Joint 48th IEEE Conference on Decision and Control and
28th Chinese Control Conference
Shanghai, P.R. China, December 16-18, 2009

Closed-loop subspace identification of Hammerstein-Wiener models

Jan-Willem van Wingerden and Michel Verhaegen

Abstract— **In this paper we present a novel algorithm to identify MIMO Hammerstein-Wiener systems under open and closed-loop conditions. We reformulate a linear regression problem, commonly used as the first step in closed loop subspace identification, as an intersection problem which can be solved by using canonical correlation analysis (CCA). This makes it possible to utilize ideas from machine learning to estimate the static nonlinearities of Hammerstein-Wiener systems, using kernel canonical correlation analysis (KCCA). In the second step the state sequence is estimated and consequently the dynamic part can be identified. The effectiveness of the approach is illustrated with a closed-loop simulation example.**

Index Terms— **Subspace identification, Hammerstein-Wiener systems, System identification**

I. Introduction

Hammerstein, Wiener, and Hammerstein-Wiener systems are a particular class of nonlinear systems, which are linear time-invariant (LTI) models with a static nonlinearity at the input, output, and output and input, respectively. Although, Hammerstein and Wiener system identification attracted considerable attention in the past few years (see [1], [2] and references therein), Hammerstein-Wiener system identification did not [2]. Still, the identification of Hammerstein-Wiener models is of interest since this model structure appears in a large number of applications [2].

To efficiently handle Multiple-Input Multiple-Output (MIMO) Hammerstein-Wiener systems and to exploit the numerical properties of subspace techniques, the focus of this paper is on subspace-based Hammerstein-Wiener system identification. In the area of Hammerstein-Wiener model identification only one subspace method exist [3], [4]. In this work they formulate the subspace identification problem as an intersection problem of the past and the future using Kernel Canonical Correlation Analysis (KCCA). From this intersection they compute the state sequence and in the second step they apply Least-Squares Support Vector Machines (LS-SVM) [5] to obtain the nonlinearities and the system matrices. The proposed method only applies for data generated in open loop, while from a practical point of view it is necessary to look at closed-loop system identification. Since, system identification is often used to identify a part of a complex model, either because the complete model is too complex or parts of the model are well-understood and it is not necessary to identify these. In Fig. 1 a sketch of a submodel is given, where the user is interested in the dynamics between u and y. In such cases, typically the output of the unknown (sub)model is correlated with the input of the other (sub)model(s). For this situation it is well known that, for the LTI case, the projector type of subspace algorithms (*e.g.* MOESP [6] and N4SID [7]) give biased estimates if the identification data is generated under closed-loop conditions. The recently developed predictor-based subspace identification methods (*e.g.* PBSID [8] and SSARX [9]) do not suffer from this drawback.

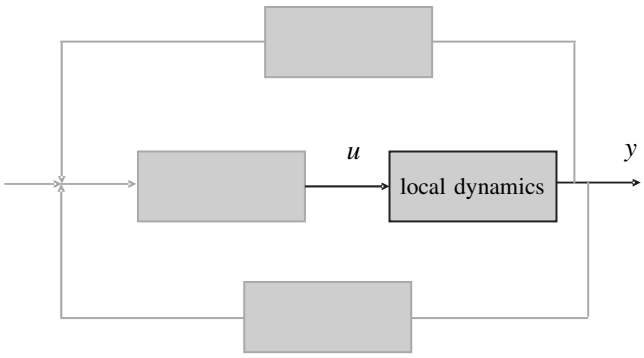

Fig. 1. Schematic representation of a coupled dynamic block.

In this paper we develop a novel subspace identification algorithm for Hammerstein-Wiener systems for data generated in open and closed loop that first estimates the nonlinearities and then the system matrices of the LTI part.

The outline of this paper is as follows; we start in Section II with the problem formulation and assumptions. In Section III we reformulate the $\text{PBSID}_{\text{opt}}$ identification scheme in a CCA setting. In Section IV we replace the CCA problem by a KCCA problem which enables us to identify Hammerstein-Wiener models. In Section V a simulation example is presented. We end this paper with our conclusions.

II. Problem formulation

In this section we present the problem formulation and the assumptions we make.

A. Problem formulation

For the derivation of the algorithm we consider the following Hammerstein-Wiener system[1]:

$$x_{k+1} = Ax_k + Bf(u_k) + Ke_k, \quad (1)$$
$$g^{-1}(y_k) = Cx_k + Df(u_k) + e_k, \quad (2)$$

The authors are with the Delft Center for Systems and Control (DCSC), Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands J.W.vanWingerden@tudelft.nl m.verhaegen@moesp.org

[1]Note that we use g^{-1} in the notation to emphasis the relation between the common formulation of Hammerstein-Wiener model structures. Further, for the identification algorithm we do not require the assumption that g^{-1} is invertible.

