Chapter 15
NARX Identification of Hammerstein Systems Using Least-Squares Support Vector Machines

Ivan Goethals, Kristiaan Pelckmans, Tillmann Falck,
Johan A.K. Suykens, and Bart De Moor

15.1 Introduction

This chapter describes a method for the identification of a SISO and MIMO Hammerstein systems based on Least Squares Support Vector Machines (LS-SVMs). The aim of this chapter is to give a practical account of the works [14] and [15], adding to this material new insights published since. The identification method presented in this chapter gives estimates for the parameters governing the linear dynamic block represented as an ARX model, as well as for the unknown static nonlinear function. The method is essentially based on Bai's overparametrisation technique, and combines this with a regularisation framework and a suitable model description which fits nicely within the LS-SVM framework with primal and dual model representations. This technique is found to cope effectively (i) with the ill-conditioning typically occurring in overparametrisation approaches, and (ii) with cases where no stringent assumptions can be made about the nature of the nonlinearity except for a certain degree of smoothness.

Consider the task of modelling the nonlinear dynamic relation between an *input* signal $(u_t \in \mathbb{R})_t$ and an *output* signal $(y_t \in \mathbb{R})_t$, both indexed over discrete time instants $t \in \mathbb{Z}$. The main body of the paper will be concerned with SISO Hammerstein

Ivan Goethals
ING Life Belgium, Sint Michielswarande 70, B-1040 Etterbeek, Belgium
e-mail: ivan.goethals@gmail.com

Kristiaan Pelckmans
Uppsala University, Department of Information Technology, division Syscon,
Box 337, SE-751 05, Uppsala, Sweden
e-mail: kp@it.uu.se

Tillmann Falck, Johan A.K. Suykens, and Bart De Moor
Katholieke Universiteit Leuven, ESAT-SCD-SISTA,
Kasteelpark 10, B-3001 Leuven (Heverlee), Belgium
e-mail: tillmann.falck, johan.suykens@esat.kuleuven.be,
bart.demoor@esat.kuleuven.be

F. Giri & E.-W. Bai (Eds.): Block-oriented Nonlinear System Identification, LNCIS 404, pp. 241–258.
springerlink.com © Springer-Verlag Berlin Heidelberg 2010

systems consisting of a (fixed but unknown) static nonlinearity $f : \mathbb{R} \rightarrow \mathbb{R}$ followed by a (fixed but unknown) ARX linear subsystem with a transfer function of orders $m > 0$ and $n > 0$ of the numerator and denominator respectively. The parameters of the ARX subsystem are denoted as $\omega = (a_1, \ldots, a_n, b_0, \ldots, b_m)^T \in \mathbb{R}^{n+m+1}$, assuming the following model

$$y_t = \sum_{i=1}^{n} a_i y_{t-i} + \sum_{j=0}^{m} b_j f(u_{t-j}) + e_t \quad , \forall t \in \mathbb{Z}. \tag{15.1}$$

The equation error e_t is assumed to be white and zero mean. An extension of the ideas developed in this chapter to systems exhibiting multiple inputs and outputs will be presented in Subsection 15.4.2.

For a general survey of different existing techniques for identification of Hammerstein systems, we refer to the relevant chapters in this book. In brief, identification of systems of the form (15.1) is often performed by describing the non-linear function f using a finite set of parameters and identifying those parameters together with the linear parameters ω. Parametric approaches found in the literature express the static non-linearity as a sum of (orthogonal or non-orthogonal) basis functions [22, 24, 25], as a finite number of cubic spline functions [9] or as a set of piecewise linear functions [32, 38]. Another form of parametrisation lies in the use of neural networks [19].

Regardless of the parametrisation scheme that is chosen, the final cost function will involve cross products between parameters describing the static nonlinearity f, and the parameters θ describing the linear dynamical system. Employing a maximum likelihood criterion results in a so-called bi-convex optimisation problem where global convergence is not guaranteed [28]. Several approaches have been proposed to solve the bi-convex optimisation problems typically encountered in Hammerstein system-identification, such as iterative approaches [24] and stochastic approaches [2, 5, 6, 7, 17, 38]. Moreover, in order to find a good optimum for these techniques, a proper initialisation is crucial [7] in practical applications.

In this chapter, we will focus on one particular approach, the so-called overparametrisation approach and demonstrate that the key ideas behind the overparametrisation approach can conveniently be combined with the method of Least Squares-Support Vector Machines regression to yield reliable ARX and even subspace identification algorithms for Hammerstein systems. The main practical benefits of this approach will include: (i) the increased (numerical) robustness due to the presence of a well-defined regularisation mechanism, (ii) the user is not required to restrict the form ('parametrisation') of the nonlinearity *a priori*, but reliable results will be obtained whenever the nonlinearity can be assumed to be 'smooth' (as will be defined later) to some degree and (iii) the framework of LS-SVMs will make it possible to confine the overparametrised model class more than was the case in [1]. Those advantages will support the claim of practical efficiency as illustrated on a number of case-studies in [14] and [15]. Since the publication of the works [14] and [15] some progress has been made towards more general model structures, other loss functions or recursive identification schemes for such models (see Section 15.6).

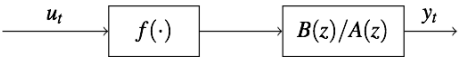


Fig. 15.1: General structure of a Hammerstein system, consisting of a static nonlinearity f and a linear subsystem with transfer function $B(z)/A(z)$

The outline of this chapter is as follows: in Section 15.2 the basic ideas behind overparametrisation are briefly reviewed. The use of LS-SVMs for static function estimation is described in Section 15.3. In Section 15.4 the ideas of Sections 15.2 and 15.3 are combined into the Hammerstein identification algorithm under study. Section 15.5 gives an illustrative example, and Section 15.6 will discuss extensions towards other block-structure models and towards a Hammerstein subspace identification algorithm. Section 15.7 gives concluding remarks.

15.2 Hammerstein Identification Using an Overparametrisation Approach

This section focuses on classical overparametrisation techniques applied in Hammerstein identification such as presented in [1, 22, 24, 25, 35, 37]. The key idea behind overparametrisation is to transform the bi-convex optimisation problem into a convex one by replacing every crossproduct of unknowns by new independent parameters [1, 4]. In a second stage the obtained solution is projected onto the Hammerstein model class. A technical implementation of this idea is presented here below.

