Subspace-Based Identification
for Linear and Nonlinear Systems

Harish J. Palanthandalam-Madapusi, Seth Lacy, Jesse B. Hoagg and Dennis S. Bernstein

1. Introduction

Mathematical models describe the dynamic behavior of a system as a function of time, and arise in all scientific disciplines. These mathematical models are used for simulation, operator training, analysis, monitoring, fault detection, prediction, optimization, control system designs and quality control.

System identification is the process of constructing mathematical models of dynamical systems using measured data. The identified models can be used for output prediction, system analysis and diagnostics, system design, and control. Identification of linear models has been well studied and numerous frequency-domain and time-domain methods are available [21]. Among the available time-domain methods, subspace algorithms are an important class of algorithms for identifying linear state-space models.

Subspace methods are naturally applicable to multi-input multi-output (MIMO) systems and identify the system in the state-space form. These methods are computationally tractable and robust since they use QR factorization and singular value decompositions. Another advantage of subspace algorithms is that no a priori information about the system is needed, and thus they are widely applicable. Since the number of user choices is small, the software is user-friendly. All these factors make subspace identification a valuable tool.

Overviews of linear subspace identification are presented in detail in [5, 34, 45]. Detailed examination of subspace-based identification of bilinear and linear parameter-varying (LPV) systems is done in [36]. Subspace algorithms have also been applied to numerous industrial and real-life applications ranging from glass tube manufacturing processes to bridge dynamics. A few application-based papers are [9], Sections 6.3 and 6.4 of [34] and [31]. Section 6.4 of [34] references several other interesting applications. In this paper, we introduce the basic ideas behind subspace algorithms to the uninitiated reader and provide an overview of linear, time-varying, nonlinear and closed-loop methods.

The basic idea behind subspace is that the Kalman filter state sequence can directly be estimated from the input/output observations. Once the state estimates are available, the state space matrices are estimated using a linear least squares fit. This is contrary to traditional identification

methods, where the model is identified first and then the states are generated using the identified model.

This paper is organized in the following way. Section 2 deals with the basic subspace algorithm for time-invariant systems. A simplified proof of the fact that the state sequence and/or the observability matrix of the dynamical system can be determined directly from input-output data is provided in Section 2.2. Variations of the basic algorithm and implementation issues are covered in Sections 2.5-2.7. In Section 3, we present some existing identification algorithms for linear time-varying systems. Section 4 covers the bulk of the existing subspace-based nonlinear identification algorithms including Hammerstein and nonlinear feedback identification (Section 4.1), Hammerstein-Wiener identification (Section 4.2), identification for Wiener systems (Section 4.3), linear parameter-varying system identification (Section 4.4) and bilinear system identification (Section 4.5). A brief overview of identification of closed-loop systems is presented in Section 5 and some illustrative examples are presented in Section 6.

2. Subspace Identification of Linear Time-Invariant Systems

Consider the linear time-invariant system

$$
\begin{aligned}
x_{k+1} &= Ax_k + Bu_k + w_k, & (2.1) \\
y_k &= Cx_k + Du_k + v_k, & (2.2)
\end{aligned}
$$

with

$$
\mathbf{E}\left[\begin{pmatrix} w_p \\ v_p \end{pmatrix} \begin{pmatrix} w_q^{\mathrm{T}} & v_q^{\mathrm{T}} \end{pmatrix} \right] = \begin{bmatrix} Q & S \\ S^{\mathrm{T}} & R \end{bmatrix} \delta_{pq} \geq 0, \tag{2.3}
$$

where $x, w \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $y, v \in \mathbb{R}^l$, $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{l \times n}$, $D \in \mathbb{R}^{l \times m}$, $Q \in \mathbb{R}^{n \times n}$, $S \in \mathbb{R}^{n \times l}$ and $R \in \mathbb{R}^{l \times l}$. Here $\mathbf{E}$ is the Expected value and δ is the kronecker delta. The identification problem can then be stated as

Problem statement:

Given N consecutive measurements of inputs u and outputs y, estimate the order n of the system and the system matrices A, B, C, D, Q, R, and S.

2.1. Notation

Let i and j be user-defined integers such that $i \geq n$ and $j \gg i$. Then we define the input block Hankel matrix as

S. Lacy is with the Air Force Base, Albuquerque, New Mexico

H. J. Palanth, J. B. Hoagg and D. S. Bernstein are with the department of Aerospace Engineering at the University of Michigan, Ann Arbor, MI 48109-2140. {hpalanth,hoagg,dsbaero}@umich.edu

$$U_{0|2i-1} \triangleq \begin{pmatrix} u_0 & u_1 & \cdots & u_{j-1} \\ u_1 & u_2 & \cdots & u_j \\ \vdots & \vdots & \ddots & \vdots \\ \hline u_{i-1} & u_i & \cdots & u_{i+j-2} \\ \hline u_i & u_{i+1} & \cdots & u_{i+j-1} \\ u_{i+1} & u_{i+2} & \cdots & u_{i+j} \\ \vdots & \vdots & \ddots & \vdots \\ u_{2i-1} & u_{2i} & \cdots & u_{2i+j-2} \end{pmatrix} \quad (2.4)$$

$$= \left(\frac{U_{0|i-1}}{U_{i|2i-1}} \right) = \left(\frac{U_p}{U_f} \right). \quad (2.5)$$

Also, when partitioned differently

$$U_{0|2i-1} = \begin{pmatrix} u_0 & u_1 & \cdots & u_{j-1} \\ u_1 & u_2 & \cdots & u_j \\ \vdots & \vdots & \ddots & \vdots \\ u_{i-1} & u_i & \cdots & u_{i+j-2} \\ \hline u_i & u_{i+1} & \cdots & u_{i+j-1} \\ \hline u_{i+1} & u_{i+2} & \cdots & u_{i+j} \\ \vdots & \vdots & \ddots & \vdots \\ u_{2i-1} & u_{2i} & \cdots & u_{2i+j-2} \end{pmatrix} \quad (2.6)$$

$$= \left(\frac{U_{0|i}}{U_{i+1|2i-1}} \right) = \left(\frac{U_p^+}{U_f^-} \right). \quad (2.7)$$

The subscript p denotes the 'past' and the subscript f denotes 'future'.

The output block Hankel matrices $Y_{0|2i-1}$, Y_p, Y_f, Y_p^+ and Y_f^- are defined similar to equations (2.4)-(2.7) with u replaced by y.

The extended observability matrix is

$$\Gamma_i \triangleq \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{i-1} \end{bmatrix}. \quad (2.8)$$

Under the assumption that (A, C) is observable, the extended observability matrix Γ_i has rank n (note that $i \geq n$).

The state sequence X_i is

$$X_i \triangleq \begin{pmatrix} x_i & x_{i+1} & \cdots & x_{i+j-2} & x_{i+j-1} \end{pmatrix}. \quad (2.9)$$

And finally, the notation F/G represents the orthogonal projection of the row space of F onto the row space of G. Similarly $F/_H G$ denotes the oblique projection of the row space of F along the row space of H onto the row space of G. The procedure for computing the orthogonal projection and the oblique projection are discussed in Sections 2.3 and 2.4, for more details see [5, 34].

The most important idea behind subspace algorithms is that the state sequence X_i can be directly estimated as an intersection (explained later) of the input and output block Hankel matrices. In the next subsection, we provide a simplified proof of why this is true for the case in which no noise terms are present. And in the subsequent subsections we will elaborate on the actual methodology for computing the state sequence, the order of the system, and the state-space matrices.