where $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^r$, $y_k \in \mathbb{R}^\ell$, are the state, input and output vectors. $e_k \in \mathbb{R}^\ell$ denotes the zero-mean white innovation process. The matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times r}$, $C \in \mathbb{R}^{\ell\times n}$, $D \in \mathbb{R}^{\ell\times r}$, $K \in \mathbb{R}^{n\times \ell}$ are the local system, input, output, direct feed through, and the noise matrices and finally $f(.) : \mathbb{R}^r \to \mathbb{R}^r$ and $g(.)^{-1} : \mathbb{R}^\ell \to \mathbb{R}^\ell$ are static smooth nonlinear functions with as additional constraint that $f(0) = 0$ and $g(0) = 0$. We can rewrite (1)-(2) in the predictor form as:

$$x_{k+1} = \tilde{A}x_k + \tilde{B}f(u_k) + Kg^{-1}(y_k), \quad (3)$$
$$g^{-1}(y_k) = Cx_k + Df(u_k) + e_k, \quad (4)$$

with $\tilde{A} = A - KC$, and $\tilde{B} = B - KD$. It is well-known that an invertible linear transformation of the state does not change the input-output behavior of the dynamic part of a Hammerstein-Wiener system. A similar linear transformation appears between the nonlinearities and the dynamic part. So, the static nonlinearities can be estimated up to a square invertible transformation. These transformations are given by: $T^{-1}AT$, $T^{-1}BT_u$, $T^{-1}KT_y$, $T_y^{-1}CT$, $T_y^{-1}DT_u$, $T_y^{-1}g(.)^{-1}$, and $T_u^{-1}f(.)$ with $T \in \mathbb{R}^{n\times n}, T_y \in \mathbb{R}^{\ell\times\ell}, T_u \in \mathbb{R}^{r\times r}$.

The identification problem can now be formulated as: given the input sequence u_k and the output sequence y_k over a time $k = \{0,\ldots,N-1\}$; find all, if they exist, the system matrices A, B, D, C, K and static nonlinearities, $f(.)$ and $g(.)$, up to the mentioned similarity transformations.

B. Assumptions and notation

We define a past window denoted by p. This window is used to define the following stacked vector:

$$f(\overline{U}_P) = \begin{bmatrix} f(U_0) \\ f(U_1) \\ \vdots \\ f(U_{p-1}) \end{bmatrix}, \qquad g^{-1}(\overline{Y}_P) = \begin{bmatrix} g^{-1}(Y_0) \\ g^{-1}(Y_1) \\ \vdots \\ g^{-1}(Y_{p-1}) \end{bmatrix},$$

and

$$f(\overline{U}_F) = f(U_p), \qquad g^{-1}(\overline{Y}_F) = g^{-1}(Y_p),$$

with:

$$f(U_i) = \begin{bmatrix} f(u_i), & f(u_{i+1}), & \cdots, & f(u_{i+N-p}) \end{bmatrix},$$
$$g^{-1}(Y_i) = \begin{bmatrix} g^{-1}(y_i), & g^{-1}(y_{i+1}), & \cdots, & g^{-1}(y_{i+N-p}) \end{bmatrix},$$
$$X_i = \begin{bmatrix} x_i, & x_{i+1}, & \cdots, & x_{i+N-p} \end{bmatrix}.$$

Considering the LTI part of the model structure, we assume that the state sequence, X_p, has full row rank and the extended observability matrix given by:

$$\Gamma = \begin{bmatrix} C^T, & (C\tilde{A}^{(1)})^T, & \ldots, & (C\left(\tilde{A}^{(1)}\right)^{f-1})^T \end{bmatrix}^T,$$

has full column rank and where f is referred to as the future window and $f \leq p$. We also define the extended controllability matrix:

$$\mathscr{K} = \begin{bmatrix} \mathscr{K}_u, & \mathscr{K}_y \end{bmatrix},$$

with:

$$\mathscr{K}_u = \begin{bmatrix} \tilde{A}^{p-1}\tilde{B}, & \tilde{A}^{p-2}\tilde{B}, & \cdots, & \tilde{B} \end{bmatrix},$$
$$\mathscr{K}_y = \begin{bmatrix} \tilde{A}^{p-1}K, & \tilde{A}^{p-2}K, & \cdots, & K \end{bmatrix}.$$

In the next section we revise the idea of predictor based subspace identification, in particular we will look at the PBSID$_{\text{opt}}$ algorithm [8]. Although, we will formulate the first step of the algorithm as an intersection problem which can be solved using Canonical Correlation Analysis (CCA). This step is required to estimate the static nonlinearities in the Hammerstein-Wiener setting.

III. Predictor-Based Subspace IDentification

It is well known that the projector type of subspace algorithms (*e.g.* MOESP [6] and N4SID [7]) give biased estimates if the identification data set is generated under closed-loop conditions. The main reason for the bias is the constraint that for the projector type of algorithms the noise and the input should be uncorrelated. This assumption is clearly violated if there is a feedback loop present (as clearly explained in [10]). Predictor-based subspace identification (*e.g.* PBSID [8] and SSARX [9]) methods do not suffer from this drawback. In this section we introduce the PBSID$_{\text{opt}}$ algorithm and for now we assume that the nonlinearity is known. In the next section we reformulate the problem and we explicitly estimate the nonlinearities.

