31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Nonlinear Modeling and Identification of Stretch Reflex Dynamics Using Support Vector Machines

Mujahed Al Dhaifallah and David T. Westwick

Abstract— **In this work, an algorithm that identifies Hammerstein models with support vector machine nonlinearities and output-error linear dynamics is proposed. This algorithm is used to identify a Hammerstein model of stretch reflex EMG dynamics from experimental data.**

I. INTRODUCTION

System identification techniques have been used widely to find mathematical descriptions for physiological systems from measured input/output data. Since these systems often contain hard nonlinearities, block structured models, cascades of static nonlinearities and dynamic linear systems, can be used to represent them [1]. The main advantage of such models over other nonlinear models is that they retain much of the simplicity of linear models, but can nevertheless be used to approximate many nonlinear systems very accurately. The simplest of these is the Hammerstein cascade: a memoryless nonlinearity followed by a dynamic linear element.

The stretch reflex is the involuntary contraction of a muscle in response to a perturbation of its length. In the case of the ankle, it can be treated as the dynamic relationship between the angular velocity of the ankle and the resulting electromyogram (EMG), measured over the Gactrocnemius-Soleus (GS) [2]. Kearney and Hunter [3] suggested a Hammerstein structure to model such dynamics and showed that the static nonlinearity resembles a half-wave rectifier. Afterward, Westwick and Kearney [1] used polynomials to represent the nonlinearity because they are computationally easy to use. Nevertheless, they are not suitable to fit hard nonlinearities. So, Dempsey and Westwick [4] considered cubic splines, which can represent nonlinearities containing hard and smooth curves, as the nonlinearity in the Hammerstein cascade. However, cubic spline functions are defined by a series of knot points which must either be chosen a-priori, or treated as model parameters and included in the (non-convex) optimization.

Many algorithms have been proposed to identify Hammerstein systems. Although the earliest algorithms assumed an output error model nonlinear structure [5], [6], most recent algorithms assume that the systems under consideration have either FIR [1], [4] or ARX [7], [8] linear elements, as these models are linear in their parameters, which simplifies their identification. This restricts the plant and noise transfer functions have common denominators. From a physical point of view it may seem more natural to parametrize these transfer functions independently. In which case, an output error (OE) model would be appropriate. Several techniques have been suggested to deal with linear OE models; instrumental variables, repeated least squares, subspace methods, and the Steiglitz-McBride (S-M) algorithm [9].

Recently, support vector machines (SVMs) and least squares support vector machines (LS-SVMs) have shown powerful abilities in approximating linear and nonlinear functions [10], [11]. They provide much greater flexibility in modeling nonlinearities than is possible with a fixed basis expansion. The SVM has additional advantages over the LS-SVM: sparseness of the solution and robustness to outliers, but requires increased computational effort. Both SVMs and LS-SVMs are fit by solving convex optimization problems, and neither requires *a-priori* structural information [10].

The main contribution of this paper is to extend the S-M iterative algorithm to the identification of the output error Hammerstein models with SVM nonlinearities. This algorithm will be demonstrated by using it to identify a model of the stretch reflex EMG dynamics model from experimental data.

II. IDENTIFICATION OF OUTPUT ERROR MODEL

The Hammerstein output-error system consists of a nonlinear memoryless element followed by a linear output-error model. The linear dynamics are represented by an output-error model :

$$\left\{ \begin{array}{c} s_t = \dfrac{B(z)}{A(z)} x_t = \dfrac{\sum_{j=0}^{m} b_j z^{-j}}{1+\sum_{i=1}^{n} a_i z^{-i}} x_t \\ y_t = s_t + e_t = \dfrac{B(z)}{A(z)} x_t + e_t \end{array} \right. \qquad (1)$$

where $x_t,\ s_t,\ y_t \in \mathbb{R}$, are the output of the nonlinear block, the unmeasurable noise-free output, and the measured output signals, respectively, for $t = 1 \dots N$. The innovation e_t is assumed to be white and z^{-1} is a shift operator $\left[z^{-1} y_t = y_{t-1}\right]$. The static nonlinearity is assumed to have the following form:

$$x_t = f(u_t) = \mathbf{w}^T \varphi(u_t) + d_0 \qquad (2)$$

where $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^{n_H}$ denotes a mapping to high dimensional feature space which can be infinite dimensional, $\mathbf{w}$ is a vector of weights in this feature space, and d_0 represents the bias term. Then the Hammerstein nonlinear output-error model