2.2. Deterministic Case

Here we consider a linear system without any unmeasured disturbances and noise terms

$$\begin{align} x_{k+1} &= Ax_k + Bu_k, & (2.10) \\ y_k &= Cx_k + Du_k. & (2.11) \end{align}$$

From (2.10) and (2.11), we can write

$$\begin{align} Y_{0|i-1} &= \Gamma_i X_0 + H_i U_{0|i-1} & (2.12) \\ Y_{i|2i-1} &= \Gamma_i X_i + H_i U_{i|2i-1}, & (2.13) \end{align}$$

where $H_i \in \mathbb{R}^{li \times mi}$ is the lower block triangular Toeplitz matrix and is defined as

$$H_i = \begin{bmatrix} D & 0 & 0 & \cdots & 0 \\ CB & D & 0 & \cdots & 0 \\ CAB & CB & D & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ CA^{i-2}B & CA^{i-3}B & \cdots & \cdots & D \end{bmatrix}. \quad (2.14)$$

From (2.12), we have

$$\begin{bmatrix} Y_{0|i-1} \\ U_{0|i-1} \end{bmatrix} = \begin{bmatrix} \Gamma_i & H_i \\ 0 & I_{mi} \end{bmatrix} \begin{bmatrix} X_0 \\ U_{0|i-1} \end{bmatrix} \quad (2.15)$$

and hence,

$$\text{rank} \begin{bmatrix} Y_{0|i-1} \\ U_{0|i-1} \end{bmatrix} = \text{rank} \begin{bmatrix} X_0 \\ U_{0|i-1} \end{bmatrix}. \quad (2.16)$$

Now, under the 'persistent excitation' assumption (i.e. $U_{0|i-1}$ is of full row rank, there is no intersection between the row spaces of X_0 and $U_{0|i-1}$ and the state sequence X_0 is of full row rank), we have

$$\text{rank} \begin{bmatrix} Y_{0|i-1} \\ U_{0|i-1} \end{bmatrix} = mi + n. \quad (2.17)$$

Similarly, it can also be shown that

$$\begin{align} \text{rank} \begin{bmatrix} Y_{i|2i-1} \\ U_{i|2i-1} \end{bmatrix} &= mi + n, & (2.18) \\ \text{rank} \begin{bmatrix} Y_{0|2i-1} \\ U_{0|2i-1} \end{bmatrix} &= 2mi + n. & (2.19) \end{align}$$

Next we can also relate X_i and X_0 as

$$X_i = A^i X_0 + \Delta_i^r U_{0|i-1}, \quad (2.20)$$

where

$$\Delta_i^r \triangleq \begin{bmatrix} A^{i-1}B & A^{i-2}B & \cdots & B \end{bmatrix} \quad (2.21)$$

is the reversed extended observability matrix. Assuming that the model is observable and noting that $i \geq n$, it follows from (2.13) that

$$X_i = \begin{bmatrix} -\Gamma_i^\dagger H_i & \Gamma_i^\dagger \end{bmatrix} \begin{bmatrix} U_{i|2i-1} \\ Y_{i|2i-1} \end{bmatrix} \quad (2.22)$$

which implies that the row space of X_i is contained in the row space of $\begin{bmatrix} U_f \\ Y_f \end{bmatrix}$. Similarly, from (2.20) and (2.12) we can write

$$\begin{aligned} X_i &= A^i \begin{bmatrix} \Gamma_i^\dagger Y_{0|i-1} - \Gamma_i^\dagger H_i U_{0|i-1} \end{bmatrix} + \Delta_i^r U_{0|i-1} \\ &= \begin{bmatrix} \Delta_i^r - A^i \Gamma_i^\dagger H_i & A^i \Gamma_i^\dagger \end{bmatrix} \begin{bmatrix} U_{0|i-1} \\ Y_{0|i-1} \end{bmatrix}, \quad (2.23) \end{aligned}$$

which implies that the row space of X_i is also contained in the row space of $\begin{bmatrix} U_p \\ Y_p \end{bmatrix}$.

Now, using the Grassmann's dimension theorem ([2] Theorem 2.3.1) gives (under the assumption of persistant excitations)

$$\begin{aligned} \dim &\left(\text{row space} \begin{bmatrix} U_p \\ Y_p \end{bmatrix} \bigcap \text{row space} \begin{bmatrix} U_f \\ Y_f \end{bmatrix} \right) \\ &= \text{rank} \begin{bmatrix} U_p \\ Y_p \end{bmatrix} + \text{rank} \begin{bmatrix} U_f \\ Y_f \end{bmatrix} - \text{rank} \begin{bmatrix} U_p \\ Y_p \\ U_f \\ Y_f \end{bmatrix} \\ &= (mi+n) + (mi+n) - (2mi+n) = n \end{aligned}$$

Hence, we have shown that any basis for the intersection between 'past' and 'future' represents a valid state sequence X_i. There are several ways to compute the intersection of matrices, one popular method is by computing the singular value decomposition of stacked up matrix containing the past and future block Hankel matrices. The state sequence X_{i+1} can be obtained analogously. Once the state sequence is available, the system matrices are estimated in a least squares step elaborated in Section 2.4.

2.3. Output-Only Identification

Consider the case in which there are no inputs u_k, the only signals that drive the equations are the noise terms w_k and v_k. Although the signals w_k and v_k are not available, identification of the state space matrices A, C, Q, S and R can be performed from output only data. This scenario arises in applications such as identification of the dynamics of bridges or structures excited by ambient disturbances. The system equations are thus

$$\begin{aligned} x_{k+1} &= A x_k + w_k, \quad (2.24) \\ y_k &= C x_k + v_k. \quad (2.25) \end{aligned}$$

The LQ decomposition of the block Hankel matrix $Y_{0|2i-1}$ is computed and is partitioned as

$$\begin{bmatrix} Y_{0|i-1} \\ Y_{i|i} \\ Y_{i+1|2i-1} \end{bmatrix} = \begin{bmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \\ Q_3^T \end{bmatrix} \quad (2.26)$$

Then the orthogonal projection Y_f/Y_p of the future output space onto the past output space is

$$\mathcal{O}_i \triangleq Y_f/Y_p = \begin{bmatrix} L_{21} \\ L_{31} \end{bmatrix} Q_1^T. \quad (2.27)$$

It can be shown that the matrix $\mathcal{O}_i$ is equal to the product of the extended observability matrix and the Kalman filter state sequence $\hat{X}_i$

$$\mathcal{O}_i = \Gamma_i \hat{X}_i. \quad (2.28)$$

Similarly it can be shown that the projection $\mathcal{O}_{i+1} \triangleq Y_f^-/Y_p^+$ is equal to

$$\mathcal{O}_{i+1} = \Gamma_{i-1} \hat{X}_{i+1} \quad (2.29)$$

where Γ_{i-1} is the matrix obtained by discarding the last l rows from Γ_i. Now, the Kalman filter state sequence can be computed by performing a singular value decomposition of $\mathcal{O}_i$. Once the estimates of the state sequence are available, a linear least squares problem for estimating the state space matrices A and C can be solved. The residuals from the least squares problem are then used to estimate Q, S and R. The details of the above procedure are presented in the next subsection.

2.4. Complete Stochastic-Deterministic Case

Consider the complete equations for the linear system (2.1) and (2.2). When there are no noise terms present, i.e. when $v_k \equiv 0$ and $w_k \equiv 0$, the state sequence can be determined by computing the intersection of the 'past' and 'future' data matrices, $\begin{bmatrix} U_p \\ Y_p \end{bmatrix}$ and $\begin{bmatrix} U_f \\ Y_f \end{bmatrix}$. In the presence of unmeasured noise terms, we obliquely project the row space of Y_f along the row space of U_f onto the row space of $\begin{bmatrix} U_p \\ Y_p \end{bmatrix}$. Then we compute the state sequence by computing a singular value decomposition of the projection and retaining the first few dominant singular values.