A. Predictors

The first objective of predictor based subspace algorithms is to reconstruct the state sequence up to a similarity transformation. The state X_p is given by:

$$X_p = \tilde{A}^p X_0 + \begin{bmatrix} \mathscr{K}_u, & \mathscr{K}_y \end{bmatrix} \begin{bmatrix} f(\overline{U}_P) \\ g^{-1}(\overline{Y}_P) \end{bmatrix}.$$

The key approximation in this algorithm is that we assume that $\tilde{A}^j \approx 0$ for all $j \geq p$. It can be shown that if the system in (3)-(4) is uniformly exponentially stable, the approximation error can be made arbitrarily small by making p sufficiently large [8]. With this assumption the state X_p is approximately given by:[2] [3]

$$X_p \approx \begin{bmatrix} \mathscr{K}_u, & \mathscr{K}_y \end{bmatrix} \begin{bmatrix} f(\overline{U}_P) \\ g^{-1}(\overline{Y}_P) \end{bmatrix}.$$

In a number of other LTI subspace methods it is well known to make this step (*e.g.* N4SID, SSARX, PBSID$_{\text{opt}}$, CVA). The input-output behavior is now approximately given by:

$$g^{-1}(Y_F) \approx C \begin{bmatrix} \mathscr{K}_u, & \mathscr{K}_y \end{bmatrix} \begin{bmatrix} f(\overline{U}_P) \\ g^{-1}(\overline{Y}_P) \end{bmatrix} + Df(U_F) + e_k.$$

This equation can be looked at as a regression problem and when the nonlinearities are known, an estimate of the matrices $C\mathscr{K}$, and D can be obtained by solving a least squares problem. However, the nonlinearities are not known and that is why we introduce in the next section a kernel-CCA (KCCA) problem to estimate the nonlinearities. In this section we continue by assuming that we already have

[2]Remark: If the matrix $(A - KC)$ is nilpotent the assumption can be removed. Since, in that case it holds that there exists a p such that $(A - KC)^p = 0$.

[3]Remark: In the case that there is no noise present the observer gain will be a deadbeat observer and under the conditions stated the following will hold: $(A - KC)^p = 0$ for $p \geq n$.

the nonlinearities and we show how to obtain the system matrices. But first we show how to obtain a normal CCA problem if we know the nonlinearities.

B. Intersection problem

We can formulate the following intersection problem, better known as a CCA problem:[4]

$$\min_{V,W} ||V\Phi - W\Psi||_F^2, \tag{5}$$

$$\begin{aligned} \text{Subject to} \quad V\Phi\Phi^{\mathrm{T}}V^{\mathrm{T}} &= I_\ell, \\ W\Psi\Psi^T W^T &= I_\ell, \end{aligned}$$

with:

$$\Phi = \begin{bmatrix} -g^{-1}(\overline{Y}_P) \\ g^{-1}(Y_F) \end{bmatrix}, \qquad \Psi = \begin{bmatrix} f(\overline{U}_P) \\ f(U_F) \end{bmatrix}.$$

The solution is the estimate of the following matrices:

$$V = T_y^{-1}\left[C\mathscr{K}_y, \quad I\right], \qquad W = T_y^{-1}\left[C\mathscr{K}_u, \quad D\right].$$

Observe that we retrieve some unknown similarity transformation which is defined in the problem formulation. In the case that we know the nonlinearities we can get rid of this nonlinearity by putting the last $\ell \times \ell$ rows of V equal to the identity. So, we end up with estimates of $C\mathscr{K}_u$, $C\mathscr{K}_y$, and D.

The problem formulated in (5) can be interpreted as attempting to determine the intersection between the row spaces of two known matrices, or as a generalization of angles between subspaces. The constraints in (5) ensure orthogonality of the projections. Different methods for solution exist on which we will not elaborate further [11], [12]. It is however important to note that when the dimension of the target space (the number of columns in (5)) is relatively small compared to the dimension of the signal space (the number of rows of Φ and Ψ), CCA methods may obtain solution directions which are determined mainly by noise, which in our case is present because we have some measurement noise. This is also known as overfitting. A method to decrease the variance of the solutions, at the cost of introducing a bias, is to employ regularization as presented in [13]. It can be shown that a solution of (5) can be found by solving the regularized generalized eigenvalue problem:

$$\begin{bmatrix} 0 & \Phi\Psi^T \\ \Psi\Phi^T & 0 \end{bmatrix} \begin{bmatrix} V^T \\ W^T \end{bmatrix} = \lambda \begin{bmatrix} \Phi\Phi^T + \nu_1 I & 0 \\ 0 & \Psi\Psi^T + \nu_2 I \end{bmatrix} \begin{bmatrix} V^T \\ W^T \end{bmatrix},$$

where λ is referred to as the canonical correlation coefficient and will be between 0 and 1, indicating the closeness of the subspaces. The generalized eigenvectors corresponding to the ℓ highest canonical coefficients are chosen as solutions. The regularization parameters ν_1, ν_2 should be chosen appropriately as will be discussed in Section V.

