SUPPORT VECTOR MACHINE IDENTIFICATION OF STRETCH REFLEX DYNAMICS

Mujahed Al Dhaifallah And David T. Westwick
Department of Electrical and Computer Engineering,
Schulich School of Engineering, Calgary, Canada

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. 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.

Recently, support vector machines (SVMs) and least squares support vector machines (LS-SVMs) have shown powerful abilities in approximating linear and nonlinear functions [5], [6]. 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. SVMs and LS-SVMs are fit by solving convex optimization problems, and do not require a-priori structural information [6]. Al Dhaifallah and Westwick [7] formulated an algorithm to identify NARX Hammerstein models with nonlinearities represented using SVMs. In this paper, this algorithm will be used to identify the stretch reflex dynamics model.

SUPPORT VECTOR MACHINES FOR FUNCTION ESTIMATION

Basically, to construct a support vector machine for real-valued function estimation problems, the input data are mapped into a high-dimensional feature space where a linear function is constructed. A kernel function is used to avoid constructing this mapping explicitly.

STANDARD SVM REGRESSION

Consider the nonlinear regression model $y = f(\mathbf{x})+v$ where $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is an unknown scalar-valued function and v is an additive white noise term. x_i is a sample value of the input vector $\mathbf{x}$ and y_i is the corresponding value of the model output y. In the primal space, the following model is assumed for $f(\mathbf{x})$

$$f(\mathbf{x})=\mathbf{w}^T\boldsymbol{\varphi}(\mathbf{x})+d_0 \tag{1}$$

Where $\boldsymbol{\varphi}: \mathbb{R}^d \rightarrow \mathbb{R}^{nH}$ 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. Now, to find an estimate of the dependence of y on $\mathbf{x}$ in the standard SVM sense, a cost function consisting of a weighted average of ε-insensitive cost function and the L-2 norm of the weight vector is minimized,

$$\min_{w,\xi,\xi^*} \frac{1}{2} w^T w + c\sum_{i=1}^{l}\left(\xi_i + \xi_i^*\right) \tag{2}$$

subject to

$$\begin{aligned} y_i - w^T\varphi(x) - d_0 &\le \varepsilon + \xi_i \\ w^T\varphi(x) + d_0 - y_i &\le \varepsilon + \xi_i^* \end{aligned} \tag{3}$$

$$\xi_i, \xi_i^* \geq 0, \mathrm{i} = 1,...,l$$

Where ε is the accuracy level of the approximation, c > 0 is a constant that determines the relative weighting of the two terms, and ξ_i and ξ_i^* are the errors in the ε-insensitive cost function, which are treated as slack variables in the optimization problem.

The optimization problem just described is the primal problem for regression. To formulate the corresponding dual problem, we write the Lagrangian function L. Then, we minimize L with respect to the weight vector **w** and slack variables **ξ** and $\boldsymbol{\xi}^*$ and maximize with respect to the Lagrange multipliers. By carrying out this optimization we can write **w** in terms of the Lagrange multipliers. Finally, we can substitute the value of **w** and simplify to get the following dual problem (see [6] for details)

$$\min_{\alpha,\alpha^*} \frac{1}{2}\begin{bmatrix}\alpha & \alpha^*\end{bmatrix}\begin{bmatrix} KM & -KM \\ -KM & KM \end{bmatrix}\begin{bmatrix}\alpha \\ \alpha^*\end{bmatrix} + \begin{bmatrix}-y_{1:l}^T + \varepsilon \cdot 1_l^T & y_{1:l}^T + \varepsilon \cdot 1_l^T\end{bmatrix}\begin{bmatrix}\alpha \\ \alpha^*\end{bmatrix} \tag{4}$$

subject to

$$\begin{bmatrix}1_l^T & -1_l^T\end{bmatrix}\begin{bmatrix}\alpha \\ \alpha^*\end{bmatrix} \leq 0$$
$$0_{2l} \leq \begin{bmatrix}\alpha \\ \alpha^*\end{bmatrix} \leq c1_{2l}^T \tag{5}$$

With $KM(i,j) = K(\mathrm{x}_i, \mathrm{x}_j)$

Where **α** and **α***** are vectors of Lagrange multipliers. This maximization is a quadratic program, which can be solved using standard tools [5], [6]. Finally, the nonlinear function model takes the form

$$\mathrm{f(x)} = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)K(\mathrm{x}, \mathrm{x}_i) + d_0 \tag{6}$$

Where $K(x_i,x_j)= \varphi^T(\mathrm{x}_i)\varphi(\mathrm{x}_j)$ is a kernel function used to represent the inner product in the feature space. They can be any symmetric function satisfying Mercer's condition [6]. Typical examples are the use of a polynomial kernel or the radial basis function (RBF) kernel.