15.2.1 Implementation of Overparametrisation

The idea of overparametrisation for Hammerstein systems is implemented here by substituting the product $b_j f(u_{t-j})$ in (15.1) by separate non-linear functions $g_j(u_{t-j})$ for all $j = 1,\ldots,m$. This results in the following overparametrised model:

$$y_t = \sum_{i=1}^{n} a_i y_{t-i} + \sum_{j=0}^{m} g_j(u_{t-j}) + e_t. \tag{15.2}$$

Note that this equation is linear in the parameters a_i, $i \in \{1,\ldots,n\}$ and the nonlinear functions g_j. When the $\{g_j\}_j$ are appropriately parametrised, (15.2) can be solved for a_i, $i \in \{1,\ldots,n\}$ and g_j, $j \in \{0,\ldots,m\}$ using an ordinary least squares approach. In a second stage, we are interested in recovering the parameters $b_{(m)} = (b_0,\ldots,b_m) \in \mathbb{R}^m$ from the estimated functions $\{\hat{g}_j : \mathbb{R} \to \mathbb{R}\}_j$. In order to do so, we concentrate on the function-values these functions take on the samples $(u_t)_t$, rather than on the functions themselves. This idea allows us to work further with tools from linear algebra, rather than working in a context of functional analysis. Hence, let the matrix $G \in \mathbb{R}^{(m+1)\times(N-2m-1)}$ be defined as

$$G = \begin{bmatrix} g_0(u_m) & \dots & g_0(u_{N-m}) \\ \vdots & & \vdots \\ g_m(u_m) & \dots & g_m(u_{N-m}) \end{bmatrix}, \tag{15.3}$$

and where $\hat{G} \in \mathbb{R}^{(m+1)\times(N-2m+1)}$ is the same matrix formed by the functions $\{\hat{g}_j\}_j$ estimated before. Now the key observation is that $G = b_{(m)} f_N^T$. here $f_N \in \mathbb{R}^{N-2m+1}$ is defined as $f_N = (f(u_m), \dots, f(u_{N-m}))^T$. Hence in the ideal case the matrix G is a rank-one matrix where the left- and right singular vector corresponding to the nonzero singular value, is proportional to f_N and $b_{(m)}$ respectively. The practical way to proceed is now to replace G by $\hat{G}$, and to use the best rank-one decomposition to give an estimate $\hat{b}_{(m)}$ of $b_{(m)}$. So far no specific parametrisation was assumed in the derivation above.

If the $m+1$ functions g_j have a common parametrisation, one can perform this projection in the parameter space instead as follows. A common parametrisation involves writing the original static non-linearity f in (15.1) as a linear combination of n_f general non-linear basis functions f_k, each with a certain weight c_k such that $f(u_t) = \sum_{k=1}^{n_f} c_k f_k(u_t)$. The functions f_1, f_2, and f_{n_f} are thereby chosen beforehand. Note that this amounts to parametrising the functions g_j in (15.2) as $g_j(u_t) = \sum_{k=1}^{n_f} \theta_{j,k} f_k(u_t)$, with $\theta_{j,k} = b_j c_k$. Hence, The original model (15.1) is rewritten as

$$y_t = \sum_{i=1}^{n} a_i y_{t-i} + \sum_{j=0}^{m} \sum_{k=1}^{n_f} b_j c_k f_k(u_{t-j}) + e_t \tag{15.4}$$

$$= \sum_{i=1}^{n} a_i y_{t-i} + \sum_{j=0}^{m} \sum_{k=1}^{n_f} \theta_{j,k} f_k(u_{t-j}) + e_t, \tag{15.5}$$

which can be solved for $\theta_{j,k}$, $j = 0, \dots, m$, $k = 1, \dots, n_f$ using *e.g.* a least squares algorithm. Denoting the estimates for $\theta_{j,k}$ by $\hat{\theta}_{j,k}$, estimates for the b_j and c_k are thereafter recovered from the SVD of:

$$\hat{\Theta} = \begin{bmatrix} \hat{\theta}_{0,1} & \hat{\theta}_{0,2} & \dots & \hat{\theta}_{0,n_f} \\ \hat{\theta}_{1,1} & \hat{\theta}_{1,2} & \dots & \hat{\theta}_{1,n_f} \\ \vdots & \vdots & & \vdots \\ \hat{\theta}_{m,1} & \hat{\theta}_{m,2} & \dots & \hat{\theta}_{m,n_f}, \end{bmatrix}. \tag{15.6}$$

15.2.2 Potential Problems in Overparametrisation

Estimating individual components in a sum of non-linearities is not without risks. Suppose for instance that $m = 1$, then Equation (15.2) can be rewritten as:

$$y_t = \sum_{i=1}^{n} a_i y_{t-i} + g_0(u_t) + g_1(u_{t-1}) + e_t \tag{15.7}$$

$$= \sum_{i=1}^{n} a_i y_{t-i} + g_0(u_t) + \delta + g_1(u_{t-1}) - \delta + e_t \tag{15.8}$$

$$= \sum_{i=1}^{n} a_i y_{t-i} + g_0'(u_t) + g_1'(u_{t-1}) + e_t, \tag{15.9}$$

with δ an arbitrary constant and $g_0'(u_t) = g_0(u_t) + \delta$, $g_1'(u_{t-1}) = g_1(u_{t-1}) - \delta$.

Similarly, note that for any set of variables $\varepsilon_k, k = 1, \ldots, n_f$ with $\forall u \in \mathbb{R}$, $\sum_{k=1}^{n_f} \varepsilon_k f_k(u) =$ Constant and any set $\alpha_j, j = 0, \ldots, m$ such that $\sum_{j=0}^{m} \alpha_j = 0$, $\theta'_{j,k} = \theta_{j,k} + \alpha_j \varepsilon_k$ is also a solution to (15.5) [18].

Hence, given a sequence of input/output measurements, all non-linearities estimated on these measurements will only be determined up to a set of constants. This problem is often overlooked in existing overparametrisation techniques and may lead to conditioning problems and destroy the low-rank property of (15.6). In fact, many published overparametrisation approaches applied to more complex Hammerstein systems lead to results which are far from optimal if no measures are taken to overcome this problem [14]. Following the parametric notation one possible solution is to calculate:

$$\hat{G} = \begin{bmatrix} \hat{\theta}_{0,1} & \hat{\theta}_{0,2} & \ldots & \hat{\theta}_{0,n_f} \\ \hat{\theta}_{1,1} & \hat{\theta}_{1,2} & \ldots & \hat{\theta}_{1,n_f} \\ \vdots & \vdots & & \vdots \\ \hat{\theta}_{m,1} & \hat{\theta}_{m,2} & \ldots & \hat{\theta}_{m,n_f} \end{bmatrix} \begin{bmatrix} f_1(u_1) & \ldots & f_1(u_N) \\ f_2(u_1) & \ldots & f_2(u_N) \\ \vdots & & \vdots \\ f_{n_f}(u_1) & \ldots & f_{n_f}(u_N) \end{bmatrix}, \tag{15.10}$$

subtract the mean of every row in $\hat{G}$ and take the SVD of the remaining matrix, from which estimates for the b_j can be extracted. Estimates for the c_k can then be found in a second round by solving (15.4). This identifiability issue is dealt with properly in the LS-SVM approach as described in the next section.

15.3 Function Approximation Using Least Squares Support Vector Machines

In this section, we review some elements of Least Squares Support Vector Machines (LS-SVMs) for static function approximation. The theory reviewed here will be extended to the estimation of Hammerstein systems in Section 15.4. This framework has strong connections to research on RBF- and regularisation networks, Gaussian processes, smoothing splines and dual ridge regression amongst others, see *e.g.* [30] for a thorough overview.