[4]Remark: In the noise and approximation free case the following equality should hold:

$$\operatorname{rank}\left(\begin{bmatrix} \Phi \\ \Psi \end{bmatrix}\right) + \ell = \operatorname{rank}(\Phi) + \operatorname{rank}(\Psi).$$

In the case that there is noise or an approximation the rank conditions are hard to check but an approximation can be made by using an SVD.

C. Extended observability times controllability matrix

The product $\mathscr{K}\left[f(U_P)^T \quad g^{-1}(Y_P)^T\right]^T$ that represents by definition the state sequence, X_p, can not be estimated directly. In the predictor-based identification algorithms, $C\mathscr{K}$ is used to construct the extended observability matrix times the extended controllability matrix. This matrix is given by:

$$\begin{aligned} \Gamma\mathscr{K} &= \begin{bmatrix} C\mathscr{K}_u & C\mathscr{K}_y \\ C\tilde{A}\mathscr{K}_u & C\tilde{A}\mathscr{K}_y \\ \vdots & \vdots \\ C\tilde{A}^{f-1}\mathscr{K}_u & C\tilde{A}^{f-1}\mathscr{K}_y \end{bmatrix}, \\ &= \begin{bmatrix} \Gamma\mathscr{K}_u & \Gamma\mathscr{K}_y \end{bmatrix}. \end{aligned}$$

If we look into more detail at for instance $\Gamma\mathscr{K}_u$ term we have for the PBSID$_{\text{opt}}$ algorithm (we assume $f = p$):

$$\Gamma\mathscr{K}_u \approx \begin{bmatrix} C\tilde{A}^{p-1}\tilde{B} & C\tilde{A}^{p-2}\tilde{B} & \cdots & C\tilde{B} \\ 0 & C\tilde{A}^{p-1}\tilde{B} & \cdots & C\tilde{A}\tilde{B} \\ \vdots & & \ddots & \\ 0 & & & C\tilde{A}^{p-1}\tilde{B} \end{bmatrix}.$$

The zeros appear in this equation based on the approximation that $\tilde{A}^j \approx 0$ for all $j \geq p$. Although, in [8] they showed that these zeros appear in the solution based on the result of an optimization problem. Observe that from the solution of the CCA problem formulated in (5) we can construct this matrix. A similar thing can be done for the matrix $\Gamma\mathscr{K}_y$.

From the constructed matrix $\Gamma\mathscr{K}$ we can compute:

$$\Gamma\mathscr{K}\begin{bmatrix} f(\overline{U}_P) \\ g^{-1}(\overline{Y}_P) \end{bmatrix},$$

which equals by definition the extended observability times the state sequence, ΓX_p. By computing a Singular Value Decomposition (SVD) of this estimate we can estimate the state sequence and the order of the system. We will use the following SVD:

$$\widehat{\Gamma\mathscr{K}}\begin{bmatrix} f(\overline{U}_P) \\ g^{-1}(\overline{Y}_P) \end{bmatrix} = \begin{bmatrix} \mathscr{U} & \mathscr{U}_\perp \end{bmatrix} \begin{bmatrix} \Sigma_n & 0 \\ 0 & \Sigma \end{bmatrix} \begin{bmatrix} \mathscr{V} \\ \mathscr{V}_\perp \end{bmatrix}, \tag{6}$$

where Σ_n is the diagonal matrix containing the n largest singular values and $\mathscr{V}$ is the corresponding row space. Note that we can find the largest singular values by detecting a gap between the singular values [14]. The state is now estimated by:

$$\widehat{X}_p = \Sigma_n \mathscr{V}. \tag{7}$$

It is well known that when the state, input, output, and nonlinearities are known the system matrices can be estimated [14]. The approach presented so far assumed that we know the nonlinearities. In the next section we present how we can solve the problem without knowing the nonlinear functions beforehand.

IV. Hammerstein-Wiener Predictor-Based Subspace IDentification

In the previous section we showed that the first step in the PBSID$_{\text{opt}}$ algorithm can be replaced by an intersection problem. The remainder of that particular section was used

to explain the other steps of the algorithm. In this section we will show that when we exploit tools derived in machine learning we can estimate the nonlinearities. Instead of formulating the intersection problem as a CCA problem we now formulate a kernel CCA problem. From which we directly estimate the nonlinearities and the solution can be used to construct the extended controllability times the observability matrix. From that point on similar steps can be taken as in the previous section.

A. Kernel Canonical Correlation Analysis (KCCA)

With a number of substitutions the problem in (5) can be rewritten as:

$$\min_{\alpha,\beta} ||\alpha\tilde{\Phi}^T\Phi - \beta\tilde{\Psi}^T\Psi||_F^2, \tag{8}$$

Subject to:

$$\alpha\tilde{\Phi}^T\Phi\Phi^T\tilde{\Phi}\alpha^T = I_\ell, \qquad \beta\tilde{\Psi}^T\Psi\Psi^T\tilde{\Psi}\beta^T = I_\ell,$$

where $\tilde{\Psi}$ and $\tilde{\Phi}$ can be interpreted as basis data which we will define later and $V = \alpha\tilde{\Psi}^T$ and $W = \beta\tilde{\Phi}^T$.