Mujahed Al Dhaifallah and David Westwick are with the Department of Electrical and Computer Engineering, Schulich School of Engineering, University of Calgary, Calgary, Canada [maldhaif,dwestwic]@ucalgary.ca

978-1-4244-3296-7/09/$25.00 ©2009 IEEE

may be expressed as

$$\begin{cases} s_t = \frac{B(z)}{A(z)}\left(\mathbf{w}^T\varphi(u_t)+d_0\right) \\ y_t = s_t + e_t = \frac{B(z)}{A(z)}\left(\mathbf{w}^T\varphi(u_t)+d_0\right)+e(t) \end{cases} \quad (3)$$

Note that y_t is nonlinear function of the parameters b_j, $\mathbf{w}$, d_0, and a_i, which makes the identification problem difficult to solve. Thus, a related overparameterized problem is proposed by rewriting First, an overparameterized model is (3) as

$$y_t = \frac{W(z)}{A(z)}\varphi(u_t) + d + e(t) \quad (4)$$

where

$$\begin{aligned} W(z) &= \left(\mathbf{w}_0^T + \mathbf{w}_1^T z^{-1} + \mathbf{w}_2^T z^{-2} + \cdots + \mathbf{w}_m^T z^{-m}\right) \\ \mathbf{w}_j &= b_j\mathbf{w},\ d = d_0\frac{\sum_{j=0}^m b_j}{1+\sum_{i=1}^n a_i} \end{aligned}$$

It is clear that y_t is linear in the parameters of the overparameterized numerator in (4), but this model class is more general than the Hammerstein model, which it includes as a special case (when $\mathbf{w}_j = b_j\mathbf{w}$ for $j = 1..m$). The strategy will be to identify this model first, and then use a low-rank projection to force the estimated model to be a Hammerstein cascade [7], [8]. To identify the parameters of model (3) using SVM regression, solve the following optimization problem

$$\min_{\mathbf{w}_j,\boldsymbol{a},d,\boldsymbol{\xi}} \frac{1}{2}\sum_{j=0}^{m}\mathbf{w}_j^T\mathbf{w}_j + \frac{1}{2}\sum_{i=1}^{n}a_i^2 + c\sum_{t=r}^{N}(\xi_t+\xi_t^*) \quad (5)$$

subject to

$$\sum_{t=1}^{N}\mathbf{w}_j^T\varphi(u_t) = 0, \quad j = 0,\ldots,m \quad (6)$$

$$\begin{aligned} & y(t) - \frac{W(z)}{A(z)}\varphi(u_t) + d \le \varepsilon + \xi_t \\ & \frac{W(z)}{A(z)}\varphi(u_t) + d - y(t) \le \varepsilon + \xi_t^* \end{aligned} \quad (7)$$

$$\xi_t, \xi_t^* \ge 0, \quad t = r,\ldots,N \quad (8)$$

Note that (5) is a standard SVM objective function, consisting of a weighted average, with the weighting controlled by the parameter c, of the 2 norm of the parameters ($\mathbf{w}$ and $\boldsymbol{a}$) and the Vapnik ε-insensitive cost function applied to the residuals. The constraints in (7) are derived by modifying the constraints of the standard SVM to include the dynamics of the output error model. From (7), it is evident that errors smaller than ε, a user selected tuning parameter, do not contribute to the cost function. Constraints (6) were added to center the nonlinear functions $\mathbf{w}_j^T\varphi(\cdot)$, $j = 0,\ldots,m$ around their average over the training set [7], [8]. Unfortunately, the optimization problem (5)-(8) is not a standard quadratic programming problem as is the case when the linear dynamics are represented by an ARX model [8]. However, one can carry out the minimization (5)-(8) iteratively. The idea is to fit a NARX Hammerstein model for the data using the algorithm proposed in [8] as initial starting point. The result is a "first estimate" of $A(z)$, $B(z)$, $f(u_t)$, denoted $A^1(z)$, $B^1(z)$, $f^1(u_t)$. Then a S-M inspired iteration is used. Thus at iteration l, the previous denominator $A^{l-1}(z)$ is used to pre-filter the input and outputs of the linear part. Then $A^l(z)$,$W^l(z)$ are found by solving the following minimization