Let $\{(x_t, y_t)\}_{t=1}^{N} \subset \mathbb{R}^d \times \mathbb{R}$ be a set of input/output training data (x_t, y_t) with an input $x_t \in \mathbb{R}^d$ and output $y_t \in \mathbb{R}$. Consider the regression model $y_t = f(x_t) + e_t$ where $x_1, \ldots, x_N$ are deterministic points, $f : \mathbb{R}^d \to \mathbb{R}$ is an unknown real-valued function and $e_1, \ldots, e_N$ are uncorrelated random errors with $E[e_t] = 0$, $E[e_t^2] = \sigma_e^2 < \infty$. In recent years, Support Vector Machines (SVMs) [33] have been used for the purpose of estimating the non-linear f. The following model is assumed:

$$f(x) = w^T \varphi(x) + b,$$

where $\varphi(x) : \mathbb{R}^d \rightarrow \mathbb{R}^{n_H}$ denotes a potentially infinite ($n_H = \infty$) dimensional feature map, $w \in \mathbb{R}^{n_H}$, $b \in \mathbb{R}$. The regularised cost function of the Least Squares SVM (LS-SVM) [30] is given as

$$\begin{aligned} \min_{w,b,e} \mathscr{J}(w,e) &= \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{t=1}^{n} e_t^2, \\ \text{subject to : } y_t &= w^T \varphi(x_t) + b + e_t, \quad t = 1,\ldots,N. \end{aligned}$$

The relative importance between the smoothness of the solution and the data fitting is governed by the scalar $\gamma \in \mathbb{R}_0^+$, referred to as the regularisation constant. The optimisation performed corresponds to ridge regression [16] in feature space. In order to solve the constrained optimisation problem, a Lagrangian is constructed:

$$\mathscr{L}(w,b,e;\alpha) = \mathscr{J}(w,e) - \sum_{t=1}^{N} \alpha_t \{w^T \varphi(x_t) + b + e_t - y_t\},$$

with α_t for $t = 1,\ldots,N$ the Lagrange multipliers. The conditions for optimality are given as:

$$\frac{\partial \mathscr{L}}{\partial w} = 0 \quad \rightarrow \quad w = \sum_{t=1}^{N} \alpha_t \varphi(x_t), \tag{15.11}$$

$$\frac{\partial \mathscr{L}}{\partial b} = 0 \quad \rightarrow \quad \sum_{t=1}^{N} \alpha_t = 0, \tag{15.12}$$

$$\frac{\partial \mathscr{L}}{\partial e_t} = 0 \quad \rightarrow \quad \alpha_t = \gamma e_t, \quad t = 1,\ldots,N, \tag{15.13}$$

$$\frac{\partial \mathscr{L}}{\partial \alpha_t} = 0 \quad \rightarrow \quad y_t = w^T \varphi(x_t) + b + e_t, \quad t = 1,\ldots,N. \tag{15.14}$$

Substituting (15.11)-(15.13) into (15.14) yields the following dual problem (*i.e.* the problem in the Lagrange multipliers):

$$\left[\begin{array}{c|c} 0 & 1_N{}^T \\ \hline 1_N & \Omega + \gamma^{-1} I_N \end{array}\right] \left[\begin{array}{c} b \\ \hline \alpha \end{array}\right] = \left[\begin{array}{c} 0 \\ \hline y \end{array}\right], \tag{15.15}$$

where $y = \begin{bmatrix} y_1 & \ldots & y_N \end{bmatrix}^T \in \mathbb{R}^N$, $1_N = \begin{bmatrix} 1 & \ldots & 1 \end{bmatrix}^T \in \mathbb{R}^N$, $\alpha = \begin{bmatrix} \alpha_1 & \ldots & \alpha_N \end{bmatrix}^T \in \mathbb{R}^N$, and the matrix $\Omega \in \mathbb{R}^{N \times N}$ where $\Omega_{ij} = K(x_i,x_j) = \varphi(x_i)^T \varphi(x_j)$, $\forall i,j = 1,\ldots,N$, with K the positive definite kernel function. Note that in order to solve the set of equations (15.15), the feature map φ does never have to be defined explicitly. Only its inner product, a positive definite Mercer kernel, is needed. This is called the kernel trick [27, 33]. For the choice of the kernel $K(\cdot,\cdot)$, see *e.g.* [27]. Typical examples are the use of a linear kernel $K(x_i,x_j) = x_i^T x_j$, a polynomial kernel $K(x_i,x_j) = (\tau + x_i^T x_j)^d, \tau \geq 0$ of degree d or an RBF kernel $K(x_i,x_j) = \exp(-\|x_i - x_j\|_2^2/\sigma^2)$

where σ denotes the bandwidth of the kernel. The resulting LS-SVM model for function estimation can be evaluated at a new point x_* as

$$\hat{f}(x_*) = \sum_{t=1}^{N} \alpha_t K(x_*, x_t) + b,$$

where (b, α) is the solution to (15.15). Note that in the above, no indication is given as to how to choose free parameters such as the regularisation constant γ and the bandwidth σ in an RBF kernel. These parameters, which are generally referred to as *hyper-parameters* will have to be obtained from data, *e.g.* by tuning on an independent validation dataset, or by using cross-validation [18].

Besides the function estimation case, the class of LS-SVMs also includes classification, kernel PCA (principal component analysis), kernel CCA, kernel PLS (partial least squares), recurrent networks and solutions to non-linear optimal control problems. For an overview on applications of the LS-SVM framework, the reader is referred to [30] and citations.

15.4 NARX Hammerstein Identification as a Componentwise LS-SVM

The key advantage of the use of overparametrisation as introduced in Section 15.2 is the particularly attractive convexity property. An essential problem with the overparametrisation approach however is the increased variance of the estimates due to the increased number of unknowns in the first stage.

In this section, we will demonstrate that the ideas behind the overparametrisation approach can conveniently be combined with the method of LS-SVM regression which (i) features an inherent regularisation framework to deal with the increased number of unknowns, and (ii) enables one to deal properly with the identifiability issue discussed in Subsection 2.2. For instructive purposes, we will again focus on systems in SISO form and deal with the extension to MIMO systems later.