The problem formulated in (8) is still nonlinear. We will now approximate the matrices $\tilde{\Phi}^T\Phi$ and $\tilde{\Psi}^T\Psi$ by centered component-wise kernels. These kernels are defined as follows[5]:

Definition 1 (Centered component wise kernels): Given two data matrices:

$$E = \begin{bmatrix} e_1 \\ \vdots \\ e_\nu \end{bmatrix} \in \mathbb{R}^{\nu\times N_\nu}, \quad \text{and} \qquad \mathrm{F} = \begin{bmatrix} f_1 \\ \vdots \\ f_\tau \end{bmatrix} \in \mathbb{R}^{\tau\times \mathrm{N}_\tau}.$$

We define the follow component wise kernel:

$$K(E,F) = \overline{K}(e_1,f_1) + ... + \overline{K}(e_\nu,f_\tau),$$

with $\overline{K}(a,b) \in \mathbb{R}^{N_\nu\times N_\tau}$ defined as a centered kernel with $a \in \mathbb{R}^{N_\nu}$, $b \in \mathbb{R}^{N_\tau}$ and defined as:

$$\overline{K}(a,b) = \begin{bmatrix} (q_{a_1,b_1} - q_{a_1,0}), & \cdots, & (q_{a_1,b_{N_\tau}} - q_{a_1,0}) \\ (q_{a_2,b_1} - q_{a_2,0}), & \cdots, & (q_{a_2,b_{N_\tau}} - q_{a_2,0}) \\ \vdots & \ddots & \vdots \\ (q_{a_{N_\nu},b_1} - q_{a_{N_\nu},0}), & \cdots, & (q_{a_{N_\nu},b_{N_\tau}} - q_{a_{N_\nu},0}) \end{bmatrix}$$

where $q_{i,j}$ is a kernel function with the arguments i, and j. For instance we have radial kernels, $q_{i,j} = \exp -(i-j)^2/\sigma$ (σ is a kernel constant), and linear kernels, $q_{i,j} = i \times j$. The matrix $\overline{K}$ has the property that $\overline{K}(a,0) = 0$ and this is referred to as the centering property.

With the definition of the kernels we can redefine (8) in a KCCA problem. However, we still have the degree of freedom to pick the 'basis' data. There is a large amount of literature written on how to choose these 'basis' functions (see [5] and references therein) but we continue by taking the whole data set as basis data:

$$\min_{\alpha,\beta} \left|\left| \alpha K\left(\begin{bmatrix}\overline{Y}_P \\ Y_F\end{bmatrix}, \begin{bmatrix}\overline{Y}_P \\ Y_F\end{bmatrix}\right) - \beta K\left(\begin{bmatrix}\overline{U}_P \\ U_F\end{bmatrix}, \begin{bmatrix}\overline{U}_P \\ U_F\end{bmatrix}\right) \right|\right|_F^2, \tag{9}$$

Subject to:

$$\alpha K\left(\begin{bmatrix}\overline{Y}_P \\ Y_F\end{bmatrix}, \begin{bmatrix}\overline{Y}_P \\ Y_F\end{bmatrix}\right) K\left(\begin{bmatrix}\overline{Y}_P \\ Y_F\end{bmatrix}, \begin{bmatrix}\overline{Y}_P \\ Y_F\end{bmatrix}\right)^T \alpha^T = I_\ell,$$
$$\beta K\left(\begin{bmatrix}\overline{U}_P \\ U_F\end{bmatrix}, \begin{bmatrix}\overline{U}_P \\ U_F\end{bmatrix}\right) K\left(\begin{bmatrix}\overline{U}_P \\ U_F\end{bmatrix}, \begin{bmatrix}\overline{U}_P \\ U_F\end{bmatrix}\right)^T \beta^T = I_\ell.$$

This formula represents the kernel counterpart of the CCA problem from the previous section. Observe that regularization is a must to solve this problem (see discussion in the previous section).

The key observation that we need for the remainder of the paper is that the Markov parameters times the static nonlinearities can be expressed in the solution of (9); $\hat{\alpha}$ and $\hat{\beta}$. Since by definition we have:

$$\alpha K\left(\begin{bmatrix}\overline{Y}_P \\ Y_F\end{bmatrix}, .\right) = Vg^{-1}(.) = T_y^{-1}\left[C\mathscr{K}_y, \quad I\right] g^{-1}(.),$$
$$\beta K\left(\begin{bmatrix}\overline{U}_P \\ U_F\end{bmatrix}, .\right) = Wf(.) = T_y^{-1}\left[C\mathscr{K}_u, \quad D\right] f(.),$$

and due to the component-wise formulation of the kernels the following equalities also hold:

$$\alpha K(Y_i,.) = T_y^{-1}C\tilde{A}^{p-i-1}Kg^{-1}(.),$$
$$\beta K(U_i,.) = T_y^{-1}C\tilde{A}^{p-i-1}\tilde{B}f(.),$$

for all $i \in \{0,\cdots,p-1\}$ and:

$$\alpha K(Y_F,.) = T_y^{-1}g^{-1}(.), \qquad \beta K(U_F,.) = T_y^{-1}Df(.).$$

B. Estimation of the Wiener and Hammerstein nonlinearities

The Wiener nonlinearity is directly a product of the kernel CCA problem that we formulated. We have:

$$\widehat{T_y^{-1}g(.)^{-1}} = \hat{\alpha}K(Y_F,.), \tag{10}$$

which follows directly by the definition of the component wise centered kernels as stated at the end of the previous subsection. To obtain an estimate of the Hammerstein nonlinearity we have to solve an SVD. We have the following equality:

$$\begin{bmatrix} T_y^{-1}CB \\ T_y^{-1}C\tilde{A}B \\ \vdots \\ T_y^{-1}C\tilde{A}^{p-1}B \end{bmatrix} f(U_F) = \begin{bmatrix} \beta K(U_0,U_F) \\ \beta K(U_1,U_F) \\ \vdots \\ \beta K(U_{p-1},U_F) \end{bmatrix}.$$

The right-hand side of the equation can be estimated and since it is a low rank matrix we can obtain the Hammerstein nonlinearity by performing an SVD:

$$\begin{bmatrix} \hat{\beta}K(U_0,U_F) \\ \hat{\beta}K(U_1,U_F) \\ \vdots \\ \hat{\beta}K(U_{p-1},U_F) \end{bmatrix} = \begin{bmatrix} \mathscr{U}_f & \mathscr{U}_{f,\perp} \end{bmatrix} \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathscr{V}_f \\ \mathscr{V}_{f,\perp} \end{bmatrix}.$$

[5]These kernels are only suited for Hammerstein-Wiener models without cross-coupling between the different inputs or outputs. The cross-coupling can be included by changing the definition of the used kernels. For sake of simplicity this is not included in this paper.

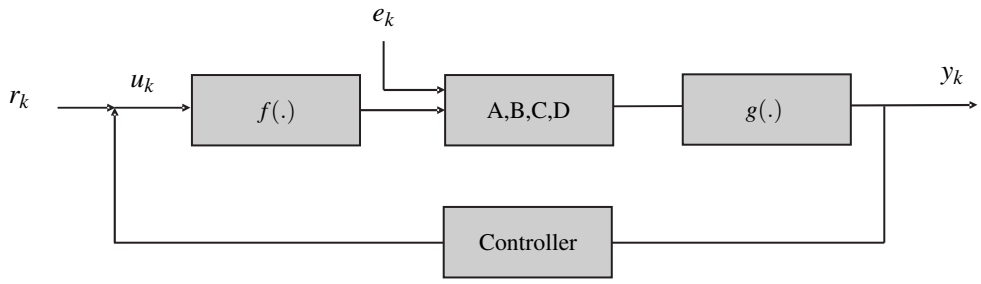

Fig. 2. Schematic representation of the simulation scheme

$$A=\begin{bmatrix} 0.67 & 0.67 & 0 & 0 \\ -0.67 & 0.67 & 0 & 0 \\ 0 & 0 & -0.67 & -0.67 \\ 0 & 0 & 0.67 & -0.67 \end{bmatrix}, \quad B=\begin{bmatrix} 0.6598 & -0.5256 \\ 1.9698 & 0.4845 \\ 4.3171 & -0.4879 \\ -2.6436 & -0.3416 \end{bmatrix}, \quad C=\begin{bmatrix} 0.3749 & 0.0751 & -0.5225 & 0.5830 \\ -0.8977 & 0.7543 & 0.1159 & 0.0982 \end{bmatrix},$$

$$K=\begin{bmatrix} -0.6968 & -0.1474 \\ 0.1722 & 0.5646 \\ 0.6484 & -0.4660 \\ -0.9400 & 0.1032 \end{bmatrix}, \quad F_b=\begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}, \quad f(u_k)=\begin{bmatrix} \sin\left(u_k^{(1)}\right) u_k^{(1)} \\ \sin\left(u_k^{(2)}\right) \end{bmatrix}, \quad g(y_k)=\begin{bmatrix} \text{sign}(y_k^{(1)})\sqrt{y_k^{(1)}} \\ \text{sign}(y_k^{(2)})\sqrt{y_k^{(2)}} \end{bmatrix},$$

TABLE I

SIMULATION PARAMETERS

Using the column space of this matrix we now have an estimate of the Hammerstein nonlinearity and this one is given by:

$$\widehat{T_u^{-1} f(.)} = \mathscr{U}_f^{\dagger} \begin{bmatrix} \hat{\beta} K(U_0, U_F) \\ \hat{\beta} K(U_1, U_F) \\ \vdots \\ \hat{\beta} K\left(U_{f-1}, U_F\right) \end{bmatrix},$$

where $\dagger$ represents the pseudo inverse and T_u^{-1} the unknown transformation.

