CONFIDENTIAL. Limited circulation. For review only.

Identification of Wiener-Hammerstein Systems using LS-SVMs

Tillmann Falck *, **Kristiaan Pelckmans** *,
Johan A.K. Suykens *, **Bart De Moor** *

* *Katholieke Universiteit Leuven - ESAT - SCD/SISTA,*
Kasteelpark Arenberg 10,
B-3001 Leuven (Heverlee), Belgium
email: {tillmann.falck,kristiaan.pelckmans,johan.suykens,bart.demoor}
@esat.kuleuven.be

Abstract: This paper extends the overparametrization techinque as used for Hammerstein systems employing nonlinear Least-Squares Support Vector Machines (LS-SVMs) towards the identification of Wiener-Hammerstein systems. We present some practical guidelines as well as empirical results on the performance of the method with respect to various deficiencies of the excitation signal. Finally we apply our method to the SYSID2009 Wiener-Hammerstein benchmark data set.

Keywords: Wiener-Hammerstein systems, Identification, Over-parametrization.

1. INTRODUCTION

Wiener-Hammerstein systems (Bilings and Fakhouri [1978]) are an important special class of nonlinear systems consisting of a concatenation of two linear subsystems with a static nonlinearity in between them as shown in Figure 1. There exists extensive literature on the estimation of the easier Hammerstein systems, see for example Narendra and Gallman [1966], Pawlak [1991] or Bai and Fu [2002]. Yet for the case of Wiener-Hammerstein systems all known methods are limited to rather simple models restricting attention to the case of low order linear dynamic blocks and simple nonlinearities, see e.g.Tan and Godfrey [2002], Bershad et al. [2001], Enqvist and Ljung [2005] and Greblicki and Pawlak [2008]. One popular approach for identification of Hammerstein systems known in literature is overparmatrization, dividing the identification in two stages Chang and Luus [1971], Bai [1998]: identification of a slightly broader model class, and projection of the estimate on the strict Hammerstein class. This paper extends this technique to handle input dynamics.

The proposed algorithm extends the result in Goethals et al. [2005] where the static nonlinearity is modeled by Least Squares - Support Vector Machines (or LS-SVMs). LS-SVMs (Suykens et al. [2002]) are a variation of the classical Support Vector Machines (SVMs) (Vapnik [1998], Schölkopf and Smola [2002]) which have proven their use in the field of machine learning. The technique is closely related to the technique of smoothing splines Wahba [1990] but generalizes to the the use of arbitrary kernel functions. Both techniques perform nonlinear regression by projecting the input into a high (possibly infinite) dimensional space and then solving a regularized linear regression problem in that feature space. The mapping into the feature space itself is typically not known explicitly, but is induced by a proper positive definite kernel function. A main advantage of doing so is that the dual problem corresponding

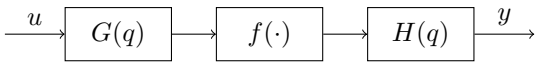


Fig. 1. General structure of a Wiener-Hammerstein system

to the original (primal) optimization (estimation) problem can be expressed in terms of a finite number of unknowns, and allows for efficient solvers to be used. An important difference between SVMs and LS-SVMs is that the latter can be solved as a linear system, while the former needs to solve a convex quadratical program.

The contribution of this paper is to show that the combination of overparametrization with LS-SVMs as in Goethals et al. [2005] yields a powerful algorithm to identify Wiener-Hammerstein systems. The key idea is to jointly model the input linear dynamic block with the nonlinearity, and apply an overparametrization technique to recover the parameters of the linear output system $H(q)$. The novelty of this approach is firstly that it assumes no prior knowledge of any of the blocks. Secondly, it does not rely on inversion at any stage. Thus the class of static nonlinearities is not restricted and there are no restrictions on the zeros or poles of the second linear system. The only assumption is that $G(q)$ has finite memory. From a practical perspective the model order of $G(q)$ should not be too high. This can be seen as follows: if the expressive power of the function covering $H(q)$ and f would get too large, one would be able to capture the output dynamics with this model, and the output parameters cannot be estimated accurately. In general, as we consider the identification of the first linear block and the nonlinearity jointly, the technique of regularization becomes even more crucial to avoid (numerical) ill-conditioning than in the simpler Hammerstein identification methods. We investigate and compare the use of the proposed estimation scheme with respect to prediction performance for cases where the input signal is suboptimal (colored, short), and would result in bad

Preprint submitted to 15th IFAC Symposium on System Identification.
Received November 8, 2008.

estimates in the purely linear case Söderström and Stoica [1989]. Furthermore we study the influence of different cross-validation schemes and report the performance of the proposed method on the SYSID2009 benchmark data Schoukens et al. [2008].