15.4.1 SISO Systems

In line with LS-SVM function approximation, we replace every function $g_j(u_{t-j})$ in (15.2) by $w_j^T \varphi(u_{t-j})$ with $\varphi : \mathbb{R} \to \mathbb{R}^{n_H}$ a fixed, potentially infinite ($n_H = \infty$) dimensional feature map and for every $j = 0, \ldots, m$, $w_j \in \mathbb{R}^{n_H}$. Adding an additional constant d, the reason of which will become clear below, the overparametrised model in Equation (15.2) can be written as

$$y_t = \sum_{i=1}^{n} a_i y_{t-i} + \sum_{j=0}^{m} w_j^T \varphi(u_{t-j}) + d + e_t. \tag{15.16}$$

With $r = max(m, n) + 1$, the regularised cost function of LS-SVM is given as:

$$\min_{w_j,a,d,e} \mathscr{J}(w_j,e) = \frac{1}{2}\sum_{j=0}^{m} w_j^T w_j + \gamma\frac{1}{2}\sum_{t=r}^{N} e_t^2, \tag{15.17}$$

subject to

$$\sum_{j=0}^{m} w_j^T \varphi(u_{t-j}) + \sum_{i=1}^{n} a_i y_{t-i} + d + e_t - y_t = 0, \quad t = r,\ldots,N, \tag{15.18}$$

$$\sum_{k=1}^{N} w_j^T \varphi(u_k) = 0, \quad j = 0,\ldots,m. \tag{15.19}$$

The problem (15.17)-(15.19), is known as a component-wise LS-SVM regression problem and was described in [26], and may be traced back to earlier research on additive models using smoothing splines, see *e.g.* [36] and citations. The term component-wise refers to the fact that the output is ultimately written as the sum of a set of linear and non-linear components. As will be seen shortly, the derivation of a solution to a component-wise LS-SVM problem follows the same approach within the LS-SVM setting with primal and dual model representations.

Note the additional constraints (15.19) to centre the non-linear functions $w_j^T\varphi(\cdot)$ for all $j = 0,\ldots,m$ around their average over the training set. These constraints resolve the identifiability issue described in Subsection 15.2.2: they remove the uncertainty resulting from the fact that any set of constants can be added to the terms of the additive non-linear function (15.16), as long as the sum of the constants is zero. Observe that the constraints (15.19) correspond to removing the mean of every row in $\hat{G}$ in (15.10). Removing the mean will facilitate the extraction of the parameters b_j in (15.1) later. Furthermore, the constraints enable us to give a clear meaning to the bias parameter d, namely $d = \sum_{j=0}^{m} b_j \left(\frac{1}{N}\sum_{k=1}^{N} f(u_k)\right)$. Now it is seen that inclusion of constraints of the form (15.19) can naturally be included in the objective function of the LS-SVM, avoiding the need for a separate normalisation step in the identification procedure.

Lemma 15.1. *Let again $r = max(m,n)+1$. Given the system (15.16), the LS-SVM estimates for the non-linear functions $w_j^T\varphi : \mathbb{R} \to \mathbb{R}$, $j = 0,\ldots,m$, are given as:*

$$w_j^T\varphi(u_*) = \sum_{t=r}^{N} \alpha_t K(u_{t-j},u_*) + \beta_j \sum_{t=1}^{N} K(u_t,u_*) \tag{15.20}$$

where the parameters $\alpha_t, t = r,\ldots,N$, $\beta_j, j = 0,\ldots,m$, as well as the linear model parameters $a_i, i = 1,\ldots,n$ and d are obtained from the following set of linear equations:

$$\left[\begin{array}{c|c|c|c} 0 & 0 & 1^T & 0 \\ \hline 0 & 0 & \mathscr{Y}_p & 0 \\ \hline 1 & \mathscr{Y}_p^T & \mathscr{K}+\gamma^{-1}I & K^0 \\ \hline 0 & 0 & {K^0}^T & (1_N^T\Omega 1_N)I_{m+1} \end{array}\right] \left[\begin{array}{c} d \\ \hline a \\ \hline \alpha \\ \hline \beta \end{array}\right] = \left[\begin{array}{c} 0 \\ \hline 0 \\ \hline \mathscr{Y}_f \\ \hline 0 \end{array}\right], \tag{15.21}$$

with

$$a = \begin{bmatrix} a_1 & \dots & a_n \end{bmatrix}^T, \tag{15.22}$$

$$\beta = \begin{bmatrix} \beta_0 & \dots & \beta_m \end{bmatrix}^T, \tag{15.23}$$

$$K^0(p,q) = \sum_{t=1}^{N} \Omega_{t,r+p-q} = \sum_{t=1}^{N} K(u_t, u_{r+p-q}), \tag{15.24}$$

$$\mathscr{K}(p,q) = \sum_{j=0}^{m} \Omega_{p+r-j-1,q+r-j-1} = \sum_{j=0}^{m} K(u_{p+r-j-1}, u_{q+r-j-1}), \tag{15.25}$$

$$\mathscr{Y}_p = \begin{bmatrix} y_{r-1} & y_r & \dots & y_{N-1} \\ y_{r-2} & y_{r-1} & \dots & y_{N-2} \\ \vdots & \vdots & & \vdots \\ y_{r-n} & y_{r-n+1} & \dots & y_{N-n} \end{bmatrix}, \tag{15.26}$$

$$\mathscr{Y}_f = \begin{bmatrix} y_{r+1} & \dots & y_N \end{bmatrix}^T, \tag{15.27}$$

and 1_N *is a column vector of length N with elements* 1.

Proof. The proof is found in [14]. □

Note that the matrix $\mathscr{K}$, which appears on the left hand side of (15.21) and plays a similar role as the kernel matrix Ω in (15.15), actually represents a sum of kernels in (15.21). This follows as a typical property of the solution of component-wise LS-SVM problems [26]. It is instrumental that ill-conditioning only arises if the matrix $\mathscr{Y}_p$ has zero singular values. That is, the regularisation term $\gamma^{-1} I_N$ will avoid ill-conditioning even if $\mathscr{K}$ is singular, as would be the case when the signal $(u_t)_t$ is constant (and not persistently exciting of any order).

Projecting onto the class of ARX Hammerstein models

The projection of the obtained model onto (15.1) goes as follows. Estimates for the autoregressive parameters $a_i, i = 1, \dots, n$ are directly obtained from (15.21). Furthermore, for the training input sequence $\begin{bmatrix} u_1 & \dots & u_N \end{bmatrix}$, we have:

$$\begin{bmatrix} b_0 \\ \vdots \\ b_m \end{bmatrix} \begin{bmatrix} \hat{\underline{f}}(u_1) \\ \vdots \\ \hat{\underline{f}}(u_N) \end{bmatrix}^T = \begin{bmatrix} \alpha_N & \dots & \alpha_r & & 0 \\ & \alpha_N & \dots & \alpha_r & \\ & & \ddots & & \ddots \\ 0 & & & \alpha_N & \dots & \alpha_r \end{bmatrix}$$
$$\times \begin{bmatrix} \Omega_{N,1} & \Omega_{N,2} & \dots & \Omega_{N,N} \\ \Omega_{N-1,1} & \Omega_{N-1,2} & \dots & \Omega_{N-1,N} \\ \vdots & \vdots & & \vdots \\ \Omega_{r-m,1} & \Omega_{r-m,2} & \dots & \Omega_{r-m,N} \end{bmatrix} + \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_m \end{bmatrix} \sum_{t=1}^{N} \begin{bmatrix} \Omega_{t,1} \\ \vdots \\ \Omega_{t,N} \end{bmatrix}^T, \tag{15.28}$$

with $\hat{\underline{f}}(u)$ an estimate for

$$\underline{f}(u) = f(u) - \frac{1}{N}\sum_{t=1}^{N} f(u_t). \tag{15.29}$$

Hence, estimates for b_j and the static non-linearity f evaluated in $\{u_t\}_{t=1}^N$ can be obtained from a rank 1 approximation of the right hand side of (15.28), for instance using a singular value decomposition. Again, this is the equivalent of the SVD-step that is generally encountered in overparametrisation methods [1, 4]. Once all the elements b_j are known, $\sum_{k=1}^N f(u_k)$ can be obtained as $\sum_{k=1}^N f(u_k) = \frac{Nd}{\sum_{j=0}^m b_j}$. In a second step, a parametric estimation of f can be obtained by applying classical LS-SVM function approximation on the couples $\{(u_t, \hat{f}(u_t))\}_{t=1}^N$.

