) equal to 10 dB. The results obtained for this system are presented in the next section.

III. RESULTS

The system structure in this case is known, therefore we selected the same structure for our model in order to facilitate comparisons. However, note that the model order selection procedure outlined above yields values that are close to the true ones and the results are not affected much. The results for the estimated coefficients are shown in Figs. (1) - (2) for two different values of λ, i.e. $\lambda = 0.993 and \lambda = 1$. The latter case corresponds to regular least squares, whereby all the previous input values are taken into account in the cost function with equal weights. We can see that for $\lambda = 0.993$ the estimated coefficients track the real coefficients well for both inputs - we show four representative coefficients for each input. Note that the true coefficients vary with different frequencies for the two inputs. On the other hand, for $\lambda = 1$ the coefficients gradually converge to values that are close to zero, i.e. to the mean value of the true coefficients. Therefore, the results obtained are not correct.

This is illustrated by computing the normalized mean-squared error (NSME) of the estimated time-varying kernels relative to the true time-varying kernels, defined as the square of the ratio between the second norm of the difference vector between them to the second norm of the true kernel vector. In both cases, the time-varying kernels are given by eq. (2) but they are updated every time step according to the (true or estimated values) of the coefficient vector $\hat{\mathbf{c}}(n)$. The estimated linear time-varying kernels are shown in Figs. (3)-(4) for both both inputs for $\lambda = 0.993$ and $\lambda = 1$ respectively, where it can be seen that in the former case the kernels are tracked with a frequency that corresponds to the oscillation frequency of the true coefficients (Figs. (1)-(2), whereas in the second case they gradually converge to a "mean kernel". The corresponding NMSE values for both values of λ are shown in Fig. (5). It is evident that the recursive scheme with $\lambda = 0.993$ provides much better estimates of the true time-varying kernels. Note that when the number of observation increases the recursive scheme with $\lambda = 1$ converges to the least-squares estimate that would be obtained using all the input-output observations (hence in this case the recursive scheme is equivalent to regular least-squares as mentioned above).

IV. DISCUSSION

We have presented a recursive least-squares scheme for estimating nonlinear, multiple-input models, based on Laguerre expansions of Volterra kernels and illustrated its performance on a simulated system. One of the main issues in nonlinear systems identification is the number of the required free

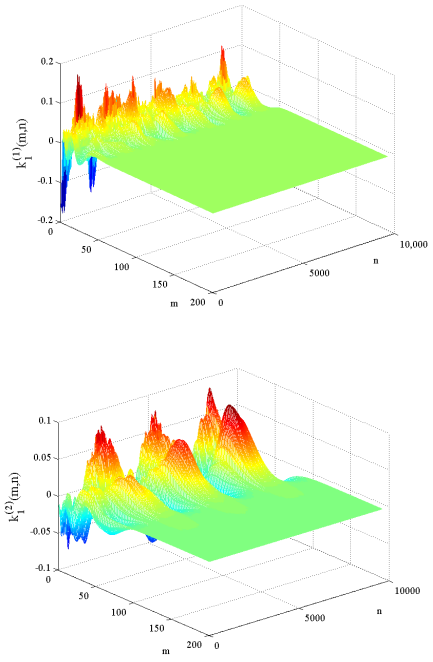

Fig. 3. Estimated time-varying linear kernels ($k_1^{(x_i)}$, $i = 1, 2$ in eq. (1)) for both inputs (left panel: input 1, right panel: input 2) as a function of time for $\lambda = 0.993$.

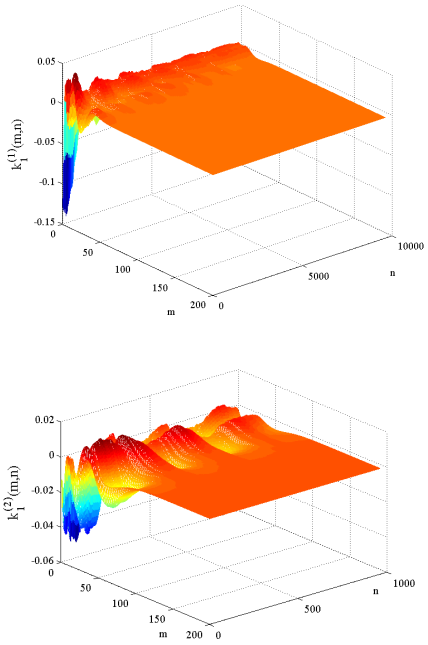

Fig. 4. Estimated time-varying linear kernels ($k_1^{(x_i)}$, $i = 1, 2$ in eq. (1)) for both inputs (top panel: input 1, bottom panel: input 2) as a function of time for $\lambda = 1$.

parameters. In case of a Q-th order system with memory M, the number of free parameters in the standard least-squares formulation of the Volterra model is equal to M^Q. Utilizing function expansions reduces this number to L^Q, with the difference being more pronounced for nonlinear systems, since typically $L << M$. Recursive least-squares require a fixed model structure; in order to determine the most representative structure we propose a scheme utilizing the BIC to select the model structure.

Obtaining accurate estimates of nonstationarities in physiological systems is particularly important, as these systems exhibit a high degree of complexity and are affected by many physiological factors that exert their effects over time scales that may be widely different. Therefore, in real-time applications it is important to track these nonstationarities in a reliable manner. In this context, selection of the parameter λ is important as it determines the adaptive properties of the estimation algorithm. As the value of λ reduces, the effective memory of the estimation algorithm reduces as well, and the estimates are affected relatively more by the immediately preceding input values. Therefore, smaller values of λ are more suitable for systems that are rapidly varying and/or have small memory, whereas larger values are suitable for slowly varying systems and systems with large memory.We aim to examine the selection of λ in a more systematic manner; however an empirical rule is that it should correspond to the memory of each system input. For instance for relatively small values of λ, the effective memory of the estimation is small (for example $0.95^{20} = 0.36$), which implies that

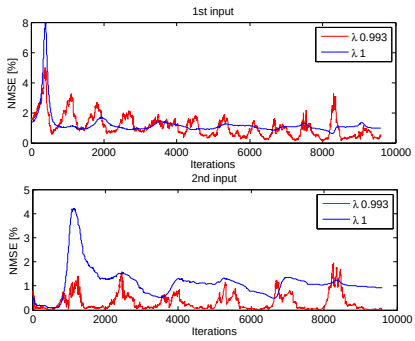

Fig. 5. Normalized mean squared error between true and estimated time-varying kernels for both inputs (top panel: input 1, bottom panel: input 2) for both values of λ considered. The estimates are much better for $\lambda = 0.993$

a small number of points affect the current estimates and that such a value should be used only for systems with short memory. The application of the proposed method to cebebral hemodynamics is currently underway and will be reported in subsequent studies.

REFERENCES

[1] V. Marmarelis, Nonlinear Dynamic Modeling of Physiological Systems. Piscataway, NJ: Wiley-Interscience & IEEE Press, 2004.

[2] ——, "Identification of nonlinear biological systems using laguerre expansions of kernels," Annals of Biomedical Engineering, vol. 21, pp. 573–589, 1993.

[3] H. Ogura, "Estimation of wiener kernels of a nonlinear system and a fast algorithm using digital laguerre filters," in Proc. 15th NIBB Conference, Okazaki, Japan, 1985, pp. 14–62.

[4] L. Ljung, System identification: Theory for the user. Upper Saddle River, NJ: Prentice Hall, 1999.