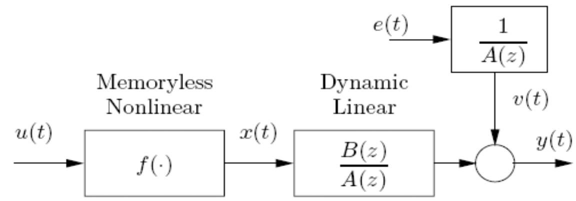

Figure 1: Block diagram of an ARX-Hammerstein cascade. The investigator is assumed to have access to the input, u(t), and the output, y(t), but not the intermediate signal, x(t) or the innovation, e(t).

IDENTIFICATION OF NONLINEAR ARX HAMMERSTEIN MODELS

The Hammerstein cascade, a static nonlinearity followed by a linear filter as shown in Figure 1, is often used to represent certain higher-order nonlinear systems.

The output of the NARX Hammerstein model is given by:

$$y_t = \sum_{i=1}^{n} a_i y_{t-i} + \sum_{j=0}^{m} b_j f(u_{t-j}) + e_t \tag{7}$$

Where u_t, y_t $\in \mathbb{R}$, are the input and output measurements, respectively, for t=1,…,N. The noise e_t is assumed to be white and *m* and *n* denote the order of the numerator and denominator in the transfer function of the linear model. The static nonlinearity is assumed to have the form (1).

Following Goethals [8] overparameterization approach, an overparameterized, but linear in the parameters, model [8], [9] is initially identified. Hence, (7) can be rewritten as

$$y_t = \sum_{i=1}^{n} a_i y_{t-i} + \sum_{j=0}^{m} \mathrm{w}_j^{\mathrm{T}} \varphi(u_{t-j}) + d + e_t \tag{8}$$

Where

$$\mathrm{w}_{\mathrm{j}} = b_j \mathrm{w}, d = d_0 \sum_{j=0}^{m} b_j$$

Note that models of this form can be uniquely identified, 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.

Now, to identify the linear and nonlinear parts, solve the following optimization problem

$$\min_{w_j, a_i, \xi, \xi^*} \frac{1}{2}\sum_{j=0}^{m} w_j^T w_j + \frac{1}{2}\sum_{i=1}^{n} a_i^2 + c\sum_{t=1}^{N}\left(\xi_t + \xi_t^*\right) \quad (9)$$

subject to

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

$$y_t - \sum_{i=1}^{n} a_i y_{t-i} - \sum_{j=0}^{m} w_j^T \varphi(u_{t-j}) - d \le \varepsilon + \xi_t$$

$$\sum_{i=1}^{n} a_i y_{t-i} + \sum_{j=0}^{m} w_j^T \varphi(u_{t-j}) + d - y_t \le \varepsilon + \xi_t^* \quad (11)$$

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

Note that (9) is a standard SVM objective function, consisting of the 2 norm of the parameters (w and a) and the Vapnik ε-insensitive cost function (2) applied to the residuals. The constraints in (11) are derived by modifying the constraints of the standard SVM to include the dynamics of the ARX model. Constraints (10) were added to center the nonlinear functions $\mathbf{w}_j^T\varphi(\cdot)$, j=0,…,m around their average over the training set [7], [8].

The dual optimization problem can be written as

$$\min_{\alpha,\alpha^*} \frac{1}{2}\begin{bmatrix}\gamma & \alpha & \alpha^*\end{bmatrix}\begin{bmatrix} -SI_{m+1} & 0 & 0 \\ 0 & K & -K \\ 0 & K & K \end{bmatrix}\begin{bmatrix}\gamma \\ \alpha \\ \alpha^*\end{bmatrix} + \quad (13)$$

$$\begin{bmatrix}0_{m+1} & -y_{r:N}^T + \varepsilon\cdot 1_{N-r+1}^T & y_{r:N}^T + \varepsilon\cdot 1_{N-r+1}^T\end{bmatrix}\begin{bmatrix}\gamma \\ \alpha \\ \alpha^*\end{bmatrix}$$

subject to

$$\begin{bmatrix}1_{N-r+1}^T & -1_{N-r+1}^T\end{bmatrix}\begin{bmatrix}\alpha \\ \alpha^*\end{bmatrix} \le 0$$

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

$$0_{2(N-r+1)} \le \begin{bmatrix}\alpha \\ \alpha^*\end{bmatrix} \le c1_{2(N-r+1)}^T$$