Algorithm 15.1. LS-SVM Hammerstein identification – a SISO algorithm

1. Choose a kernel K and regularisation constant γ
2. Calculate the componentwise kernel matrix $\mathscr{K}$ as a sum of individual kernel matrices as described in (15.25)
3. Solve (15.21) for d, a, α, β
4. Apply (15.21) to a validation set or use cross-validation. Go to step 1 and change kernel parameters and or γ until optimal performance is obtained
5. Take the SVD of the right hand side of (15.28) to determine the linear parameters $b_0, \ldots, b_m$
6. Obtain estimates $\{\hat{f}(u_t)\}_{t=1}^N$ from (15.28) and (15.29)
7. If a parametric estimate for f is needed, apply LS-SVM function estimation on $\{(u_t, \hat{f}(u_t))\}_{t=1}^N$

15.4.2 Identification of Hammerstein MIMO Systems

Conceptually, an extension of the method presented in the previous section towards the MIMO case is straightforward, but the calculations involved are quite extensive. Assuming a MIMO Hammerstein system of the form:

$$y_t = \sum_{i=1}^{n} A_i y_{t-i} + \sum_{j=0}^{m} B_j f(u_{t-j}) + e_t, \tag{15.30}$$

with $y_t, e_t \in \mathbb{R}^{n_y}$, $u_t \in \mathbb{R}^{n_u}$, $A_i \in \mathbb{R}^{n_y \times n_y}$, $B_j \in \mathbb{R}^{n_y \times n_u}$, $t = 1, \ldots, N$, $i = 1, \ldots, n$, $j = 0, \ldots, m$, and $f : \mathbb{R}^{n_u} \to \mathbb{R}^{n_u} : u \to f(u) = \begin{bmatrix} f_1(u) & \ldots & f_{n_u}(u) \end{bmatrix}^T$, we have for every row s in (15.30), that

$$y_t(s) = \sum_{i=1}^{n} A_i(s,:) y_{t-i} + \sum_{j=0}^{m} B_j(s,:) f(u_{t-j}) + e_t(s). \tag{15.31}$$

Note that for every non-singular matrix $V \in \mathbb{R}^{n_u \times n_u}$, and for any $j = 0, \ldots, m$:

$$B_j(s,:) f(u_{t-j}) = B_j(s,:) V V^{-1} f(u_{t-j}), \tag{15.32}$$

with $B_j(s,:)$ denoting row s in the matrix B_j. Hence, any model of the form (15.30) can be replaced with an equivalent model by applying a linear transformation on the components of f and the columns of B_j. This will have to be taken into account when identifying models of the form (15.30) without any prior knowledge of the non-linearity involved.

Substituting $f(u) = \begin{bmatrix} f_1(u) & \ldots & f_{n_u}(u) \end{bmatrix}^T$ in (15.31) leads to:

$$y_t(s) = \sum_{i=1}^{n} A_i(s,:) y_{t-i} + \sum_{j=0}^{m} \sum_{k=1}^{n_u} B_j(s,k) f_k(u_{t-j}) + e_t(s). \tag{15.33}$$

By replacing $\sum_{k=1}^{n_u} B_j(s,k) f_k(u_{t-j})$ by $w_{j,s}^T \varphi(u_{t-j}) + d_{s,j}$ this reduces to

$$y_t(s) = \sum_{i=1}^{n} A_i(s,:) y_{t-i} + \sum_{j=0}^{m} \omega_{j,s}^T \varphi(u_{t-j}) + d_s + e_t(s). \tag{15.34}$$

where

$$d_s = \sum_{j=0}^{m} d_{s,j}. \tag{15.35}$$

The primal problem that is subsequently obtained is the following:

$$\min_{\omega_{j,s},e} \mathscr{J}(\omega_{j,s}, e) = \sum_{j=0}^{m} \sum_{s=1}^{n_y} \frac{1}{2} \omega_{j,s}^T \omega_{j,s} + \sum_{s=1}^{n_y} \sum_{t=r}^{N} \frac{\gamma_s}{2} e_t(s)^2. \tag{15.36}$$

subject to (15.34) and $\sum_{k=1}^{N} w_{j,s}^T \varphi(u_k) = 0, \quad j = 0, \ldots, m, \quad s = 1, \ldots, n_y$.

Lemma 15.2. *Given the primal problem (15.36), the LS-SVM estimates for the non-linear functions $w_{j,s}^T \varphi : \mathbb{R} \to \mathbb{R}$, $j = 0, \ldots, m$, $s = 1, \ldots, n_y$, are given as:*

$$w_{j,s}^T \varphi(u_*) = \sum_{t=r}^{N} \alpha_{t,s} K(u_{t-j}, u_*) + \beta_{j,s} \sum_{t=1}^{N} K(u_t, u_*), \tag{15.37}$$

where the parameters $\alpha_{t,s}, t = r, \ldots, N$, $s = 1, \ldots, n_y$, $\beta_{j,s}, j = 0, \ldots, m$, $s = 1, \ldots, n_y$ as well as the linear model parameters $A_i, i = 1, \ldots, n$ and d_s, $s = 1, \ldots, n_y$ are obtained from the following set of linear equations:

$$\begin{bmatrix} L_1 & & \\ & \ddots & \\ & & L_{n_y} \end{bmatrix} \begin{bmatrix} X_1 \\ \vdots \\ X_{n_y} \end{bmatrix} = \begin{bmatrix} R_1 \\ \vdots \\ R_{n_y} \end{bmatrix}, \tag{15.38}$$

where

$$L_s = \left[\begin{array}{c|c|c|c} 0 & 0 & 1^T & 0 \\ \hline 0 & 0 & \mathscr{Y}_p & 0 \\ \hline 1 & \mathscr{Y}_p^T & \mathscr{K}+\gamma_s^{-1}I & \mathscr{S} \\ \hline 0 & 0 & \mathscr{S}^T & \mathscr{T} \end{array}\right], \quad X_s = \left[\begin{array}{c} d_s \\ \hline \mathscr{A}_s \\ \hline \overline{\alpha}_s \\ \hline \overline{\beta}_s \end{array}\right], \tag{15.39}$$