The paper uses the following notation. Vector valued variables are written as lowercase, boldface letters whereas matrices are denoted by uppercase, boldface characters. All variables that are not boldfaced are scalar valued. Constants are usually put as capital letters. The backshift operator q is defined as $qx_t = x_{t-1}$. With abuse of notation the subscript of a variable might refer to both the element of a vector or the element of a sequence. Functions, denoted by lowercase names, are real valued and always have real domains. Systems are specified by their transfer functions which are in q and typeset in uppercase letters. Estimated values are highlighted with a hat.

The paper is organized as follows. Section 2 describes the proposed method. In Section 3 the method is applied to numerical data. First it is demonstrated on several interesting toy examples and then evaluated on the benchmark data. The next section is concerned with the analysis with respect to different excitation signal properties. Conclusions are given in Section 5.

2. IDENTIFICATION OF WIENER-HAMMERSTEIN SYSTEMS

The technique presented in Goethals et al. [2005] is extended towards the convex identification of another nonlinear block oriented model, defined as follows.

Definition 1. (Wiener-Hammerstein System). The system shown in Figure 1 system consisting of a sequence of a linear dynamic - static nonlinearity and a dynamic linear system is called Wiener-Hammerstein system. It is described by

$$y_t = H(q)(f(G(q)u_t)) \tag{1}$$

where f possesses typically a smoothness property. Let $G(q)$ and $H(q)$ be parameterized as a FIR and ARX model respectively. The FIR part has order $P \in \mathbb{N}$ and the ARX part having orders $R_1, R_2 \in \mathbb{N}$, respectively with parameters $\boldsymbol{a} = [a_0, \ldots, a_{P-1}]^T \in \mathbb{R}^P$, $\boldsymbol{b} = [b_0, \ldots, b_{R_1-1}]^T$ and $\boldsymbol{c} = [c_0, \ldots, c_{R_2-1}]^T$.

Given a set of measurements $\{(u_t, y_t)\}_{t=1}^T$ of inputs and outputs the proposed method estimates the parameters $\boldsymbol{b}$ and $\boldsymbol{c}$ for $H(q)$ and a dynamical nonlinear block describing $f(G(q)u_t)$. For notational ease define vectors of lagged inputs $\boldsymbol{u}_t = [u_t, \ldots, u_{t-D+1}]^T$ and outputs $\boldsymbol{y}_t = [y_{t-1}, \ldots, y_{t-R_2}]^T$. Then the full model is given by

$$y_t = \sum_{r=0}^{R_1-1} b_r f(\boldsymbol{a}^T \boldsymbol{u}_{t-r}) + \boldsymbol{c}^T \boldsymbol{y}_t. \tag{2}$$

For further clarity let the index τ be defined as $\tau = \max(P + R_1, R_2)$.

The identification of a Wiener-Hammerstein system based on observations can be formalized as

$$\min_{\boldsymbol{a},\boldsymbol{b},\boldsymbol{c},f} \mathcal{J}(f, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}) = \sum_{t=\tau}^{T} \left(y_t - \sum_{r=0}^{R_1-1} b_r f(\boldsymbol{a}^T \boldsymbol{u}_{t-r}) + \boldsymbol{c}^T \boldsymbol{y}_t \right)^2. \tag{3}$$

Definition 2. (Overparametrisation). The method of overparametrisation starts from representing the Wiener-Hammerstein model as

$$y_t = \sum_{r=0}^{R_1-1} f_r(\boldsymbol{u}_{t-r}) + \boldsymbol{c}^T \boldsymbol{y}_t + e_t \tag{4}$$

where we assume implicitly that $f_r(\boldsymbol{x}) = b_r f(\boldsymbol{a}^T \boldsymbol{x})$ for all $r = 0, \ldots, R_1 - 1$. After identifying the unknowns $\{f_r\}$ and $\{c_r\}$, the obtained model is projected back onto the Wiener-Hammerstein class. In particular, we decompose the functions $\{f_r\}$ into a global f and linear parameters $\{b_r\}$ by computing the best rank one approximation.