$$\min_{\mathbf{w}_j,\boldsymbol{a}^l,d,\boldsymbol{\xi}} \frac{1}{2}\sum_{j=0}^{m}\mathbf{w}_j^{l^T}\mathbf{w}_j^l + \frac{1}{2}\sum_{i=1}^{n}\left(a_i^l\right)^2 + c\sum_{t=r}^{N}(\xi_t+\xi_t^*) \quad (9)$$

subject to

$$\sum_{t=1}^{N}\mathbf{w}_j^{l^T}\varphi(u_t) = 0, \quad j = 0,\ldots,m \quad (10)$$

$$\begin{aligned} & A^l(z)\,y_f(t) - \frac{W^l(z)}{A^{l-1}(z)}\varphi(u_t) - d^l \le \varepsilon + \xi_t \\ & \frac{W^l(z)}{A^{l-1}(z)}\varphi(u_t) + d^l - A^l(z)\,y_f(t) \le \varepsilon + \xi_t^* \end{aligned} \quad (11)$$

$$\xi_t, \xi_t^* \ge 0, \quad t = r,\ldots,N \quad (12)$$

where $y_f(t) = \frac{y_t}{A^{l-1}(z)}$. Note that as $A^{l-1}(z)$ approaches $A^l(z)$ problem (9)-(12) approaches the optimization problem (5)-(8). Problem (9)-(12) can be solved as follows. First, write the Lagrangian of (9)-(12)

$$\begin{aligned} L\left(\mathbf{w}_j^l, d^l, \boldsymbol{\xi}, \boldsymbol{\xi}^*, \boldsymbol{a}^l; \boldsymbol{\alpha}, \boldsymbol{\alpha}^*, \boldsymbol{\beta}, \boldsymbol{\beta}^*, \boldsymbol{\gamma}\right) &= \tfrac{1}{2}\sum_{j=0}^{m}\mathbf{w}_j^{l^T}\mathbf{w}_j^l \\ &+ \tfrac{1}{2}\sum_{i=1}^{n}\left(a_i^l\right)^2 + c\sum_{t=r}^{N}(\xi_t+\xi_t^*) \\ &- \sum_{j=0}^{m}\gamma_j\left(\sum_{t_1=1}^{N}\mathbf{w}_j^{l^T}\varphi(u_{t_1})\right) - \sum_{t=r}^{N}\alpha_t\Bigg(\frac{W^l(z)}{A^{l-1}(z)}\varphi(u_t) \\ &+ d^l - A^l(z)\,y_f(t) + \varepsilon + \xi_t\Bigg) \\ &- \sum_{t=r}^{N}\alpha_t^*\Bigg(A^l(z)\,y_f(t) - \frac{W^l(z)}{A^{l-1}(z)}\varphi(u_t) - \\ &d^l + \varepsilon + \xi_t^*\Bigg) - \sum_{t=r}^{N}(\beta_t\xi_t + \beta_t^*\xi_t^*) \end{aligned} \quad (13)$$

where α_i, α_i^*, β_i, β_i^* are non-negative Lagrange multipliers and $\gamma_j \in \mathbb{R}$. Then, the dual problem can be formulated by finding the stationary point of the Lagrangian (13). Setting $\frac{\partial L}{\partial \mathbf{w}_j}$ to zero yields

$$\mathbf{w}_j^l = \gamma_j\sum_{t=1}^{N}\varphi(u_t) + \sum_{t=r}^{N}(\alpha_t - \alpha_t^*)\left(\frac{1}{A^{l-1}(z)}\varphi(u_{t-j})\right) \quad (14)$$

One of the key concepts in SVM regression, is the so-called kernel trick, whereby inner-products of the nonlinear basis functions are replaced with a kernel [10]. Thus, let $K(u_i,u_j) = \varphi(u_i)^T\varphi(u_j)$ be the kernel function, and let K be $N \times N$ matrix whose entries are $K_{i,j} = K(u_i,u_j)$. From (14),

$$\begin{aligned}
\mathbf{w}_j^{l^T}\varphi\left(u_{t_k}\right) &= \gamma_j\sum_{t_1=1}^{N}\varphi^T\left(u_{t_1}\right)\varphi\left(u_{t_k}\right)+\sum_{t=r}^{N}\left(\alpha_t-\alpha_t^*\right)\\
&\quad\times\left(\frac{1}{A^{l-1}(z)}\varphi^T\left(u_{t-j}\right)\right)\varphi\left(u_{t_k}\right)\\
&= \gamma_j\sum_{t_1=1}^{N}K\left(u_{t_1},u_{t_k}\right)\\
&\quad+\sum_{t_1=r}^{N}\left(\alpha_{t_1}-\alpha_{t_1}^*\right)K_{f,0}\left(u_{t_{1-j}},u_{t_k}\right)
\end{aligned} \tag{15}$$

where $K_{f,0}$ is the result of filtering each column of the matrix K with the previous denominator, $\frac{1}{A^{l-1}(z)}$. Similarly, define $K_{0,f}$ and $K_{f,f}$ as the results obtained by filtering the rows, and rows and columns, of K, respectively. Note that the second equation in (15) involves the kernel, but does not use the nonlinear basis functions, φ.