$$R_s = \left[\begin{array}{c|c|c|c} 0 & 0 & \mathscr{Y}_{f,s}^T & 0 \end{array}\right]^T, \quad \mathscr{Y}_{f,s} = \begin{bmatrix} y_r(s)^T & \dots & y_N(s)^T \end{bmatrix}^T, \tag{15.40}$$

$$\mathscr{A}_s = \begin{bmatrix} A_1(s,:)^T \\ \vdots \\ A_m(s,:)^T \end{bmatrix}, \quad \overline{\alpha}_s = \begin{bmatrix} \alpha_{r,s} \\ \vdots \\ \alpha_{N,s} \end{bmatrix}, \quad \mathscr{S}(p,q) = \sum_{t=1}^{N} \Omega_{t,r+p-q}, \tag{15.41}$$

$$\overline{\beta}_s = \begin{bmatrix} \beta_{0,s} & \dots & \beta_{m,s} \end{bmatrix}^T, \quad \Omega_{p,q} = \varphi(u_p)^T \varphi(u_q), \tag{15.42}$$

$$\mathscr{K}(p,q) = \sum_{j=0}^{m} \Omega_{p+r-j-1,q+r-j-1}, \quad \mathscr{T} = 1_N^T \Omega 1_N \cdot I_{m+1}. \tag{15.43}$$

Proof. The proof is found in [14]. □

Note that the matrices $L_s, s = 1,\dots,n_y$ in (15.38) are almost identical, except for the different regularisation constants γ_s. In many practical cases, however, and if there is no reason to assume that a certain output is more important than another, it is recommended to set $\gamma_1 = \gamma_2 = \dots = \gamma_{n_y}$. This will speed up the estimation algorithm since $L_1 = L_2 = \dots = L_{n_y}$ needs to be calculated only once, but most importantly, it will reduce the number of hyper-parameters to be tuned.

Projection onto the class of ARX Hammerstein models

The projection of the obtained model onto (15.33) is similar as in the SISO case. Estimates for the autoregressive matrices $A_i, i = 1,\dots,n$ are directly obtained from (15.38). For the training input sequence $\begin{bmatrix} u_1 & \dots & u_N \end{bmatrix}$ and every $k = 1,\dots,n_u$, we have:

$$\begin{bmatrix} B_0(1,:) \\ \vdots \\ \hline B_m(1,:) \\ \vdots \\ \hline B_0(n_y,:) \\ \vdots \\ B_m(n_y,:) \end{bmatrix} \begin{bmatrix} \underline{\hat{f}}^T(u_1) \\ \vdots \\ \underline{\hat{f}}^T(u_N) \end{bmatrix}^T = \begin{bmatrix} \beta_{0,1} \\ \vdots \\ \hline \beta_{m,1} \\ \vdots \\ \hline \beta_{0,n_y} \\ \vdots \\ \beta_{m,n_y} \end{bmatrix} \sum_{t=1}^{N} \begin{bmatrix} \Omega_{t,1} \\ \vdots \\ \Omega_{t,N} \end{bmatrix}^T + \begin{bmatrix} \mathscr{A}_1 \\ \hline \vdots \\ \hline \mathscr{A}_{n_y} \end{bmatrix} \times \begin{bmatrix} \Omega_{N,1} & \Omega_{N,2} & \dots & \Omega_{N,N} \\ \Omega_{N-1,1} & \Omega_{N-1,2} & \dots & \Omega_{N-1,N} \\ \vdots & \vdots & & \vdots \\ \Omega_{r-m,1} & \Omega_{r-m,2} & \dots & \Omega_{r-m,N} \end{bmatrix}, \tag{15.44}$$

with $\underline{\hat{f}}(u)$ an estimate for

$$\underline{f}(u) = f(u) - g, \tag{15.45}$$

and g a constant vector such that:

$$\sum_{j=0}^{m} B_j g = \begin{bmatrix} d_1 \\ \vdots \\ d_{n_y} \end{bmatrix}. \tag{15.46}$$

Estimates for $\underline{f}$ and the $B_j, j = 0, \ldots, m$, can be obtained through a rank-n_u approximation of the right hand side of (15.44). If a singular value decomposition is used, the resulting columns of the left hand side matrix of (15.44) containing the elements of $B_j, j = 0, \ldots, m$, can be made orthonormal, effectively fixing the choice of V in (15.32).

From estimates for $\underline{f}$ in (15.45) and g in (15.46), finally, an estimate for the nonlinear function f can be obtained. Note that if the row-rank of $\sum_{j=0}^{m} B_j$ is smaller than the column-rank, multiple choices for g are possible. This results as an inherent property of blind MIMO Hammerstein identification. The choice of a particular g is left to the user.

Algorithm 15.2. LS-SVM Hammerstein identification – a MIMO algorithm

1. Choose a kernel K and regularisation constants $\gamma = \gamma_1 = \ldots = \gamma_{n_y}$
2. Calculate the componentwise kernel matrix $\mathscr{K}$ as a sum of individual kernel matrices as described in (15.43)
3. Solve (15.38) for $\alpha, \beta, d_1, \ldots, d_{n_y}$ and $A_1, \ldots, A_n$
4. Apply (15.38) to a validation set or use cross-validation. Go to step 1 and change kernel parameters and or γ until optimal performance is obtained
5. Take the SVD (rank-n_u approximation) of the right hand side of (15.44) to determine the linear parameters $B_0, \ldots, B_m$
6. Obtain estimates $\{\hat{f}(u_t)\}_{t=1}^{N}$ from (15.38), (15.45) and (15.46)
7. If a parametric estimate for f is needed, apply LS-SVM function estimations on $\{(u_t, \hat{f}_k(u_t))\}_{t=1}^{N}, k = 1, \ldots, n_y$

15.5 Example

This section illustrates the importance of the centring constraints and the careful selection of model parameters. Therefore consider the following example; $f(u) = \mathrm{sinc}(u)u^2$ is the true nonlinearity and the linear subsystem is of 6^{th} order. The linear subsystem is described by $A(z) = (z - 0.98e^{\pm i})(z - 0.98e^{\pm 1.6i})(z - 0.97e^{\pm 0.4i})$ and $B(z) = (z - 0.2)(z + 0.94)(z - 0.93e^{\pm 0.7i})(z - 0.96e^{\pm 1.3i})$. The importance of the centring constraints is most visible for systems with "challenging numerators", in this example characterised by system zeros close to the unit circle.