First to compute the oblique projection, we partition the LQ decomposition of $\begin{bmatrix} U \\ Y \end{bmatrix}$ as follows

$$\begin{bmatrix} U_{0|i-1} \\ U_{i|i} \\ U_{i+1|2i-1} \\ Y_{0|i-1} \\ Y_{i|i} \\ Y_{i+1|2i-1} \end{bmatrix} = \begin{bmatrix} L_{11} & 0 & 0 & 0 & 0 & 0 \\ L_{21} & L_{22} & 0 & 0 & 0 & 0 \\ L_{31} & L_{32} & L_{33} & 0 & 0 & 0 \\ L_{41} & L_{42} & L_{43} & L_{44} & 0 & 0 \\ L_{51} & L_{52} & L_{53} & L_{54} & L_{55} & 0 \\ L_{61} & L_{62} & L_{63} & L_{64} & L_{65} & L_{66} \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \\ Q_3^T \\ Q_4^T \\ Q_5^T \\ Q_6^T \end{bmatrix} \quad (2.30)$$

Now, the oblique projection $\mathcal{O}_i$ is

$$\begin{aligned} \mathcal{O}_i &= Y_f /_{U_f} \begin{bmatrix} U_p \\ Y_p \end{bmatrix}, \quad (2.31) \\ &= L_{U_p} L_{11} Q_1^T + L_{Y_p} \begin{bmatrix} L_{41} & L_{42} & L_{43} & L_{44} \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \\ Q_3^T \\ Q_4^T \end{bmatrix}, \quad (2.32) \end{aligned}$$

where

$$
\begin{bmatrix} L_{U_p} & L_{U_f} & L_{Y_p} \end{bmatrix}
\begin{bmatrix}
L_{11} & 0 & 0 & 0 \\
L_{21} & L_{22} & 0 & 0 \\
L_{31} & L_{32} & L_{33} & 0 \\
L_{41} & L_{42} & L_{43} & L_{44}
\end{bmatrix}
$$

$$
= \begin{bmatrix}
L_{51} & L_{52} & L_{53} & L_{54} \\
L_{61} & L_{62} & L_{63} & L_{64}
\end{bmatrix}. \tag{2.33}
$$

Similarly the projection $Y_f^- /_{U_f^-} \begin{bmatrix} U_p^+ \\ Y_p^+ \end{bmatrix}$ denoted by $\mathcal{O}_{i+1}$ is

$$
\mathcal{O}_{i+1} = L_{U_p^+} \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} Q_1^{\mathrm{T}} \\ Q_2^{\mathrm{T}} \end{bmatrix} \tag{2.34}
$$

$$
+ L_{Y_p^+} \begin{bmatrix}
L_{41} & L_{42} & L_{43} & L_{44} & 0 \\
L_{51} & L_{52} & L_{53} & L_{54} & L_{55}
\end{bmatrix}
\begin{bmatrix}
Q_1^{\mathrm{T}} \\
Q_2^{\mathrm{T}} \\
Q_3^{\mathrm{T}} \\
Q_4^{\mathrm{T}} \\
Q_5^{\mathrm{T}}
\end{bmatrix}, \tag{2.35}
$$

where

$$
\begin{bmatrix} L_{U_p^+} & L_{U_f^-} & L_{Y_p^+} \end{bmatrix}
\begin{bmatrix}
L_{11} & 0 & 0 & 0 & 0 \\
L_{21} & L_{22} & 0 & 0 & 0 \\
L_{31} & L_{32} & L_{33} & 0 & 0 \\
L_{41} & L_{42} & L_{43} & L_{44} & 0 \\
L_{51} & L_{52} & L_{53} & L_{54} & L_{55}
\end{bmatrix}
$$

$$
= \begin{bmatrix} L_{61} & L_{62} & L_{63} & L_{64} & L_{65} \end{bmatrix}. \tag{2.36}
$$

Under the observations that:

1) the process noise w_k and measurement noise v_k are uncorrelated with the inputs u_k,
2) the inputs u_k are persistently exciting of order $2i$, i.e. the block input Hankel matrix $U_{0|2i-1}$ is of full row rank,
3) the sample size goes to infinity: $j \to \infty$,
4) the process noise w_k and the measurement noise v_k are not identically zero,

it has been shown that the oblique projection $\mathcal{O}_i$ is equal to the product of the extended observability matrix Γ_i and a sequence of Kalman filter states

$$
\mathcal{O}_i = \Gamma_i \hat{X}_i. \tag{2.37}
$$

1) Order Determination and State Estimates: The order of the system can be determined by examining the singular values of the matrix $\mathcal{O}_i$. In the absence of noise terms, there will be only n non-zero singular values of $\mathcal{O}_i$. However, when noise terms are present, there will generally be more than n non-zero singular values of $\mathcal{O}_i$. In this case the dominant singular values are retained, the rest are ignored, and the order of the system is estimated as the number of dominant singular values.

Once the order of the system is estimated, the Kalman filter states $\hat{X}_i$ can be recovered from the singular value decomposition of $\mathcal{O}_i$. For example, let the singular value decomposition of $L_{U_p} \begin{bmatrix} L_{11} & 0 & 0 & 0 \end{bmatrix} +$

$L_{Y_p} \begin{bmatrix} L_{41} & L_{42} & L_{43} & L_{44} \end{bmatrix}$ be

$$
L_{U_p} \begin{bmatrix} L_{11} & 0 & 0 & 0 \end{bmatrix} + L_{Y_p} \begin{bmatrix} L_{41} & L_{42} & L_{43} & L_{44} \end{bmatrix}
$$

$$
= \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix} \begin{bmatrix} V_1^{\mathrm{T}} \\ V_2^{\mathrm{T}} \end{bmatrix}. \tag{2.38}
$$

where S_1 contains the dominant singular values and S_2 are the singular values to be neglected. Then from (2.32), (2.37) and (2.38), we have

$$
\Gamma_i = U_1 S_1^{1/2}, \tag{2.39}
$$

and

$$
\hat{X}_i = \Gamma_i^{\dagger} \mathcal{O}_i = S_1^{1/2} V_1^{\mathrm{T}} \begin{bmatrix} Q_1^{\mathrm{T}} \\ Q_2^{\mathrm{T}} \\ Q_3^{\mathrm{T}} \\ Q_4^{\mathrm{T}} \end{bmatrix}. \tag{2.40}
$$

Similarly $\hat{X}_{i+1}$ can be computed as $\Gamma_{i-1}^{\dagger} \mathcal{O}_{i+1}$, where Γ_{i-1} is the matrix obtained by discarding the last l rows of Γ_i.

2) State-Space Matrices: Once the state sequence is obtained, the system matrices can be solved for as a least squares problem

$$
\begin{bmatrix} X_{i+1} \\ Y_{i|i} \end{bmatrix} = \begin{bmatrix} \hat{A} & \hat{B} \\ \hat{C} & \hat{D} \end{bmatrix} \begin{bmatrix} X_i \\ U_{i|i} \end{bmatrix}. \tag{2.41}
$$

The residuals of the above least squares solution are then used to calculate the noise covariances $\hat{Q}$, $\hat{S}$, and $\hat{R}$. It should be noted here that the state sequence and the state-space matrices can be estimated to only within a basis transformation of the actual matrices. Hence the estimated states cannot be ascribed any physical significance.