With

$$K(p,q) = \sum_{j=0}^{m} K\left(u_{p+r-j-1}, u_{q+r-j-1}\right)$$

$$+ \sum_{i=1}^{n} y_{p+r-i-1} y_{q+r-i-1}$$

$$K^0(t, j) = \sum_{t_1=1}^{N} K\left(u_{t_1}, u_{t+r-j}\right)$$

$$S = \sum_{t_1=1}^{N}\sum_{t_2=1}^{N} K\left(u_{t_1}, u_{t_2}\right)$$

For a detailed derivation of (13), the interested is referred to [7]. By solving (13), one gets **α**, **α***, and **γ**. Hence, a_i is given by $a_i = \sum_{t=r}^{N}\left(\alpha_t - \alpha_t^*\right)y_{t-i}$ and *d* can be computed based on the Karush-Kuhn-Tuker (KKT) conditions as follows. If (α_i or α_i^*)∈(0,c) then

$$dv_i = y_i - \sum_{j=0}^{m}\left(\gamma_j \sum_{t=1}^{N} K\left(u_t, u_{t-j}\right) + \sum_{t=r}^{N}\left(\alpha_t - \alpha_t^*\right)\right. \quad (14)$$

$$\left.\times K\left(u_{t-j}, u_{i-j}\right)\right) - \sum_{l=1}^{n} a_l y_{i-l} \pm \varepsilon$$

Finally, *d* is calculated as $d = \frac{\sum_{i=1}^{N-r} dv_i}{N-r}$.

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,⋯,u_N] and the extracted sequence of the nonlinearity responses, we can train a SVM to represent the nonlinear part of the Hammerstein system.

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 (see Kearney and Hunter [3] for details regarding the experimental procedure). The relationship between the ankle velocity, obtained by numerically differentiating the measured position, and the GS-EMG, modeled as a Hammerstein system, was identified using the first 1000 data points. The nonlinear part was represented by a SVM and the linear part was modeled by an ARX model of order two. 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 (see Eq. (2)). 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 there performance on the validation data while keeping the others fixed [10]. 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=29.

Figure 2 shows the elements of the identified SVM-ARX Hammerstein system. Note that the nonlinearity (lower panel) resembles a half-wave rectifier, but includes a threshold (at about 0.4 rad/sec) and smooth transition from the inactive to the active regions, and the beginnings of a saturation at the highest velocities tested. The linear dynamics are less well defined, perhaps because of the need to use a low-order ARX model in the present algorithm. As evident in Figure 1, the ARX structure includes a noise model which shares the same poles as the deterministic system. The ARX structure was used in this study because it is compact and linear in the variables. Future work will consider the extension of this SVM based identification technique to include the output-error class of linear system models.

CONCLUSION

An identification algorithm for Hammerstein models consisting of a Support Vector Machine nonlinearity followed by an ARX model for the linear dynamics was developed, and used to construct a model of the relationship between the ankle angular velocity and the EMG measured over the Gactrocnemius-Soleus muscles. The SVM was able to model a complex nonlinearity, without requiring any a-priori assumptions regarding its structure. The present algorithm is limited to the use of ARX linear dynamics, which may not have been suitable for the system under study.

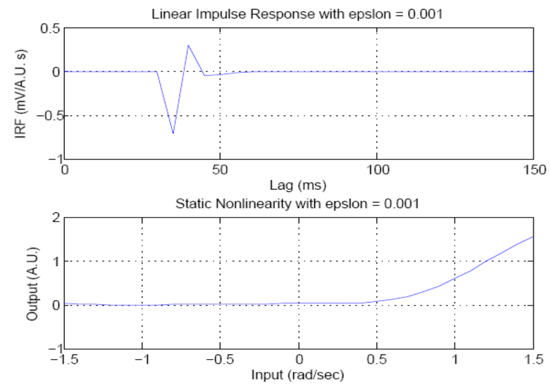

Figure 2: Identified Hammerstein Cascades of the Stretch Reflex EMG using Support Vector Machine Regression with Support Vector Machine nonlinearity.

ACKNOWLEDGEMENTS

The EMG data was provided by R. E. Kearney and M. M. Mirbagheri from the Department of Biomedical Engineering, McGill.

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, August 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, February 2004.

[5] 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.

[6] V. Vapnik. Statisical Learning Theory. John Wiley & Sons, New York, 1998.

[7] M. Al Dhaifallah and D.Westwick. Identification of NARX Hammerstein models based on support vector machines. In Proceedings of 17th IFAC World Congress, pages 4999--5004, 2008.

[8] 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, July 2005.

[9] F.H.I. Chang and R. Luus. A noniterative method for identification using Hammerstein model. IEEE Transactions on Automatic Control, 16:464--468, 1971.

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