Similarly from the centering constraints (10), one can show that

$$\begin{aligned}
\gamma_j\sum_{t_2=1}^{N}\sum_{t_1=1}^{N}K\left(u_{t_2},u_{t_1}\right)+\sum_{t=r}^{N}\sum_{t_1=1}^{N}\left(\alpha_{t_1}-\alpha_{t_1}^*\right)\\
\times K_{f,0}\left(u_{t-j},u_{t_1}\right)=0,\ j=0,\ldots,m
\end{aligned} \tag{16}$$

$$\frac{\partial L}{\partial a_i^l}=0\Rightarrow a_i^l=\sum_{t=r}^{N}\left(\alpha_t-\alpha_t^*\right)y_f\left(t-i\right) \tag{17}$$

$$\frac{\partial L}{\partial d^l}=0\Rightarrow\sum_{t=r}^{N}\left(\alpha_t-\alpha_t^*\right)=0 \tag{18}$$

$$\begin{aligned}
\frac{\partial L}{\partial \xi_t} &= 0\rightarrow\alpha_t+\beta_t=c,\ t=r,\ldots,N\\
\frac{\partial L}{\partial \xi_t^*} &= 0\rightarrow\alpha_t^*+\beta_t^*=c,\ t=r,\ldots,N
\end{aligned} \tag{19}$$

From (14)-(19), the Lagrangian function (13) can be rewritten as

$$\begin{aligned}
L\left(\boldsymbol{\alpha},\boldsymbol{\alpha}^*,\boldsymbol{\gamma}\right)=-\tfrac{1}{2}\sum_{t=r}^{N}\sum_{t_1=r}^{N}\left(\alpha_t-\alpha_t^*\right)\left(\alpha_{t_1}-\alpha_{t_1}^*\right)\\
\times\left(\sum_{j=0}^{m}K_{f,f}\left(u_{t-j},u_{t_1-j}\right)+\sum_{i=1}^{n}y_f\left(t-i\right)y_f\left(t_1-i\right)\right)\\
+\tfrac{1}{2}\sum_{j=0}^{m}\gamma_j^2\sum_{t=1}^{N}\sum_{t_1=1}^{N}K\left(u_t,u_{t_1}\right)\\
+\sum_{t=r}^{N}\left(\alpha_t-\alpha_t^*\right)y_f\left(t\right)-\sum_{t=r}^{N}\alpha_t\varepsilon-\sum_{t=r}^{N}\alpha_t^*\varepsilon
\end{aligned}$$

where $K_{f,f}$ is defined above. Hence, the dual optimization problem can be written as

$$\begin{aligned}
\min_{\alpha,\alpha^*,\gamma}\frac{1}{2}\begin{bmatrix}\boldsymbol{\gamma}^T & \boldsymbol{\alpha}^T & \boldsymbol{\alpha}^{*^T}\end{bmatrix}\begin{bmatrix}-\mathcal{S}I & 0 & 0\\ 0 & \mathcal{K} & -\mathcal{K}\\ 0 & -\mathcal{K} & \mathcal{K}\end{bmatrix}\\
\times\begin{bmatrix}\boldsymbol{\gamma}\\ \boldsymbol{\alpha}\\ \boldsymbol{\alpha}^*\end{bmatrix}+\begin{bmatrix}0 & -y_f^T\left(r:N\right) & y_f^T\left(r:N\right)\end{bmatrix}\begin{bmatrix}\boldsymbol{\gamma}\\ \boldsymbol{\alpha}\\ \boldsymbol{\alpha}^*\end{bmatrix}\\
+\begin{bmatrix}0 & \varepsilon\cdot\mathbf{1}^T & \varepsilon\cdot\mathbf{1}^T\end{bmatrix}\begin{bmatrix}\gamma\\ \boldsymbol{\alpha}\\ \boldsymbol{\alpha}^*\end{bmatrix}
\end{aligned} \tag{20}$$