2.5. Additional Variations

Several variants on the above algorithm exist. First, we note that the oblique projection $\mathcal{O}_i$ can be weighted left and right by user defined weighting matrices $W_1 \in \mathbb{R}^{li \times li}$ and $W_2 \in \mathbb{R}^{j \times j}$ respectively, which must satisfy the following conditions: W_1 is full rank and the rank of $\begin{bmatrix} U_p \\ Y_p \end{bmatrix} W_2$ is equal to the rank of $\begin{bmatrix} U_p \\ Y_p \end{bmatrix}$. Furthermore, one can distinguish between two classes of subspace identification algorithms. The first class uses the state estimates $\hat{X}_i$ (the right singular vectors of $W_1 \mathcal{O}_i W_2$) to find the system matrices. The algorithm in Section 6.2 belongs to this class. The second class of algorithms uses the extended observability matrix Γ_i (the left singular vectors of $W_1 \mathcal{O}_i W_2$) to first determine estimates of A and C and subsequently of $B; D$ and $Q; S; R$. In fact, the three subspace algorithms (N4SID [33], MOESP [42] and CVA [18]) all start from $W_1 \mathcal{O}_i W_2$ [5, 34] with a specific choice of weighting matrices W_1 and W_2 for each of the algorithms. The weightings are summarized in Table 1. From this table is clear that the algorithm described above is the N4SID algorithm ($W_1 = I_{li}$ and $W_2 = I_j$).

All of the algorithms in Table I first calculate an oblique projection $\mathcal{O}_i$, followed by an SVD of the weighted matrix

Acronym	W_1	W_2
N4SID	I_{li}	I_j
CVA	$\left(\lim_{j \to \infty} \frac{1}{j} [(Y_f/U_f)(Y_f/U_f)^{\mathrm{T}}] \right)^{-1/2}$	$\Pi_{U_f^{\perp}}$
MOESP	I_{li}	$\Pi_{U_f^{\perp}}$

TABLE I

THIS TABLE DEFINES VARIATIONS OF VARIOUS SUBSPACE
IDENTIFICATION ALGORITHMS.

$W_1 \mathcal{O}_i W_2$. The first two algorithms, N4SID and CVA, use the state estimates $\hat{X}_i$ (the right singular vectors) to find the system matrices, while MOESP is based on the extended observability matrix Γ_i (the left singular vectors). The matrix $U_f^{\perp}$ in the weights of CVA and MOESP represents the orthogonal complement of the row space of U_f.

2.6. Implementation

The basic subspace algorithm and its variants have been incorporated into several commercial software for system identification [5]. In most of the softwares, choice of the parameters i and j is automated so there are no user choices (for the advanced user, an option to change some of the parameters is provided). Hence these software implementations are easy to use.

- The System Identification Toolbox in Matlab [24], developed by Prof. L. Ljung (Linköping, Sweden): http://www.mathworks.com/products/sysid/
- The system identification package ADAPTx [20] of Adaptics, Inc, developed by Dr. W. E. Larimore: http://www.adaptics.com/
- The ISID-module in Xmath, developed by dr. P. Van Overschee and Prof. B. De Moor and licensed to ISI Inc. (now Wind River), USA: http://www.windriver.com
- The software packages RaPID and INCA of IPCOS International: http://www.ipcos.be
- The package MACEC, developed at the department of Civil Engineering of the K.U.Leuven in Belgium: http://www.kuleuven.ac.be/bwm/macec/
- Products of LMS International: http://www.lms-international.com

Other public domain software includes SLICOT (http://www.win.tue.nl/niconet/NIC2/slicot.html), the SMI toolbox of the Control Laboratory at the T.U.Delft (http://lcewww.et.tudelft.nl/ verdult/smi/) and the Cambridge University System Identification Toolbox (http://www-control.eng.cam.ac.uk/jmm/).

2.7. Notes

- The state-space matrices calculated in 2.41 and the estimated state-sequence $\hat{X}_i$ can be guaranteed only to within a basis transformation of the actual state-space matrices and the state-sequence respectively. So there exists a nonsingular matrix T such that

$$\begin{aligned} A &= T\hat{A}T^{-1}, \ B = T\hat{B}, \\ C &= \hat{C}T^{-1}, \ D = \hat{D}, \ X_i = T\hat{X}_i. \end{aligned}$$

- It is often difficult for the user to verify whether the inputs satisfy all of the assumptions. But in cases where the user is able to perform experiments on the system with inputs of his/her choice, it is better to choose inputs with high frequency content to have a better chance of being persistently exciting. In particular signals with low frequency content like sinusoids and constants should be avoided.
- One often mentioned disadvantage of subspace methods is the fact that these methods do not optimize a maximum likelihood cost function. The reason for this is that, contrary to input-output models (transfer matrices), we cannot formulate a likelihood function for identifying state space models, that also leads to an amenable optimization problem. Nevertheless, subspace identification algorithms provide surprisingly good approximations of linear models, but there is still a lot of ongoing research on how the identified model relates to a maximum likelihood formulation. Yet, it is our experience that subspace algorithms often tend to give very good linear models for real world data sets.
- Bias in inputs can be a concern when data is obtained from sensors that are not calibrated. This can be handled by using an additional artificial input with value 1 at all time steps. Consider

$$\begin{aligned} x_{k+1} &= Ax_k + B(u_k - u^0) + w_k \\ y_k &= Cx_k + D(u_k - u^0) + v_k \end{aligned}$$

where u_k is the measured inputs, while $u_k - u^0$ is the actual inputs to the system with u^0 a constant. The above equations can be rewritten as

$$\begin{aligned} x_{k+1} &= Ax_k + \begin{bmatrix} B & -Bu^0 \end{bmatrix} \begin{bmatrix} u_k \\ 1 \end{bmatrix} + w_k, \\ y_k &= Cx_k + \begin{bmatrix} D & -Du^0 \end{bmatrix} \begin{bmatrix} u_k \\ 1 \end{bmatrix} + v_k. \end{aligned}$$

Thus, a bias in inputs can be translated into an additional artificial scalar input 1. Notice here that u^0 is a vector, whereas 1 denotes a scalar.
- Subspace algorithms can be easily modified to accommodate for missing data points. In the Hankel matrices (2.4), the columns containing the missing data points are removed (the corresponding columns from $Y_{0|2i-1}$ are also removed) and the rest of the procedure is carried out without any modifications. This will yield a state sequence with jumps where the data is missing. In the least squares step (2.41), care is taken to not include the states at the jump (because $\hat{x}_{k+1}$ will correspond to a time step before the jump and $\hat{x}_k$ will correspond to a time step after the jump). The same procedure can also be used when data from several different experiments performed at different times are available.

The time between the experiments can just be regarded as missing data (even if it is several days).

3. LINEAR TIME-VARYING SYSTEMS

3.1. Arbitrarily Time-Varying Systems

In [44] for an arbitrary time-varying system, the row space of the observability matrix is estimated using a QR factorization followed by a singular value decomposition similar to the ones in section 2.3. However, here the observability matrix is different for each time step. So the block Hankel matrices defined in section 2.1 cannot be used since they include data points from different time steps. To deal with this problem, an ensemble of input-output data is used. The ensemble data set has to be collected when the system exhibits the same time-varying behavior. With this ensemble data set, the block Hankel matrices are defined in a different way. Using these block Hankel matrices, the row space of the observability matric can be estimated using a QR factorization of the block Hankel matrices followed by a singular value decomposition. For details see [44].