2.1 Identification

We consider the identification using kernel based models using the LS-SVM formalism (Suykens et al. [2002]), as illustrated in Goethals et al. [2005].

$$\min_{\boldsymbol{w}_r,\boldsymbol{c},d,e_t} \frac{1}{2} \sum_{r=0}^{R_1-1} \boldsymbol{w}_r^T \boldsymbol{w}_r + \frac{\gamma}{2} \sum_{t=\tau}^{T} e_t^2 \tag{5a}$$

subject to

$$y_t = \sum_{r=0}^{R_1-1} \boldsymbol{w}_r^T \boldsymbol{\varphi}(\boldsymbol{u}_{t-r}) + \boldsymbol{c}^T \boldsymbol{y}_t + d + e_t \quad \forall t = \tau, \ldots, T \tag{5b}$$

$$\sum_{t=\tau}^{T} \boldsymbol{w}_r^T \boldsymbol{\varphi}(\boldsymbol{u}_{t-r}) = 0 \ \forall r = 0, \ldots, R_1 - 1 \tag{5c}$$

where the $\boldsymbol{w}_r$ parameterize the nonlinear functions f_r, $\boldsymbol{\varphi} : \mathbb{R}^D \to \mathbb{R}^{n_h}$ is the feature map and $d \in \mathbb{R}$ denotes the offset. The model (4) is incorporated as (5b). The second constraint (5c) is used to center the nonlinear functions which is helpful for the final projection.

Define a data matrix $\boldsymbol{Y} = [\boldsymbol{y}_\tau, \ldots, \boldsymbol{y}_T]^T$ for the linear AR part of the model and a vector of outputs $\boldsymbol{y} = [y_\tau, \ldots, y_T]^T$. Let $\boldsymbol{1} \in \mathbb{R}^{T-\tau}$ denote a vector of all ones. After forming the Lagrangian for (5) and solving the conditions for optimality, the solution to the estimation problem can be found by solving the linear system

$$\begin{bmatrix} (\boldsymbol{\Omega}_A + \gamma_1^{-1}\boldsymbol{I}) & \boldsymbol{\Omega}_C & \boldsymbol{1} & \boldsymbol{Y} \\ \boldsymbol{\Omega}_C^T & & & \diagup \\ \boldsymbol{1}^T & & 0 & \\ \boldsymbol{Y}^T & \diagup & & \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \\ d \\ \boldsymbol{c} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{0} \\ 0 \\ \boldsymbol{0} \end{bmatrix}. \tag{6}$$

Let $\boldsymbol{\Omega}^{r,s} \in \mathbb{R}^{(T-\tau)\times(T-\tau)}$ be shifted kernel matrices defined as

$$\Omega_{i,j}^{r,s} = K(\boldsymbol{u}_{i-r}, \boldsymbol{u}_{j-s}) = \boldsymbol{\varphi}^T(\boldsymbol{u}_{i-r})\boldsymbol{\varphi}(\boldsymbol{u}_{j-s}) \tag{7}$$

for all $i, j = \tau, \ldots, T$. Then the summed kernel matrix is

$$\boldsymbol{\Omega}_A = \sum_{r=0}^{R_1-1} \boldsymbol{\Omega}^{r,r} \tag{8}$$

whereas the R_1 centering constraints are represented by

$$\boldsymbol{\omega}_C^r = \boldsymbol{\Omega}^r \boldsymbol{1} \tag{9}$$

and collated in $\boldsymbol{\Omega}_C = [\boldsymbol{\omega}_C^0, \ldots, \boldsymbol{\omega}_C^{R_1-1}]$.

Preprint submitted to 15th IFAC Symposium on System Identification.
Received November 8, 2008.

The individual estimated functions can be evaluated with

$$\hat{f}_r(\boldsymbol{x}) = \sum_{k=\tau}^{T} (\alpha_k + \beta_r)\, K(\boldsymbol{x}, \boldsymbol{u}_{k-r}) \tag{10}$$

and the global overparmeterised model can be evaluated as

$$\hat{f}(\boldsymbol{u}_t, \ldots, \boldsymbol{u}_{t-R_1+1}, \boldsymbol{y}_t) = \sum_{r=0}^{R_1-1} \hat{f}_r(\boldsymbol{u}_{t-r}) + \boldsymbol{c}^T \boldsymbol{y}_t + d. \tag{11}$$

2.2 Projection by Rank-one Approximation

The second stage of the overparametrisation approach projects the R_1 identified nonlinear functions onto the set of linear parameters $\{\hat{b}_r\}_{r=0}^{R_1-1}$ and a set of nonlinear function values $\{\hat{f}(\boldsymbol{u}_t)\}_{t=2\tau}^{T}$. Therefore we form a matrix $\boldsymbol{F}$ that contains all values for the individual functions $\hat{f}_r(\cdot)$ evaluated at $\{\boldsymbol{u}_t\}_{t=2\tau}^{T}$ as in