C. Extended observability times controllability matrix

With the known nonlinearities we can also go back to the previous section and solve the linear problem. However, it turns out that from the KCCA problem we directly can construct an estimate of the controllability time observability matrix. Again, we exploit the component wise structure of the kernels. The observability times the state is now given by:

$$\widehat{\Gamma X_p} = \begin{bmatrix} \sum_{i=0}^{p-1} \left(\hat{\alpha} K(U_i, U_i) + \hat{\beta} K(Y_i, Y_i) \right) \\ \sum_{i=1}^{p-1} \left(\hat{\alpha} K(U_i, U_{i-1}) + \hat{\beta} K(Y_i, Y_{i-1}) \right) \\ \vdots \\ \sum_{i=f-1}^{p-1} \left(\hat{\alpha} K(U_i, U_{i-f-1}) + \hat{\beta} K(Y_i, Y_{i-f-1}) \right) \end{bmatrix}.$$

With this matrix known we can compute the SVD in (6) and estimate the state (7). With an estimate of the state and the nonlinearities the system matrices can be estimated in a straightforward manner.

V. EXAMPLE

We have tested the proposed identification algorithm on a fourth-order MIMO closed-loop model with $r = 2$, and $\ell = 2$. The collected data u_k, y_k, are used for the identification algorithm. We will use the following model (as illustrated in Fig. 2):

$$\begin{aligned} x_{k+1} &= Ax_k + Bf(u)_k + Ke_k, \\ g^{-1}(y_k) &= Cx_k + e_k, \\ u_k &= F_b g^{-1}(y_k) + r_k, \end{aligned}$$

where F_b is the feedback gain and r_k the reference signal. The system matrices are given in Table IV-B. As reference signal we take a zero-mean Gaussian white noise signal with $cov(u_k) = I_r$ and for e_k we take a Gaussian white noise with the following covariance $cov(e_k) = 0.01 \cdot I_\ell$. For the identification experiment we used $N = 500$ and $p = f = 4$. The collected data (u_k and y_k) is used to identify a Hammerstein-Wiener model. The performance of the identified system is evaluated by looking at the value of the Variance-Accounted-For (VAF)[6] on a data set different from the one used for identification. To investigate the sensitivity of the identification algorithm with respect to noise, a Monte-Carlo simulation with 20 runs was carried out. For each of the 20 simulations a different realization of the input r_k is used.

The choice of the regularization parameters, ν_1 and ν_2, was made by plotting the validation VAF for 20 simulations against a logarithmic grid of different values of the parameter and choosing the value corresponding to a peak. The location of the peak is strongly related to the data set and consequently is different for every data set. With

[6]The VAF value is defined as:

$$\text{VAF}(\text{g}^{-1}(\text{y}_\text{k}), \widehat{\text{g}^{-1}(\text{y}_\text{k})}) = \max\left\{ 1 - \frac{\text{var}(\text{g}^{-1}(\text{y}_\text{k}) - \widehat{\text{g}^{-1}(\text{y}_\text{k})})}{\text{var}(\text{g}^{-1}(\text{y}_\text{k}))}, 0 \right\} * 100\%,$$

where $\widehat{g^{-1}(y_k)}$ denotes the signal obtained by simulating the identified Hammerstein-Wiener model, $g^{-1}(y_k)$ is the signal of the true Hammerstein-Wiener model, and $var()$ denotes the variance of a quasi-stationary signal.

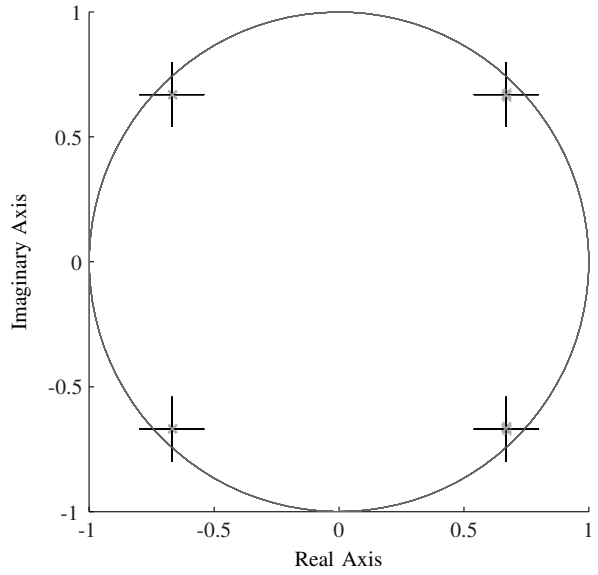

Fig. 3. True and estimated eigenvalues of A in one plot for 20 experiments.

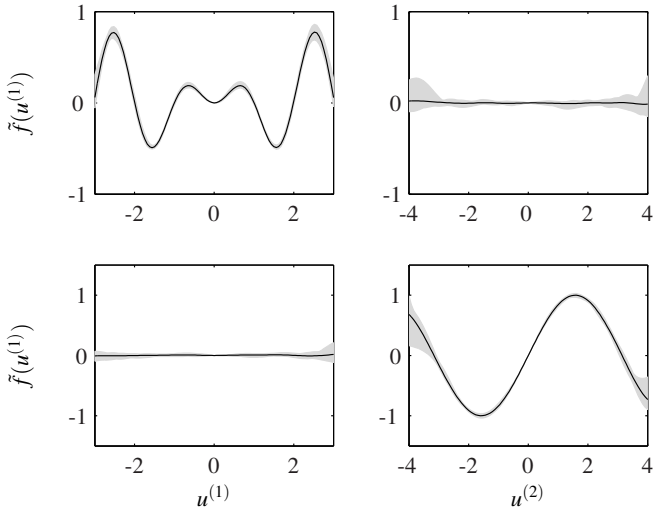

Fig. 4. Presentation of $\tilde{f}(u_k)$ based on 20 Monte-Carlo simulations. The solid line represents the estimate with the highest VAF and the gray area covers the other 19 estimates.