Subspace identification for switching systems is developed in [30]. Although several of the steps in the algorithm are derived from the basic subspace identification algorithm, neither the state sequence nor the observability matrix is computed. Instead the Markov parameters of the system are computed directly from the available data, and thus this algorithm cannot be considered a true subspace algorithm.

3.2. Periodically Time-Varying Systems

In the case of periodic systems [14] lifting can be used to express the system as a set of p time-invariant systems, where p is the period of the system. For example, consider the system

$$x_{k+1} = A_k x_k + B_k u_k + w_k, \quad (3.1)$$
$$y_k = C_k x_k + D_k u_k + v_k. \quad (3.2)$$

The matrices A_k, B_k, C_k and D_k have the same dimensions as A, B, C and D respectively in (2.1) and (2.2), but now the matrices are periodic with a period p, i.e. for all k,

$$A_k = A_{k+p}, \ B_k = B_{k+p}, \ C_k = C_{k+p}, \ D_k = D_{k+p}. \ (3.3)$$

Now, a time-invariant representation of the periodic linear system (3.1) and (3.2) can be obtained in a standard way [25]. Once $k = \kappa$ is fixed, the time-invariant system that maps x_κ to $x_{\kappa+p}$ can be written as the period-mapped system

$$\bar{x}_{k+1}[\kappa] = \bar{A}[\kappa]\bar{x}_k[\kappa] + \bar{B}[\kappa]\bar{u}_k[\kappa], \quad (3.4)$$
$$\bar{y}_k[\kappa] = \bar{C}[\kappa]\bar{x}_k[\kappa] + \bar{D}[\kappa]\bar{u}_k[\kappa]. \quad (3.5)$$

For example, the system matrices associated with the period-mapped system with $\kappa = 1$ can be written as:

$$\bar{A}[1] = A_p A_{p-1} \cdots A_1$$

$$\bar{B}[1] = \left[\begin{array}{ccccc} A_p A_{p-1} \cdots A_2 B_1 & A_p \cdots A_3 B_2 & \cdots & A_p B_{p-1} & B_p \end{array} \right]$$

$$\bar{C}[1] = \left[\begin{array}{c} C - 1 \\ C_2 A_1 \\ \vdots \\ C_p A_{p-1} \cdots A_1 \end{array} \right]$$

$$\bar{D}[1] = \left[\begin{array}{cccc} D_1 & 0 & \cdots & 0 \\ C_2 B_1 & D_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ C_{p-1} A_{p-2} \cdots A_2 B_1 & C_{p-1} A_{p-2} \cdots A_3 B_2 & \cdots & 0 \\ C_p A_{p-1} \cdots A_2 B_1 & C_p A_{p-1} \cdots A_3 B_2 & \cdots & D_p \end{array} \right]$$

Note that a state sequence of the periodic-mapped system is a subsequence of the periodic system, i.e.

$$\bar{x}_k[\kappa] = x_{\kappa+pk} \quad (3.6)$$

and the augmented input and output vectors $\bar{u}_k[\kappa]$ and $\bar{y}_k[\kappa]$ are created by stacking up the input and output vectors u_k and y_k, respectively:

$$\bar{u}_k[\kappa] = \left[\begin{array}{c} u_{\kappa+pk} \\ u_{\kappa+pk+1} \\ \cdots \\ u_{\kappa+p(k+1)-1} \end{array} \right] \quad (3.7)$$

$$\bar{y}_k[\kappa] = \left[\begin{array}{c} y_{\kappa+pk} \\ y_{\kappa+pk+1} \\ \cdots \\ y_{\kappa+p(k+1)-1} \end{array} \right] \quad (3.8)$$

Therefore the linear periodic time-varying system (3.1) and (3.2) can be written as a set of p time-invariant equations given by (3.4) and (3.5) with $\kappa = (1, 2, \ldots, p)$. Now defining the state sequence $X[\kappa]$ as

$$X_i[\kappa] = \left[\begin{array}{cccc} x_{\kappa+ip} & x_{\kappa+(i+1)p} & \cdots & x_{\kappa+(i+j-1)p} \end{array} \right], \quad (3.9)$$

the state sequence can be calculated as described in section 2.3 for $\kappa = (1, 2, \ldots, p)$. Once the state sequences are calculated, the state-space matrices can be calculated from the system of linear equations

$$\left[\begin{array}{c} X_i[\kappa+1] \\ \mathcal{Y}_i[\kappa] \end{array} \right] = \left[\begin{array}{cc} A_\kappa & B_\kappa \\ C_\kappa & D_\kappa \end{array} \right] \left[\begin{array}{c} X_i[\kappa] \\ \mathcal{U}_i[\kappa] \end{array} \right] \quad (3.10)$$

where

$$\mathcal{U}_i[\kappa] = \left[\begin{array}{cccc} u_{\kappa+ip} & u_{\kappa+(i+1)p} & \cdots & u_{\kappa+(i+j-1)p} \end{array} \right],$$
$$\mathcal{Y}_i[\kappa] = \left[\begin{array}{cccc} y_{\kappa+ip} & y_{\kappa+(i+1)p} & \cdots & y_{\kappa+(i+j-1)p} \end{array} \right].$$

This procedure involves computing $p+1$ state sequences, singular value decompositions and LQ decompositions which might be expensive for large p. In [14], a procedure for calculating $X_i[\kappa + r]$ for $r = (1, \ldots, p)$ based on the singular value decomposition of $X_i[\kappa]$ is presented. This procedure substantially reduces the computational load of calculating $p + 1$ state sequences.

Note: In [14], page 297, equation 27, the expression for $\mathcal{X}[\kappa + r]$ for $1 \leq r \leq p - 1$ is misprinted. The actual expression should read

$$\mathcal{X}[\kappa+r] = \tilde{U}_{qb}^\mathrm{T}[\kappa, r]\tilde{U}_{12}^\mathrm{T}[\kappa, r]\tilde{U}[\kappa, r](1 : pi(lu + ly), :)S_a[\kappa, r]\tilde{V}^\mathrm{T} V_b^\mathrm{T}[\kappa]$$
$$1 \leq r \leq p - 1$$

4. NONLINEAR SYSTEMS

Nonlinear identification is an increasingly active research area. The most common model structures are the Hammerstein, Wiener, nonlinear feedback, and combined Hammerstein/nonlinear feedback models shown in Figure 1, Figure 2, Figure 3, and Figure 4. These models involve the interconnection of a single linear block and a single nonlinear block. Identification with these model structures has been widely considered, see, for example, [1, 3, 11–13, 23, 27, 35, 43, 47] . A point that has not been stressed is the fact that nonlinear identification with the Wiener model structure is significantly more difficult than identification with the Hammerstein structure. The reason for this difficulty is the fact that identification of a nonlinear map is more tractable when a measurement of the input to the map is available.

Other model structures include linear parameter-varying systems and bilinear systems. These models structures cannot be written as the interconnection of a static nonlinearity and linear dynamics. Subspace identification with these model structures is more difficult and require substantial modifications of linear subspace model identification algorithms. Nonetheless, these problems have been considered and solutions to these identification problems exist [36, 37, 39, 40]. A good overview of these methods can be found in [36].