$$\left[\begin{array}{c} * \\ \hline \boldsymbol{F} \\ \hline * \end{array}\right] := \left[\begin{array}{ccccc} \hat{f}_0(\boldsymbol{u}_\tau) & 0 & \cdots & & 0 \\ \vdots & & \ddots & & \vdots \\ \hat{f}_0(\boldsymbol{u}_{2\tau-1}) & & & & 0 \\ \hline \hat{f}_0(\boldsymbol{u}_{2\tau}) & & & & \hat{f}_{R_1-1}(\boldsymbol{u}_{2\tau}) \\ \vdots & & & & \vdots \\ \hat{f}_0(\boldsymbol{u}_T) & & & & \hat{f}_{R_1-1}(\boldsymbol{u}_T) \\ \hline 0 & & & & \hat{f}_{R_1-1}(\boldsymbol{u}_{T+1}) \\ \vdots & & \ddots & & \vdots \\ 0 & & \cdots & 0 & \hat{f}_{R_1-1}(\boldsymbol{u}_{T+\tau}) \end{array}\right] \tag{12}$$

where the blocks of zeros are lower and upper diagonal respectively and the matrix $\boldsymbol{F}$ is of dimension $(T-2\tau)\times R_1$.

Then the projection can be performed by taking the best rank one approximation of $\boldsymbol{F}$ as follows

$$\boldsymbol{F} = \begin{bmatrix} \hat{f}(\boldsymbol{u}_{2\tau}) \\ \hat{f}(\boldsymbol{u}_{2\tau+1}) \\ \vdots \\ \hat{f}(\boldsymbol{u}_T) \end{bmatrix} \begin{bmatrix} c_0 & \hat{b}_1 & \ldots & \hat{b}_{R_1-1} \end{bmatrix}. \tag{13}$$

This can be computed efficiently by computing the Singular Value Decomposition (SVD), and witholding the singular vectors corresponding to the largest singular value.

2.3 Estimation of the Final Model

With the estimated values for $\{b_r\}_{r=0}^{R_1-1}$, the overparametrization and the centering constraints can be removed from the estimation problem in (5) yielding

$$\min_{\boldsymbol{w},\boldsymbol{c},d,e_t} \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \frac{\gamma}{2}\sum_{t=\tau}^{T} e_t^2 \tag{14a}$$

subject to

$$y_t = \sum_{r=0}^{R_1-1} \hat{b}_r \boldsymbol{w}^T \boldsymbol{\varphi}(\boldsymbol{u}_{t-r}) + \boldsymbol{c}^T \boldsymbol{y}_t + d + e_t \quad \forall t = \tau, \ldots, T \tag{14b}$$

The solution is again obtained by a solving a linear system

$$\begin{bmatrix} (\boldsymbol{\Omega}_B + \gamma_2^{-1}\boldsymbol{I}) & \mathbf{1} & \boldsymbol{Y} \\ \mathbf{1}^T & 0 & \mathbf{0} \\ \boldsymbol{Y}^T & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ d \\ \boldsymbol{c} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y} \\ 0 \\ \mathbf{0} \end{bmatrix} \tag{15}$$

with

$$\boldsymbol{\Omega}_B = \sum_{r=0}^{R_1-1}\sum_{s=0}^{R_1-1} \hat{b}_r \hat{b}_s \boldsymbol{\Omega}^{r,s}. \tag{16}$$

Estimates of the function are then obtained by

$$\hat{f}(\boldsymbol{u}_t, \ldots, \boldsymbol{u}_{t-R_1+1}, \boldsymbol{y}_t) = \sum_{k=\tau}^{T} \alpha_k \sum_{r=0}^{R_1-1}\sum_{s=0}^{R_1-1} K(\boldsymbol{u}_{t-r}, \tilde{\boldsymbol{u}}_{k-s}) + \boldsymbol{c}^T \boldsymbol{y}_t + d \tag{17}$$

where $\{\tilde{\boldsymbol{u}}_k\}_{k=\tau}^{T}$ is the set of inputs used during the estimation phase.

The complete approach in summarized in Algorithm 1.

Algorithm 1 Identification of Wiener-Hammerstein systems

(1) Compute the shifted kernel matrices $\boldsymbol{\Omega}^{r,s}$ in Eq. (7)
(2) Construct the summed kernel matrix $\boldsymbol{\Omega}_A$ and the centering constraints $\boldsymbol{\omega}_C^r$ in Eqs. (8) & (9)
(3) Solve the dual system for the component-wise LS-SVM in (6)
(4) Construct the matrix of predictions $\boldsymbol{F}$ in (12)
(5) Compute the SVD of $\boldsymbol{F}$ and take the rank one approximation as in Eq. (13)
(6) Compute a weighted kernel matrix $\boldsymbol{\Omega}_B$ as in Eq. (16)
(7) Solve the dual for the weighted LS-SVM in Eq. (15)

3. EMPIRICAL ASSESSMENTS

3.1 Training Procedure

Especially in nonlinear estimation techniques the model selection is critical for the model performance on independent data. In the proposed method we have chosen to parameterize the model in terms of the bandwidth σ of the used RBF kernel $k(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\|\boldsymbol{x}-\boldsymbol{y}\|^2/\sigma^2)$ and the regularization parameters γ_1 and γ_2 for the systems in (6) and (15) respectively. Additionally the model order R_1, R_2 and D has to be determined.