In contrast to the examples presented in [14] and [15] a minimal number of data points is used to illustrate the importance of proper tuning of the hyper-parameters

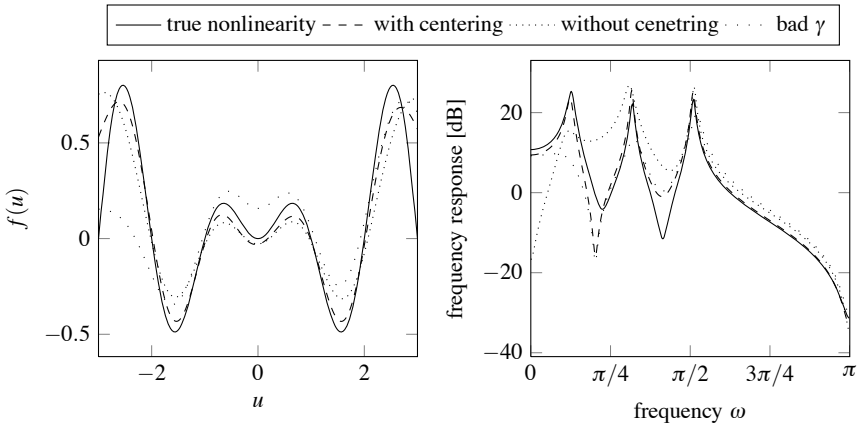


Fig. 15.2: Reconstruction of true nonlinearity $f(u) = \mathrm{sinc}(u)u^2$ (left panel, solid line) and 6th order linear subsystem (right panel, solid line) by a LS-SVM model with centring constraints (dashed line, $\sigma = 0.85$, $\gamma = 167$) and without centring constraints (dotted line, $\sigma = 1.17$, $\gamma = 13.7$). Additionally the reconstruction for a model with a badly chosen regularisation constant ($\sigma = 0.85$, $\gamma = 0.167$) is shown by the loosely dotted line. All models are estimated from 95 samples (circles) with additive white Gaussian output noise of variance 0.1^2. The model, with centring constraints and carefully selected regularisation constant, estimates the nonlinearity as well as the linear subsystem better than the other models

and the use of centring constraints. The true system is simulated for 190 time steps and white Gaussian noise with unit variance as input. The output is subject to additive white Gaussian noise with variance 0.1^2. Half of the samples are used as a validation set for model selection that is carried out using a grid search. The hyperparameters that need to selected are the regularisation constant γ and the bandwidth of the RBF kernel σ.

For illustration purposes we performed the algorithm outlined at the end of Section 15.4.1 twice, once with centring constraints and once without. Figure 15.2 shows the reconstruction of the true nonlinearity and the linear subsystem as obtained for both models. Additionally the reconstruction resulting from a badly tuned model with centring constraints is shown. It can be seen that the reconstruction using the properly tuned model with centring constraints, is substantially better than the one obtained using a model without centring constraints or badly chosen model parameters.

15.6 Extensions

Various extensions to the concepts introduced in Section 15.4 exist. We provide a brief summary below and refer the reader to the included references for further details:

- **Subspace identification:** The ARX model class is a popular one due to (i) its simple structure and (ii) the fact that its parameters can be estimated conveniently as an ordinary least squares problem. Nevertheless, ARX models are not suited for the identification of linear dynamical systems under certain experimental conditions such as the presence of heavily coloured noise on the outputs. As a result, an extension of the concepts presented in Section 15.4 to the broader and more robust class of subspace identification algorithms is desirable. Such an extension was presented in [15] and is based on the observation that the oblique projection that features in most linear subspace identification algorithms in one way or another [31] can be written as a least squares optimisation problem, not unlike the one encountered in linear ARX identification. As in the ARX case, adding a static non-linearity f ultimately transforms this least squares optimisation problem into a bi-convex optimisation that can be solved using a componentwise LS-SVM. In [20] this has been further extended to closed loop measurements.

- **Recursive subspace identification:** Over the last few years, various recursive versions of linear subspace identification algorithms have been presented [12, 21, 23]. In [3], it is shown that the Hammerstein subspace identification algorithm presented in [14] can also be transformed into recursive form, allowing for its use in on-line applications.

- **Identification of Block-structure Models:** The ideas behind the above described approach can be readily extended towards identification of more general block-structured models. The identification of Hammerstein–Wiener systems with invertible output non-linearity was described in [13]. This result is based on the observation that a so-called subspace intersection algorithms, a realisation for the internal states of a linear system is obtained as the intersection of a space spanned by measured inputs and outputs. Numerically, the intersection can be calculated using a Canonical Correlation Analysis, which in turn can be extended towards the kernel equivalent, using the so-called KCCA algorithm, see *e.g.* [34]. Identification of General Wiener–Hammerstein models was described in [10], using a slight extension of the overparametrisation technique.

- **Large-Scale Problems:** Extensions using fixed-size kernel methods for large datasets described in [30] were used to extend the kernel-based approach towards a method able to deal with $O(10^6)$ number of training data points, see *e.g.* [10] or [8].

15.7 Outlook

This chapter gave an account of an identification method for Hammerstein systems integrating kernel methods with Bai's overparametrisation technique. Illustrations of this technique on real data can be found in *e.g.* [14, 11] and [10, 8]. While the method

is not exploiting any assumption on the inputs (like whiteness) directly, the influence of persistency of excitation is not well understood in such approaches (see *e.g.* [29] for the specific case where polynomial basis functions were used). However, regularisation is found to deal effectively with the lack of persistency. A thorough theoretical understanding of this observation is still missing. A second question around this approach is the overall asymptotic performance (including bias, consistency or variance expressions). The main difficulty is that the overparametrisation technique in general lacks a global objective function. As a result the necessary conditions for the method to work well are not fully established. From a more applied perspective, the extension to identification of Wiener structures is only covered in special cases and needs more work.

Acknowledgements. Tillmann Falck, Johan Suykens and Bart De Moor are supported by Research Council KUL: GOA AMBioRICS, GOA MaNet, 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), G.0588.09 (Brain-machine); research communities (ICCoS, ANMMM, MLDM); G.0377.09 (Mechatronics MPC); IWT: PhD Grants, McKnow-E, Eureka-Flite+, SBO LeCoPro, SBO Climaqs, POM; Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, Dynamical systems, control and optimization, 2007-2011); EU: ERNSI; FP7-HD-MPC (INFSO-ICT-223854), COST intelliCIS, EMBOCOM; Contract Research: AMINAL; Other: Helmholtz: viCERP, ACCM, Bauknecht, Hoerbiger. Ivan Goethals is a senior actuary at ING Life Belgium. Johan Suykens is a professor and Bart De Moor is a full professor at the Katholieke Universiteit Leuven, Belgium.