regularization we show in Fig. 3 the poles of A for the 20 Monte-Carlo simulations and we see that we have an almost unbiased estimate. In Fig. 4 and 5 we show the nonlinearities by plotting $f(u_k)$ and $g^{-1}(y_k)$. Although, the estimates of the nonlinearities are obtained up to an unknown transformation. We can estimate in this simulation example this transformation since we know the 'real' system. These estimates are denoted by $\tilde{f}(u_k)$ and $\tilde{g}^{-1}(y_k)$, respectively. We clearly see that we can estimate the nonlinearities quite accurately for different realizations of the noise.

VI. Conclusions

In this paper we presented a novel algorithm to identify MIMO Hammerstein-Wiener systems under open and closed-loop conditions. To do so, we formulated the linear time invariant optimized predictor based subspace identification

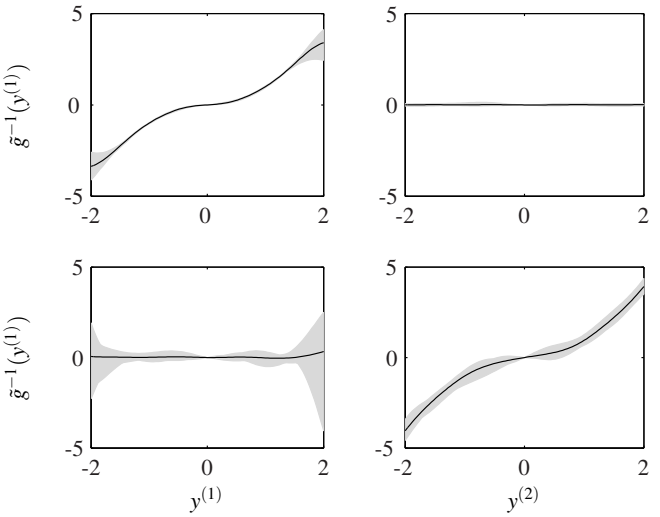

Fig. 5. Presentation of $\tilde{g}^{-1}(y_k)$ based on 20 Monte-Carlo simulations. The solid line represents the estimate with the highest VAF and the gray area covers the other 19 estimates.

algorithm as an intersection problem, better known as CCA. For Hammerstein-Wiener systems we utilized ideas from machine learning to estimate both the static nonlinearities and the Markov parameters. In the second step the state sequence is directly used to estimate the linear dynamics. The effectiveness of the approach was illustrated with a closed-loop simulation example.

References

[1] F. Ding and T. Chen. Identification of Hammerstein nonlinear ARMAX systems. *Automatica*, 41(9):1479–1489, 2005.

[2] E-W Bai. A blind approach to the Hammerstein-Wiener model identification. *Automatica*, 38(6):967–979, 2002.

[3] I. Goethals, K. Pelckmans, L. Hoegaerts, J. A. K. Suykens, and B. De Moor. Subspace intersection identification of Hammerstein-Wiener systems. In *Proceedings of the 44th Conference on Decision and Control (CDC) and the European Control Conference (ECC)*, Seville, Spain, 2005.

[4] I. Goethals. *Subspace identification for linear, Hammerstein and Hammerstein-Wiener systems*. PhD thesis, Katholieke Universiteit Leuven, 2005.

[5] J. A. K. Suykens, T. V. van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle. *Least squares support vector machines*. World Scientific, 2002.

[6] M. Verhaegen and P. Dewilde. Subspace model identification part 1: The output-error state-space model identification class of algorithms. *International Journal of Control*, 56(5):1187–1210, 1992.

[7] P. Van Overschee and B. De Moor. *Subspace Identification for linear systems*. Kluwer Academic Publishers, 1996.

[8] A. Chiuso. The role of vector auto regressive modeling in predictor based subspace identification. *Automatica*, 43(6):1034–1048, 2007.

[9] M. Jansson. A new subspace identification method for open and closed loop data. In *Proceedings of the 16th IFAC World Congress*, Prague, Czech Republic, 2005.

[10] L. Ljung and T. McKelvey. Subspace identification from closed loop data. *Signal Processing*, 52(2):209–215, 1996.

[11] G. A . F. Seber. *Multivariate Observations*. Wiley, 1984.

[12] J. Krzanowski, W. *Principles of Multivariate analysis*. Oxford University Press, 1988.

[13] T. De Bie and B. De Moor. On the regularization of canonical correlation analysis. In *Proceedings of the 4th International symposium on independent component analysis and blind signal separation (ICA2003)*, Nara, Japan, 2003.

[14] M. Verhaegen and V. Verdult. *Filtering and System Identification: A least squares approach*. Cambridge University Press, 2007.