The selection is made in two stages. At first, we tune the model parameters $(\gamma_1, \gamma_2, , \sigma)$ in Algorithm 1 such that the cross-validation performance becomes maximal. We found empirically that σ_1 and γ_1 should be chosen from a grid of possible parameters, while γ_2 in Eq. (16) can be tuned independently once γ_1 and σ_1 are fixed.

The outer level is used for model order selection. For all possible model orders (P, R_1, R_2), the above paragraph describes a method to tune $(\gamma_1, \gamma_2, \sigma_1)$. For application on the Benchmark data in Subsection 3.4. we found empirically that it is plausible to take $D \equiv R_1$. As a consequence, we only have to select proper values of R_1 and R_2 of a grid of values. For each possible pair of orders (R_1, R_2) on the grid, the performance is evaluated on an independent validation set.

Preprint submitted to 15th IFAC Symposium on System Identification.
Received November 8, 2008.

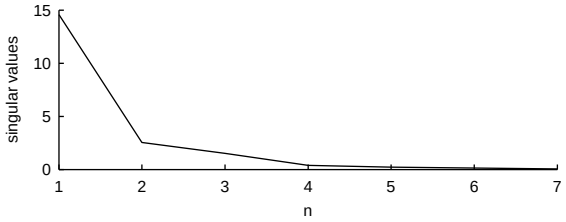


Fig. 2. The singular value spectrum for the matrix $\boldsymbol{F}$ for $G_2(q)$, $H_2(q)$ and $f_2(x)$.

3.2 Examples

Wiener-Hammerstein systems as shown in Figure 1 consist of 2 linear subsystems and the static nonlinearity. In this section we will analyze the capability of the proposed algorithm to identify models of different complexities. For the linear subsystem at the input define the systems

- $G_1(q) = 1$ (static),
- $G_2(q)$ with one zero at $z = 0.9$ and
- $G_3(q)$ with zeros at $z = 2.14 \cdot e^{\pm j0.1\pi}$, $1 \cdot e^{\pm j0.9\pi}$, $1 \cdot e^{\pm j0.7\pi}$, $1 \cdot e^{\pm j0.5\pi}$, $0.47 \cdot e^{\pm j0.1\pi}$.

At the output the following linear systems will be considered

- $H_1(q)$ with zeros at $z = 0.89 \cdot e^{\pm j0.63\pi}$, $0.84 \cdot e^{\pm j0.42\pi}$, $0.39 \cdot e^{\pm j0.5\pi}$ and poles at $p = 0.88 \cdot e^{\pm j0.21\pi}$, $0.94 \cdot e^{\pm j0.73\pi}$, $0.96 \cdot e^{\pm j0.52\pi}$ (ARX) and
- $H_2(q)$ with zeros at $z = 1j, 1 \cdot e^{\pm j0.55\pi}$ and poles $p = 0.48, 0.74 \cdot e^{\pm j0.19\pi}$ (similar to benchmark system with transmission zeros).

Finally results for the following static nonlinearities are given

- $f_1(x) = \tanh(x)$ (saturation) and
- $f_2(x) = \mathrm{sinc}(x)$.

All results in this section are obtained with a training set of 500 samples. The model parameters are selected based on cross-validation and the reported performances are given are root mean squared error (RMSE) values for an independent test set with 1000 samples. The input process used to excite the system is Gaussian white noise (GWN) with unit variance.

The performance for two different combinations of input and output stages is shown in Figure 4. The method is able to successfuly identify a model with a moderately high order input filter as well as it is capable of dealing with nontrivial linear output systems including system with transmission zeros from the output of the nonlinearity to the measured signal.

In Figure 3 the reconstructed static nonlinearities are shown as obtained by the rank one approximation for the identity input stage G_1. The reconstruction works better for the sinc function than for the saturation charateristic of the tanh. For 1-st order Wiener systems like $G_2(z)$ the identified dynamical system can still be visualized and as can be seen in the bottom panel of Figure 3 is reconstructed correctly.