subject to

$$\sum_{t=r}^{N}\left(\alpha_t-\alpha_t^*\right)=0$$

$$\gamma_j\mathcal{S}+\sum_{t=r}^{N}\left(\alpha_t-\alpha_t^*\right)K_f^0\left(t,j\right)=0,\ j=0,\ldots,m$$
$$0\leq\alpha_t^*,\alpha_t\leq c,\ t=r,\ldots,N$$

with

$$\begin{aligned}
\mathcal{K}\left(p,q\right)=\sum_{j=0}^{m}K_{f,f}\left(u_{p+r-j-1},u_{q+r-j-1}\right)\\
+\sum_{i=1}^{n}y_f\left(q+r-1-i\right)y_f\left(q+r-1-i\right),\\
\mathcal{S}=\sum_{t_1=1}^{N}\sum_{t_2=1}^{N}K\left(u_{t_1},u_{t_2}\right)\\
K_f^0\left(t,j\right)=\sum_{t_1=1}^{N}K_{0,f}\left(u_{t_1},u_{t+r-j}\right),\\
K_{0,f}\left(u_{t_1},u_t\right)=K\left(u_{t_1},u_t\right)\frac{1}{A^{l-1}(z)}
\end{aligned}$$

Then, a_i^l is given by (17) and d^l can be computed based on the Karush-Kuhn-Tucker (KKT) conditions [10] as follows. If $(\alpha_i$ or $\alpha_i^*)\in(0,c)$ then

$$\begin{aligned}
dv_i &= y_f\left(i\right)-\sum_{j=0}^{m}\Bigg(\gamma_j\sum_{t=1}^{N}K_{0,f}\left(u_t,u_{i-j}\right)\\
&\quad+\sum_{t=r}^{N}\left(\alpha_t-\alpha_t^*\right)K_{f,f}\left(u_{t-j},u_{i-j}\right)\Bigg)\\
&\quad-\sum_{h=1}^{n}a_h^l y_f\left(i-h\right)\pm\varepsilon
\end{aligned} \tag{21}$$

Finally, d^l is calculated as

$$d^l=\frac{\sum_{i=1}^{N-r}\left(dv_i\right)}{N-r} \tag{22}$$

A. Separating Numerator and Nonlinearity Parameters

To extract the numerator parameters, we use the solution presented in [7] and [8], which involves using the SVD of a m by N matrix to compute the nonlinearity output and b parameters. Then, using the training input sequence $[u_1,\ldots,u_N]$ and the extracted sequence of the nonlinearity responses, we can train a SVM to represent the nonlinear part of the Hammerstein system.

III. ILLUSTRATIVE EXAMPLE

In this section, the algorithm described above will be applied to the identification of the relationship between the ankle velocity and the GS-EMG. This problem has been studied extensively in [1] and [4]. The data were created as follows: a pulse sequence was used as the reference input for an electrohydraulic position servo. Then, the ankle position, the response to the torque produced by the position servo, and the GS-EMG were measured [3]. The relationship between the ankle velocity, obtained by numerically differentiating the measured position, and the GS-EMG, was modeled as a Hammerstein system, and identified from 5 seconds of data, sampled at 200 Hz, which resulted in 1000 input/output measurements. The nonlinear part was represented by a SVM and the linear part was modeled by an output error model with $n = 2$ and $m = 0$. The SVM training is controlled by a number of hyper-parameters: the choice of kernel function and the parameters associated with that kernel, and the regularization parameter, c. These values were selected based on cross-validation where we partitioned the data set into training and validation sets. Then, different values of the linear model order or one of the hyper parameters are compared by evaluating their performance on the validation data while keeping the others fixed [9]. For example, the regularization parameter c value was chosen by comparing the performance of the validation data on values ranged from 10 to 500 while keeping the linear model order and the other hyper parameters values fixed. The best model was obtained using an RBF-kernel with $\sigma = 1$ and a regularization parameter $c = 7$.

Fig. 1 shows the elements of the identified SVM-output error Hammerstein system. For comparison, Fig. 2 shows the results obtained using a polynomial and cubic spline based models together with FIR linear dynamics. As with the SVM Hammerstein model, these were identified using 1000 data points, and validated on the next 1000 points. In the validation segment, the polynomial, spline, and SVM models accounted for 94.6, 94.8, and 95.5 percent of the output variance, respectively.