The Hammerstein nonlinear/feedback model can be viewed as a realization of a nonlinear system. Accordingly, the representation of the nonlinearities is not unique, and thus it is not possible to uniquely identify the nonlinear maps [35]. The identified maps could differ by a scaling coefficient or a bias from the real nonlinearity.
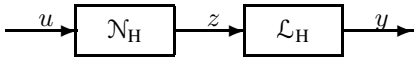

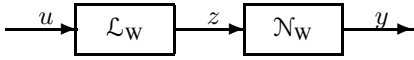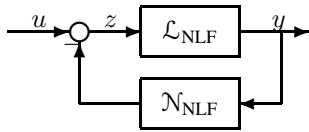Fig. 1.   Hammerstein Model

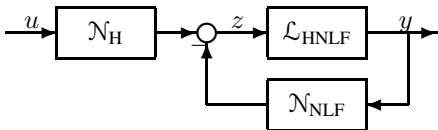Fig. 2.   Wiener Model

Fig. 3.   Nonlinear Feedback Model

Fig. 4.   Hammerstein Nonlinear Feedback Model

Subspace-based Hammerstein identification algorithms are considered in [11, 15, 35, 43]. In view of the fact that nonlinear identification is facilitated by the availability of inputs, a general formulation of nonlinear identification in this setting is considered in [15] using a subspace identification algorithm [26, 34] along with a given basis expansion for the nonlinear maps. The function expansion is chosen to be linear in the parameters, which allows the nonlinear identification problem to be recast as a linear identification problem with generalized inputs. The multivariable capability of subspace identification algorithms is essential to this approach by allowing an arbitrary number of generalized inputs.

4.1. Hammerstein and Nonlinear Feedback Systems

Consider the nonlinear discrete-time system

$$x_{k+1} = Ax_k + F(u_k, y_k), \qquad (4.1)$$
$$y_k = Cx_k + G(u_k), \qquad (4.2)$$

where $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^m$, $y_k \in \mathbb{R}^p$, $A \in \mathbb{R}^{n \times n}$, $C \in \mathbb{R}^{p \times n}$, $F \colon \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^n$, and $G \colon \mathbb{R}^m \to \mathbb{R}^p$. The functions F and G can be written in terms of their scalar-valued components as

$$F(u,y) = \begin{bmatrix} F_1(u,y) \\ \vdots \\ F_n(u,y) \end{bmatrix}, \quad G(u) = \begin{bmatrix} G_1(u) \\ \vdots \\ G_p(u) \end{bmatrix}, \quad (4.3)$$

where, for all $i = 1, \ldots, n$, $F_i \colon \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ and, for all $i = 1, \ldots, p$, $G_i \colon \mathbb{R}^m \to \mathbb{R}$. By defining

$$z \stackrel{\triangle}{=} \mathcal{N}(u,y) \stackrel{\triangle}{=} \begin{bmatrix} F(u,y) \\ G(u) \end{bmatrix}, \qquad (4.4)$$

the system (4.1), (4.2) can be illustrated as in Figure 5, where $\mathcal{N} \colon \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^{n+p}$ and $\mathcal{L}$ represents the linear system

$$x_{k+1} = Ax_k + \begin{bmatrix} I_n & 0 \end{bmatrix} z_k, \qquad (4.5)$$
$$y_k = Cx_k + \begin{bmatrix} 0 & I_p \end{bmatrix} z_k, \qquad (4.6)$$

where $z_k \stackrel{\triangle}{=} \mathcal{N}(u_k, y_k)$ is viewed as an unmeasured, exogenous input to $\mathcal{L}$.
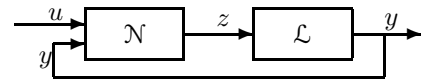

Fig. 5.   Nonlinear System with Measured-Input Nonlinearities

The main feature of the model (4.5), (4.6) is the fact that all of the inputs to $\mathcal{N}$ are measured. Therefore, the model (4.1), (4.2) includes the Hammerstein and nonlinear feedback models shown in Figure 1 and Figure 3. However, (4.1), (4.2) does not encompass the Wiener system shown in Figure 2.

Next, we assume that the components F_i and G_i can be expanded in terms of basis functions $f_1(u,y), \ldots, f_q(u,y)$, $g_1(u), \ldots, g_r(u)$, and $h_1(u), \ldots, h_s(u)$ as

$$F(u,y) = \begin{bmatrix} \sum_{i=1}^{q} b_{f1i} f_i(u,y) + \sum_{i=1}^{s} b_{h1i} h_i(u) \\ \vdots \\ \sum_{i=1}^{q} b_{fni} f_i(u,y) + \sum_{i=1}^{s} b_{hni} h_i(u) \end{bmatrix}, \quad (4.7)$$

$$G(u) = \begin{bmatrix} \sum_{i=1}^{r} d_{g1i} g_i(u) + \sum_{i=1}^{s} d_{h1i} h_i(u) \\ \vdots \\ \sum_{i=1}^{r} d_{gpi} g_i(u) + \sum_{i=1}^{s} d_{hpi} h_i(u) \end{bmatrix}. \quad (4.8)$$

The functions h_i are the basis functions that are common to both F and G. Defining $f\colon \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^q$, $g\colon \mathbb{R}^m \to \mathbb{R}^r$, and $h\colon \mathbb{R}^m \to \mathbb{R}^s$ by

$$f(u,y) = \left[\begin{array}{c} f_1(u,y) \\ \vdots \\ f_q(u,y) \end{array} \right], \; g(u) = \left[\begin{array}{c} g_1(u) \\ \vdots \\ g_r(u) \end{array} \right], \; h(u) = \left[\begin{array}{c} h_1(u) \\ \vdots \\ h_s(u) \end{array} \right],$$

it follows from (4.7) and (4.8) that

$$F(u,y) = B_f f(u,y) + B_h h(u), \qquad (4.9)$$
$$G(u) = D_g g(u) + D_h h(u), \qquad (4.10)$$

where $B_f \triangleq [b_{fij}] \in \mathbb{R}^{n \times q}$, $B_h \triangleq [b_{hij}] \in \mathbb{R}^{n \times s}$, $D_g \triangleq [d_{gij}] \in \mathbb{R}^{p \times r}$, and $D_h \triangleq [d_{hij}] \in \mathbb{R}^{p \times s}$. Thus (4.1), (4.2) can be written as

$$x_{k+1} = A x_k + B_f f(u_k, y_k) + B_h h(u_k), \qquad (4.11)$$
$$y_k = C x_k + D_g g(u_k) + D_h h(u_k), \qquad (4.12)$$

or more compactly as

$$x_{k+1} = A x_k + B \left[\begin{array}{c} f(u_k, y_k) \\ g(u_k) \\ h(u_k) \end{array} \right], \qquad (4.13)$$

$$y_k = C x_k + D \left[\begin{array}{c} f(u_k, y_k) \\ g(u_k) \\ h(u_k) \end{array} \right], \qquad (4.14)$$

where

$$B \triangleq \left[\begin{array}{ccc} B_f & 0 & B_h \end{array} \right], \quad D \triangleq \left[\begin{array}{ccc} 0 & D_g & D_h \end{array} \right]. \quad (4.15)$$

As a special case of the system shown in Figure 5, we can consider the Hammerstein system

$$x_{k+1} = A x_k + F(u_k), \qquad (4.16)$$
$$y_k = C x_k + G(u_k), \qquad (4.17)$$

where now the function F depends only on the input u. In the case that F and G are represented by a common set of basis function $h_1, \ldots, h_s$, it follows that

$$z = \left[\begin{array}{c} F(u) \\ G(u) \end{array} \right] = \left[\begin{array}{c} B \\ D \end{array} \right] h(u), \qquad (4.18)$$

where $B = B_h \in \mathbb{R}^{n \times s}$ and $D = D_h \in \mathbb{R}^{p \times s}$. Hence (4.16), (4.17) become

$$x_{k+1} = A x_k + B h(u_k), \qquad (4.19)$$
$$y_k = C x_k + D h(u_k). \qquad (4.20)$$

The goal of the nonlinear identification problem is to construct models of both $\mathcal{L}$ and $\mathcal{N}$ given measurements of (u_k, y_k) over the interval $0 \le k \le \ell$. The signal z is assumed to be unavailable. However, when $h(u)$ is approximated by $\hat{h}(u)$ and B, D are approximated by $\hat{B}, \hat{D}$, then the computed signal

$$\hat{z} \triangleq \left[\begin{array}{c} \hat{B} \\ \hat{D} \end{array} \right] h(u), \qquad (4.21)$$

is available as the input to $\mathcal{L}$.