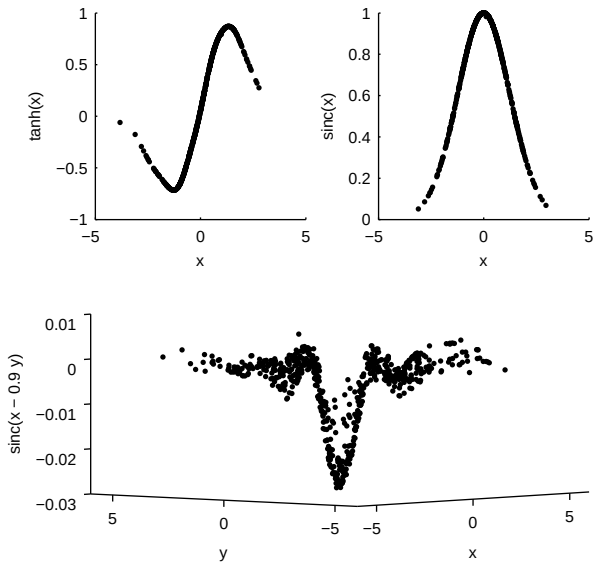


Fig. 3. The upper left panel shows the reconstructed values for the nonlinearity in case of $f_1(x)$, the upper right panel depicts $f_2(x)$. Both functions are enclosed by $G_1(q)$ and $H_1(1)$. The lower panel shows $f_2(x)$ enclosed by $G_2(q)$ and $H_1(1)$.

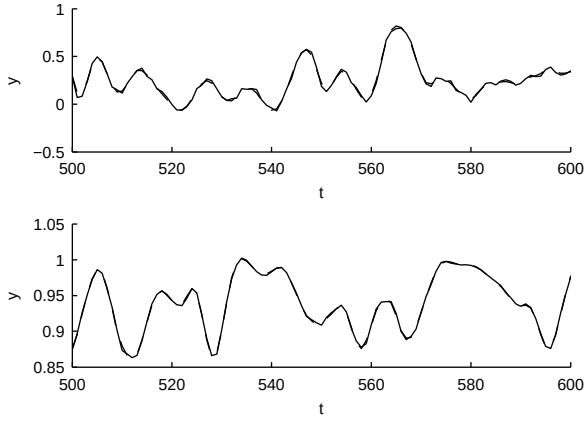


Fig. 4. The upper panel shows the system $G_2(q)$, $H_1(q)$ and $f_2(x)$ and the lower panel $G_3(q)$, $H_2(q)$ and $f_2(x)$. The solid lines are the true values whereas the dashed lines are the predictions.

3.3 Performance on the Benchmark Data

For the benchmark data a 10 training subsamples of 2000 points were trained and the best one selected for the final predictions. The models were selected using a validation set of 20,000 points. To assure that only stable models would enter the selection process, the performance evaluation is based on simulations instead of simple one-step ahead predictions. The optimal model order was found to be $D = 7$, $R_1 = 5$ and $R_2 = 7$.

The final performance on the full training set is $e_{RMSt} = 0.0117$ and on the independent evaluation set $e_{RMSe} = 0.0119$.

Preprint submitted to 15th IFAC Symposium on System Identification.
Received November 8, 2008.

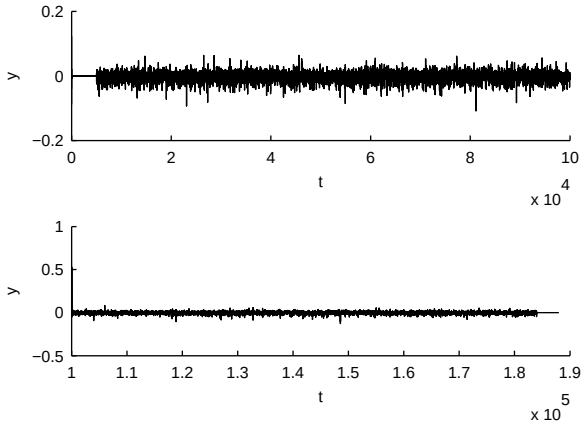


Fig. 5. Performance of the Wiener-Hammerstein model for the benchmark data. The upper panel shows the simulation residuals on the training set, whereas the lower panel shows the residuals for the evaluation set.

4. INFLUENCE OF INPUT DATA ON PREDICTION PERFORMANCE

According to the theory a system can only be identified if it is excited by an input signal that is persistenly exciting (PE) for this system (Söderström and Stoica [1989] and Stoica and Söderström [1982]). In practice this notion is not always necessary as (i) problems of ill-conditioning of the data makes the algorithm fail (e.g. for for numerical reasons), or (ii) the presence of noise renders the exact identification impossible. It is also not entirely clear which conditions on the signals are necessary or sufficient for proper approximation of the system, a question which is considerably weaker than exact identification. Approximation abilities are in general more appropriate for complex systems, where a (nonlinear) black- or grey-box approach is more plausible than coming up with models where a set of (linear) parameters are to be identified.