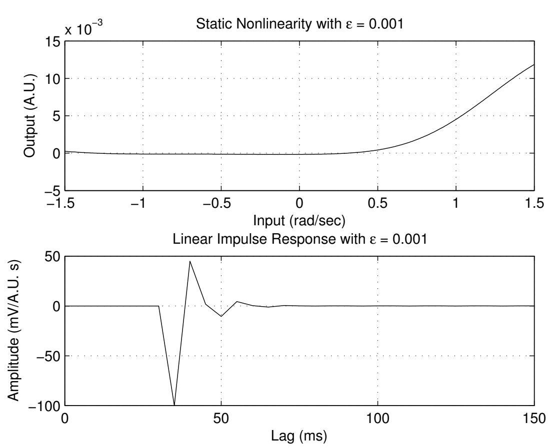

Fig. 1. Elements of the Identified SVM-OE Hammerstein Cascade model of the Stretch Reflex EMG

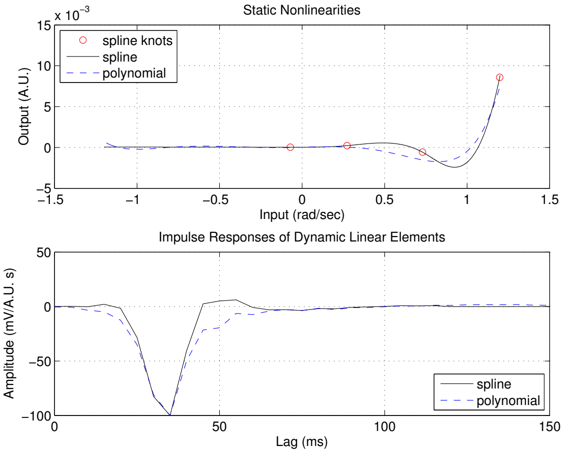

Fig. 2. Identified Hammerstein Cascades of the Stretch Reflex EMG using separable least squares with polynomial and cubic spline nonlinearities.

IV. CONCLUSION

An identification algorithm for Hammerstein models consisting of a SVM nonlinearity followed by a linear output error model was developed, and used to construct a model of the relationship between the ankle angular velocity and the EMG measured over the GS muscles. The SVM was able to model a complex nonlinearity, without requiring any a-priori assumptions regarding its structure. It is clear from the results that the SVM based approach provides better predictions of the reflex EMG than the polynomial and cubic spline based models. Furthermore, the identified nonlinearity does not contain the negative deflections present in the polynomial and spline based nonlinearities.

V. ACKNOWLEDGMENTS

The EMG data were provided by R. E. Kearney from the Department of Biomedical Engineering, McGill University.

REFERENCES

[1] D.T. Westwick and R. E. Kearney. Separable least squares identification of nonlinear Hammerstein models: Application to stretch reflex dynamics,. Ann. Biomed. Eng., 29(8):707–718, 2001.

[2] D.T. Westwick and R.E. Kearney. Identification of Nonlinear Physiological Systems. Wiley-IEEE Press, Piscataway, NJ, 2003.

[3] R.E. Kearney and I.W. Hunter. Nonlinear identification of stretch reflex dynamics. Ann. Biomed. Eng., 16:79–94, 1988.

[4] E.J. Dempsey and D.T. Westwick. Identification of Hammerstein models with cubic spline nonlinearities. IEEE Trans. Biomed. Eng., 51(2):237–245, 2004.

[5] K.S. Narendra and P.G. Gallman. An iterative method for the identification of nonlinear systems using a Hammerstein model. IEEE Trans. Aut. Cntrl., 11:546–550, 1966.

[6] F.H.I. Chang and R. Luus. A noniterative method for identification using Hammerstein model. IEEE Tran. Aut. Cntrl., 16:464–468, 1971.

[7] I. Goethals, K. Pelckmans, J. Suykens, and B. De Moor. Identification of MIMO Hammerstein models using least squares support vector machines. Automatica, 41(7): 1263–1272, 2005.

[8] M. Al Dhaifallah and D.Westwick. Identification of NARX Hammerstein models based on support vector machines. In Proc. 17th IFAC World Cong., pages 4999–5004, 2008.

[9] L. Ljung. System identification: theory for the user. 2nd ed. Prentice Hall PTR, Upper Saddle River, NJ, 1999.

[10] V. N. Vapnik. Statistical Learning Theory. Wiley, New York, 1998.

[11] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle. Least squares support vector machines. World Scientific, Singapore, 2002.