Basis-function Selection

With the basis functions $f_i(u,y)$, $g_i(u)$, $h_i(u)$ specified, subspace identification algorithms [26, 34] can be applied directly to the system (4.5), (4.6) with the computed signal $\hat{z}$ playing the role of the exogenous input. This is the approach developed in [15]. However, the choice of basis functions remains the main difficulty.

The above approach requires a fixed set of basis functions to represent the nonlinear mapping. For nonlinear mappings of several inputs, the curse of dimensionality requires an excessively large number of basis functions. In addition, the subspace identification algorithm identifies only the coefficients of the basis functions, not the basis functions themselves. Hence, without prior knowledge of the form of the nonlinear mappings, it may be necessary to employ a large number of basis functions, rendering the problem numerically intractable.

To address these issues, two techniques for iteratively refining the basis function representation of nonlinear mappings that are functions of measured inputs are given in [28].

The first technique uses selective refinement to improve the representation of the nonlinear functions. By applying a singular value decomposition to the input matrix, the dominant nonlinearities are identified for the chosen set of basis functions. Next, a random collection of basis functions is introduced to improve the representation of the dominant nonlinearities. Iteration of these steps constitutes the selective refinement process.

The second algorithm optimizes a fixed set of basis functions by means of a BFGS quasi-Newton optimization code. The representation of the nonlinear map is systematically improved by modifying the basis functions rather than by including additional basis functions. A subspace identification algorithm is used to identify the linear dynamics for a chosen set of basis functions representing the nonlinear functions. For that particular set of state space matrices, the basis functions are then optimized using a quasi-Newton optimization algorithm.

Both techniques are flexible in their implementation. For example, arbitrary basis functions such as polynomials, splines, sigmoids, sinusoids, or radial basis functions can be used. Both approaches can be used to identify nonlinear maps of multiple arguments and with multiple outputs.

We now illustrate abovementioned approaches for the Hammerstein case (4.19), (4.20).

1) Selective Refinement Algorithm: To begin, consider an initial set of basis functions $\hat{h}_1, \ldots, \hat{h}_{\hat{s}}$ with $\hat{h} \triangleq \left[\begin{array}{ccc} \hat{h}_1 & \cdots & \hat{h}_{\hat{s}} \end{array} \right]^{\mathrm{T}}$ and let $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$ denote an estimate of (A, B, C, D) provided by the subspace identification algorithm. Next, consider the singular value decomposition of $\left[\begin{array}{c} \hat{B} \\ \hat{D} \end{array} \right]$ written in standard notation as

$$\left[\begin{array}{c} \hat{B} \\ \hat{D} \end{array} \right] = \hat{U} \hat{\Sigma} \hat{V}. \qquad (4.22)$$

Then, we retain the ν largest singular values in $\hat{\Sigma}$ to obtain the approximation $\hat{\Sigma} \approx \hat{\Sigma}_0 = \hat{L}_0 \hat{R}_0$, where $\operatorname{rank} \hat{\Sigma}_0 = \nu$ and the matrices $\hat{L}_0 \in \mathbb{R}^{(n+p) \times \nu}$ and $\hat{R}_0 \in \mathbb{R}^{\nu \times s}$ have full column rank and full row rank, respectively. The retained ν largest singular values can be incorporated into either $\hat{L}_0$ or $\hat{R}_0$, yielding the approximation

$$
\begin{aligned}
\left[\begin{array}{c} \hat{B} \\ \hat{D} \end{array}\right] \hat{h}(u) &= \hat{U}\hat{\Sigma}\hat{V}\hat{h}(u) \approx \hat{U}\hat{\Sigma}_0\hat{V}\hat{h}(u) \\
&= \hat{U}\hat{L}_0\hat{R}_0\hat{V}\hat{h}(u) = \left[\begin{array}{c} \hat{B}_0 \\ \hat{D}_0 \end{array}\right] \hat{h}_0(u), \quad (4.23)
\end{aligned}
$$

where the matrix $\left[\begin{array}{c} \hat{B}_0 \\ \hat{D}_0 \end{array}\right] \triangleq \hat{U}\hat{L}_0 \in \mathbb{R}^{(n+p) \times \nu}$ and $\hat{h}_0(u) \triangleq \hat{R}_0\hat{V}\hat{h}(u)$ satisfying $\hat{h}_0 \colon \mathbb{R}^m \to \mathbb{R}^\nu$ is a column vector consisting of ν scalar-valued nonlinear functions. The motivation for this procedure is to retain only ν scalar-valued nonlinear functions each of which is a linear combination of $\hat{s}$ basis functions. Since $\nu << \hat{s}$, the ν scalar-valued components of $\hat{h}_0$ can be viewed as *dominant nonlinearities*, while the choice of ν reflects the *rank* of the nonlinear mapping $\left[\begin{smallmatrix} F \\ G \end{smallmatrix}\right]$. Hence the number of dominant nonlinearities is effectively the rank of the nonlinear map.

To refine the mapping $\hat{h}_0$ we repeat the above procedure with a new set of basis functions $\hat{h}_1', \ldots \hat{h}_{\hat{s}'}'$ with $\hat{h}' \triangleq \left[\begin{array}{ccc} \hat{h}_1' & \cdots & \hat{h}_{\hat{s}'}' \end{array}\right]^{\mathrm{T}}$, where $\hat{h}_1', \ldots, \hat{h}_\nu'$ are chosen to be the ν components of $\hat{h}_0$, and $\hat{h}_{\nu+1}', \ldots \hat{h}_{\hat{s}'}'$ are chosen randomly. Repeating the above procedure yields a new estimate $\hat{h}_0'$ and the approximation

$$
\left[\begin{array}{c} \hat{B}' \\ \hat{D}' \end{array}\right] \hat{h}'(u) \approx \left[\begin{array}{c} \hat{B}_0' \\ \hat{D}_0' \end{array}\right] \hat{h}_0'(u), \quad (4.24)
$$

where $\hat{B}'$, $\hat{D}'$ are the estimates of B and D obtained from the subspace identification algorithm at the current iteration. Note that the components of the dominant nonlinearity $\hat{h}_0'$ are now linear combinations of $\hat{s} + \hat{s}'$ basis functions. However, the number of scalar components is fixed at ν.

2) Basis Function Optimization Algorithm: In the basis function optimization algorithm we optimize a fixed set of basis functions instead of introducing additional basis functions. A convenient choice of basis functions is radial basis functions because of the ease of programming and the ability to handle multi-dimensional inputs. Radial basis functions are of the form

$$
f(u) = e^{-\alpha \|u-c\|_2^2}, \quad (4.25)
$$

where α determines the spread of the function and c decides the center of the function. For nonlinear identification, we optimize a set of radial basis functions with respect to the parameters α and c and identify the linear dynamics using a subspace identification algorithm. By optimizing a fixed set of basis functions, a more accurate representation of the nonlinear mapping is obtainable with a smaller number of basis functions than is possible with the selective refinement algorithm.