There are however a few attempts to extend the analysis of PE to block-structured models. Wiener-Hammerstein sytems are a subclass of Volterra filters for which Nowak and Van Veen [1994] have derived necessary and sufficent conditions for signals to be PE. These conditions rely on

(1) the spectrum,
(2) the record length and
(3) the amplitude richness

of the input signal. We will analyze the proposed method with respect to first two of these influences and compare it to

- a complete black box model (NARX LS-SVM),
- the intermediate componentwise model (Eq. (5)),
- linear models (estimated on unknown intermediate signals enclosing only the linear subsystems).

Unless stated otherwise we will consider the system consisting of the concatenation of $G_2(q)$, $f_3(x)$ and $H_2(z)$ in this section. The remaining experimental settings are a training set with 500 samples, GWN as input and noiseless data.

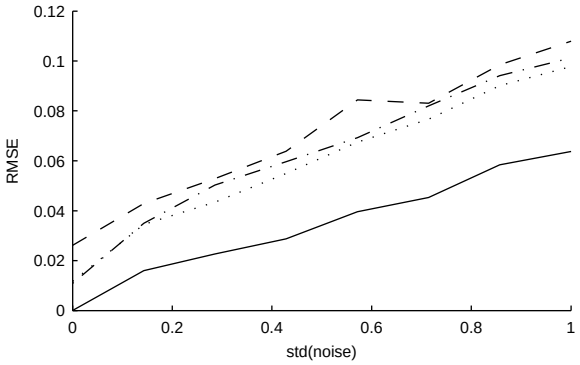


Fig. 6. Prediction performance in the presence of output noise. (solid line: linear model, dotted line: Wiener-Hammerstein model, dash-dotted line: componentwise model, dashed line: LS-SVM NARX model)

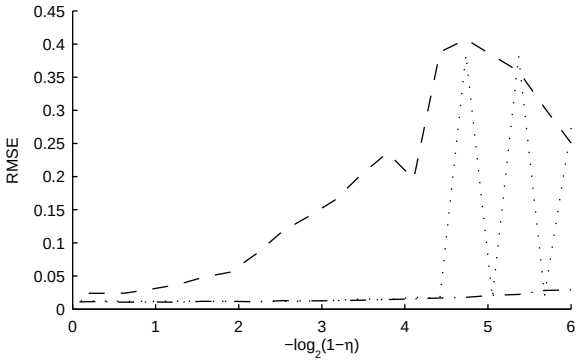
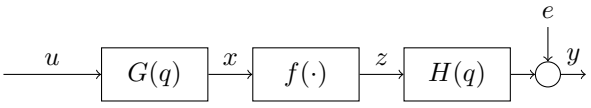


Fig. 7. Predicition performance for models that are train with colored inputs, the prediction performance is then evaluated for white inputs. (solid line: linear model, dotted line: Wiener-Hammerstein model, dash-dotted line: componentwise model, dashed line: LS-SVM NARX model)

The Hammerstein, the componentwise and the black box model are trained with $\{(u_t, y_t)\}$ as inputs. The linear models are trained with $\{(z_t, y_t)\}$.

4.1 Noise



In Figure 6 we show the prediction performance of the proposed method when the data is subject to output noise e. The difference between the component-wise model and the complete Wiener-Hammerstein approach are slim, but there is a substaintial gain over the black box model. The performance of the linear system is given as a reference.

4.2 Coloring

Identification is easiest for inputs with white spectra. In this section colored excitation signals are investigated. The colored input signal is generated by the low pass IIR filter $u_t = \eta u_{t-1} + n_t$ where n_t is a GWN process and η a control parameter to control the amount of coloring.

Preprint submitted to 15th IFAC Symposium on System Identification.
Received November 8, 2008.

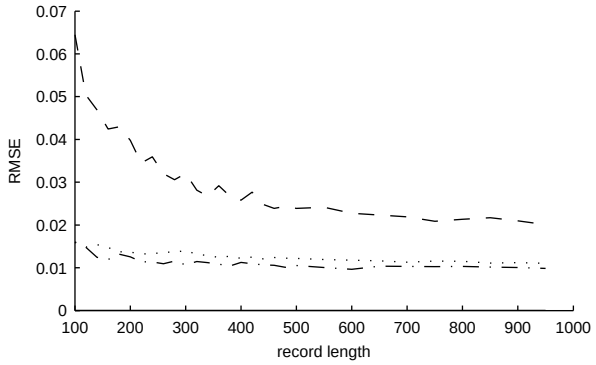


Fig. 8. Prediction performance for several model trained with different amounts of training data. (solid line: linear model, dotted line: Wiener-Hammerstein model, dash-dotted line: componentwise model, dashed line: LS-SVM NARX model)

The colored input signal is only used for the estimation phase. For the evalution of the approximation performance the system is again excited with a white input signal.