References

1. Bai, E.W.: An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems. Automatica 4(3), 333–338 (1998)
2. Bai, E.W.: A blind approach to Hammerstein model identification. IEEE Transactions on Signal Processing 50(7), 1610–1619 (2002)
3. Bako, L., Mercère, G., Lecoeuche, S., Lovera, M.: Recursive subspace identification of Hammerstein models based on least squares support vector machines. IET Control Theory & Applications 3, 1209–1216 (2009)
4. Chang, F.H.I., Luus, R.: A noniterative method for identification using the Hammerstein model. IEEE Transactions on Automatic Control 16, 464–468 (1971)
5. Crama, P.: Identification of block-oriented nonlinear models. PhD thesis, Vrije Universiteit Brussel, Dept. ELEC (2004)
6. Crama, P., Schoukens, J.: Hammerstein-Wiener system estimator initialization. In: Proc. of the International Conference on Noise and Vibration Engineering (ISMA 2002), Leuven, pp. 1169–1176 (2002)
7. Crama, P., Schoukens, J.: Initial estimates of Wiener and Hammerstein systems using multisine excitation. IEEE Transactions on Measurement and Instrumentation 50(6), 1791–1795 (2001)

8. De Brabanter, K., Dreesen, P., Karsmakers, P., Pelckmans, K., De Brabanter, J., Suykens, J.A.K., De Moor, B.: Fixed-Size LS-SVM Applied to the Wiener–Hammerstein Benchmark. In: Proceedings of the 15th IFAC Symposium on System Identification (SYSID 2009), Saint-Malo, France, pp. 826–831 (2009)
9. Dempsey, E.J., Westwick, D.T.: Identification of Hammerstein models with cubic spline nonlinearities. IEEE Transactions on Biomedical Engineering 51, 237–245 (2004)
10. Falck, T., Pelckmans, K., Suykens, J.A.K., De Moor, B.: Identification of Wiener–Hammerstein Systems using LS-SVMs. In: Proceedings of the 15th IFAC Symposium on System Identification (SYSID 2009), Saint-Malo, France, pp. 820–825 (2009)
11. Goethals, I., Hoegaerts, L., Suykens, J.A.K., Verdult, V., De Moor, B.: Hammerstein-Wiener subspace identification using kernel Canonical Correlation Analysis. Technical Report 05-30, ESAT-SISTA, K.U.Leuven, Leuven Belgium (2005), `http://ftp.esat.kuleuven.ac.be/pub/SISTA/goethals/goethals_hammer_wiener.ps`
12. Goethals, I., Mevel, L., Benveniste, A., De Moor, B.: Recursive output-only subspace identification for in-flight flutter monitoring. In: Proceedings of the 22nd International Modal Analysis Conference (IMAC-XXII), Dearborn, Michigan (2004)
13. Goethals, I., Pelckmans, K., Hoegaerts, L., Suykens, J.A.K., De Moor, B.: Subspace intersection identification of Hammerstein-Wiener systems. In: Proceedings of the 44th IEEE conference on Decision and Control, and the European Control Conference (CDC-ECC 2005), Seville, Spain, pp. 7108–7113 (2005)
14. Goethals, I., Pelckmans, K., Suykens, J.A.K., De Moor, B.: Identification of MIMO Hammerstein models using least squares support vector machines. Automatica 41(7), 1263–1272 (2005)
15. Goethals, I., Pelckmans, K., Suykens, J.A.K., De Moor, B.: Subspace identification of Hammerstein systems using least squares support vector machines. IEEE Transactions on Automatic Control, Special Issue on System Identification 50(10), 1509–1519 (2005)
16. Golub, G.H., Van Loan, C.F.: Matrix Computations. John Hopkins University Press, Baltimore (1989)
17. Greblicki, W., Pawlak, M.: Identification of discrete Hammerstein systems using kernel regression estimates. IEEE Transactions on Automatic Control 31, 74–77 (1986)
18. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, Heidelberg (2001)
19. Janczak, A.: Neural network approach for identification of Hammerstein systems. International Journal of Control 76(17), 1749–1766 (2003)
20. Kulcsar, B., Van Wingerden, J.W., Dong, J., Verhaegen, M.: Closed-loop Subspace Predictive Control for Hammerstein systems. In: Proceedings of the 48th IEEE Conference on Decision and Control held jointly with the 28th Chinese Control Conference (CDC 2009/CCC 2009), Shanghai, China, pp. 2604–2609 (2009)
21. Lovera, M., Gustafsson, T., Verhaegen, M.: Recursive subspace identification of linear and non-linear wiener state space models. Automatica 36, 1639–1650 (1998)
22. McKelvey, T., Hanner, C.: On identification of Hammerstein systems using excitation with a finite number of levels. In: Proceedings of the 13th International Symposium on System Identification (SYSID 2003), pp. 57–60 (2003)
23. Mercère, G., Lecoeuche, S., Lovera, M.: Recursive subspace identification based on instrumental variable unconstrained quadratic optimization. Adaptive Control and Signal Processing, Special issue on Subspace-based identification in adaptive control and signal processing 18, 771–797 (2004)

24. Narendra, K.S., Gallman, P.G.: An iterative method for the identification of nonlinear systems using the Hammerstein model. IEEE Transactions on Automatic Control 11, 546–550 (1966)
25. Pawlak, M.: On the series expansion approach to the identification of Hammerstein systems. IEEE Transactions on Automatic Control 36, 736–767 (1991)
26. Pelckmans, K., Goethals, I., De Brabanter, J., Suykens, J.A.K., De Moor, B.: Componentwise least squares support vector machines. In: Wang, L. (ed.) Support Vector Machines: Theory and Applications, pp. 77–98. Springer, Heidelberg (2005)
27. Schölkopf, B., Smola, A.: Learning with Kernels. MIT Press, Cambridge (2002)
28. Sjöberg, J., Zhang, Q., Ljung, L., Benveniste, A., Delyon, B., Glorennec, P., Hjalmarsson, H., Juditsky, A.: Nonlinear black-box modeling in system identification: a unified overview. Automatica 31(12), 1691–1724 (1995)
29. Stoica, P., Söderström, T.: Instrumental-Variable Methods for Identification of Hammerstein Systems. International Journal of Control 35(3), 459–476 (1982)
30. Suykens, J.A.K., Van Gestel, T., De Brabanter, J., De Moor, B., Vandewalle, J.: Least Squares Support Vector Machines. World Scientific, Singapore (2002)
31. Van Overschee, P., De Moor, B.: Subspace Identification for Linear Systems: Theory, Implementation, Applications. Kluwer Academic Publishers, Dordrecht (1996)
32. Van Pelt, T.H., Bernstein, D.S.: Nonlinear system identification using Hammerstein and nonlinear feedback models with piecewise linear static maps - part I: Theory. In: Proceedings of the American Control Conference (ACC 2000), pp. 225–229 (2000)
33. Vapnik, V.N.: Statistical Learning Theory. Wiley & Sons, Chichester (1998)
34. Verdult, V., Suykens, J.A.K., Boets, J., Goethals, I., De Moor, B.: Least squares support vector machines for kernel CCA in nonlinear state-space identification. In: Proceedings of the 16th international symposium on Mathematical Theory of Networks and Systems (MTNS 2004), Leuven, Belgium (2004)
35. Verhaegen, M., Westwick, D.: Identifying MIMO Hammerstein systems in the context of subspace model identification methods. International Journal of Control 63, 331–349 (1996)
36. Wahba, G.: Spline models for Observational data. SIAM, Philadelphia (1990)
37. Westwick, D., Verhaegen, M.: Identifying MIMO Wiener systems using subspace model identification methods. Signal Processing 52(2), 235–258 (1996)
38. Wills, A.G., Ninness, B.: Estimation of Generalised Hammerstein–Wiener Systems. In: Proceedings of the 15th IFAC Symposium on System Identification (SYSID 2009), Saint-Malo, France, pp. 1104–1109 (2009)