All models can effectively handle colored inputs during the trainig phase even for strongly colored inputs, although the black box model does not reach the same level of accuracy of the other methods. The capability to handle strongly colored inputs is due to the nonlinear function that introduces harmonics from the intermediate signal x to the signal z. Therefore the spectrum at the output of the nonlinearity can be better for linear identification than the original input.

4.3 Record Length

The amount of input data is critical for every data driven identification approach. Due to the effective parametrization of the Wiener-Hammerstein model, the estimation requires considerably less data than the estimation of an black-box model. The behavior of the different models for varying number training set sizes is shown in Figure 8

5. CONCLUSIONS

This paper studied an approach for identifying a Wiener-Hammerstein system from input/output data only. The approach integrates overparametrization with LS-SVMs, and extends (Goethals2005). The key idea is to jointly model the first linear dynamic susbsystem with the static nonlinearity. We then report numerical experiments on how the method behaves under various conditions, and give results on the SYSID2009 benchmark data.

Furthere work will be conducted with respect to a more throughly analysis of input signals and the incorporation of more training data. For the latter subsampling schemes and ensemble methods will be evaluated.

ACKNOWLEDGEMENTS

Research Council KUL: GOA AMBioRICS, CoE EF/05/006 Optimization in Engineering(OPTEC), IOF-SCORES4CHEM, several PhD/postdoc & fellow grants; Flemish Government: FWO: PhD/postdoc grants, projects G.0452.04 (new quantum algorithms), G.0499.04 (Statistics), G.0211.05 (Nonlinear), G.0226.06 (cooperative systems and optimization), G.0321.06 (Tensors), G.0302.07 (SVM/Kernel), G.0320.08 (convex MPC), G.0558.08 (Robust MHE), G.0557.08 (Glycemia2), research communities (ICCoS, ANMMM, MLDM); IWT: PhD Grants, McKnow-E, Eureka-Flite+; Helmholtz: viCERP; Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, Dynamical systems, control and optimization, 2007-2011); EU: ERNSI; FP7-HD-MPC (223854); AMINAL

Johan Suykens is a professor and Bart De Moor is a full professor at K.U. Leuven.

REFERENCES

E.-W. Bai. An optimal two stage identification algorithm for Hammerstein-Wiener nonlinear systems. In *Proc. American Control Conference the 1998*, volume 5, pages 2756–2760, 1998.

E.-W. Bai and M. Fu. A blind approach to Hammerstein model identification. *IEEE Transactions on Signal Processing*, 50(7):1610–1619, 2002.

N. J. Bershad, P. Celka, and S. McLaughlin. Analysis of stochastic gradient identification of Wiener-Hammerstein systems for nonlinearities with Hermite polynomial expansions. *IEEE Transactions on Signal Processing*, 49(5):1060–1072, 2001.

S. Bilings and S. Y. Fakhouri. Identification of a class of nonlinear systems using correlation analysis. *Proceedings of IEE*, 125(7):691–697, 1978.

F. Chang and R. Luus. A noniterative method for identification using Hammerstein model. *IEEE Transactions on Automatic Control*, 16(5):464–468, 1971.

M. Enqvist and L. Ljung. Linear approximations of nonlinear fir systems for separable input processes. *Automatica*, 41(3):459–473, Mar. 2005.

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

W. Greblicki and M. Pawlak. *Non-Parametric System Identification*. Cambridge University Press, 2008.

K. Narendra and P. Gallman. An iterative method for the identification of nonlinear systems using a Hammerstein model. *IEEE Transactions on Automatic Control*, 11(3): 546–550, 1966.

R. Nowak and B. Van Veen. Random and pseudorandom inputs for Volterra filter identification. *IEEE Transactions on Signal Processing*, 42(8):2124–2135, Aug. 1994.

M. Pawlak. On the series expansion approach to the identification of Hammerstein systems. *IEEE Transactions on Automatic Control*, 36(6):763–767, 1991.

B. Schölkopf and A. Smola. *Learning with Kernels*. MIT Press Cambridge, Mass, 2002.

J. Schoukens, J. A. K. Suykens, and L. Ljung. Wiener-hammerstein benchmark. Technical report, 2008.

T. Söderström and P. Stoica. *System Identification*. Prentice-Hall, 1989.

P. Stoica and T. Söderström. Instrumental-variable methods for identification of Hammerstein systems. *International Journal of Control*, 35(3):459–476, 1982.

J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle. *Least Squares Support Vector Machines*. World Scientific, 2002.

A. H. Tan and K. Godfrey. Identification of Wiener-Hammerstein models using linear interpolation in the frequency domain (LIFRED). *IEEE Transactions on Instrumentation and Measurement*, 51(3):509–521, 2002.

V. Vapnik. *Statistical Learning Theory*. John Wiley & Sons, 1998.

G. Wahba. *Spline Models for Observational Data*. SIAM, 1990.