The identification error is defined to be the mean square

error at the output y_k given by

$$
E(\alpha, c) = \frac{1}{2} \sum_{k=1}^{l} (y_k - \hat{y}_k)^2, \quad (4.26)
$$

where y_k and $\hat{y}_k$ are the desired and actual outputs of the identified Hammerstein system, and l is the length of the data set.

Now, writing $\hat{y}_k$ in terms of $\hat{A}, \hat{B}, \hat{C}$ and $\hat{D}$, (4.26) becomes

$$
\begin{aligned}
E(\alpha,c) = & \frac{1}{2} \sum_{k=1}^{l} (y_k - \hat{C}\hat{A}^k \hat{x}_0 \\
& - \sum_{i=0}^{k-1} \hat{C}\hat{A}^{k-i-1}\hat{B}\hat{h}(u_i) - \hat{D}\hat{h}(u_k))^2. \quad (4.27)
\end{aligned}
$$

Using a set of s radial basis functions for $\hat{h}(u)$ equation (4.27) becomes

$$
E(\alpha,c) = \frac{1}{2} \sum_{k=1}^{l} \left(y_k - \hat{C}\hat{A}^k \hat{x}_0 - \sum_{i=0}^{k-1} \hat{C}\hat{A}^{k-i-1}\hat{B} \left[\begin{array}{c} e^{-\alpha_1 \|u_i - c_1\|_2^2} \\ \vdots \\ e^{-\alpha_s \|u_i - c_s\|_2^2} \end{array}\right] \right.
$$
$$
\left. - \hat{D} \left[\begin{array}{c} e^{-\alpha_1 \|u_k - c_1\|_2^2} \\ \vdots \\ e^{-\alpha_s \|u_k - c_s\|_2^2} \end{array}\right] \right)^2. \quad (4.28)
$$

The gradient of $E(\alpha, c)$ with respect to the parameters α_j and c_j can be calculated as

$$
\begin{aligned}
\frac{\partial E}{\partial \alpha_j} = (y_k - \hat{y}_k) \sum_{k=1}^{l} \left(-\sum_{i=0}^{k-1} \hat{C}\hat{A}^{k-i-1}\hat{B} \left[\begin{array}{c} 0 \\ \vdots \\ \frac{\partial}{\partial \alpha_j} e^{-\alpha_j \|u_i - c_j\|_2^2} \\ \vdots \\ 0 \end{array}\right] \right. \\
\left. - \hat{D} \left[\begin{array}{c} 0 \\ \vdots \\ \frac{\partial}{\partial \alpha_j} e^{-\alpha_j \|u_k - c_j\|_2^2} \\ \vdots \\ 0 \end{array}\right] \right) \quad (4.29)
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial E}{\partial c_j} = (y_k - \hat{y}_k) \sum_{k=1}^{l} \left(-\sum_{i=0}^{k-1} \hat{C}\hat{A}^{k-i-1}\hat{B} \left[\begin{array}{c} 0 \\ \vdots \\ \frac{\partial}{\partial c_j} e^{-\alpha_j \|u_i - c_j\|_2^2} \\ \vdots \\ 0 \end{array}\right] \right. \\
\left. - \hat{D} \left[\begin{array}{c} 0 \\ \vdots \\ \frac{\partial}{\partial c_j} e^{-\alpha_j \|u_k - c_j\|_2^2} \\ \vdots \\ 0 \end{array}\right] \right). \quad (4.30)
\end{aligned}
$$

Since

$$\frac{\partial}{\partial \alpha_j} e^{-\alpha_j \|u - c_j\|_2^2} = -e^{-\alpha_j \|u - c_j\|_2^2} \|u - c_j\|_2^2,$$

$$\frac{\partial}{\partial c_j} e^{-\alpha_j \|u - c_j\|_2^2} = -e^{-\alpha_j \|u - c_j\|_2^2} \alpha_j \left[2c_j^T - 2u^T \right],$$

the gradients (4.29) and (4.30) can be evaluated.

By computing these gradients, a BFGS quasi-Newton optimization code is used to optimize the basis function parameters. Since the state space matrices and the basis function parameters cannot be estimated simultaneously, basis-function optimization and state space model identification are done alternately. First, an initial set of basis functions is chosen, and then the linear dynamics are identified using a subspace identification algorithm. Once the state space matrices are available, the set of basis functions is optimized. For the optimized set of basis functions, the linear dynamics is identified again, and so on.

4.2. Hammerstein-Wiener Systems

For identifying a Hammerstein-Wiener system, a two step procedure can be employed [29]. In the first step the Hammerstein subsystem is identified. Once a Hammerstein model has been identified, the output nonlinearity can be estimated by solving a standard linear least squares problem. Let $\hat{v} \in \mathbb{R}^p$ be the output of the Hammerstein model. Assuming that the output nonlinearity can be expressed as a basis expansion, a set of r basis functions

$$f(\hat{v}) = \begin{bmatrix} f_1(\hat{v}) \\ \vdots \\ f_r(\hat{v}) \end{bmatrix},$$

can be used, where $f \colon \mathbb{R}^p \to \mathbb{R}^r$ and $f_i \colon \mathbb{R}^p \to \mathbb{R}$. The above basis functions can either be fixed or can be optimized using a gradient-based scheme. Now, a least squares problem of the form

$$y = \mathcal{N}_W(v) \approx \sum_{i=1}^{r} \lambda_i f_i(\hat{v}) = \Lambda f(\hat{v}), \qquad (4.31)$$

can be considered, where y is the measured data and $\Lambda = \begin{bmatrix} \lambda_1 & \cdots & \lambda_r \end{bmatrix}$ is the coefficient matrix. Thus, the least squares solution can be calculated as

$$\Lambda = y f(\hat{v})^{\mathrm{T}} \left(f(\hat{v}) f(\hat{v})^{\mathrm{T}} \right)^{-1} \qquad (4.32)$$

This two-step procedure is not necessarily optimal. However, for the SISO case, with some assumptions on the output nonlinearity, there are results which show that iterating procedures yields convergence to the correct model [1]. Recently a direct approach to identify Hammerstein-Wiener systems using least squares support vector machines (LS-SVM) is being considered in [10].

4.3. Wiener Systems

While the literature on non-subspace Wiener-system identification is extensive [3, 12, 13, 16, 17, 23, 27, 47], these papers generally rely on the assumption that the output nonlinearity is one-to-one. Under this assumption, the identification problem can be treated as a Hammerstein system whose inputs and outputs are, respectively, the outputs and inputs of the original system. Several of the subspace-based Hammerstein identification algorithms have been extended to the Wiener case when the output nonlinearity is one-to-one and invertible. There have also been attempts at subspace-based Wiener system identification without any invertibility assumption on the output nonlinearity [46, 49]. Both these methods are direct (does not involve iterative procedures), however, [46] assumes that the inputs u_k is a Gaussian random sequence and the nonlinearity has a parallel structure. [49] extends the work of [46] to the closed-loop case.

4.4. Linear Parameter-Varying Systems

Identification of linear parameter-varying systems is dealt with in [36, 39]. The model structure is of the form

$$\begin{aligned} x_{k+1} &= A_0 x_k + \begin{bmatrix} A_1 & A_2 & \cdots & A_s \end{bmatrix} (p_k \otimes x_k) \\ &\quad + B_0 u_k + \begin{bmatrix} B_1 & B_2 & \cdots & B_s \end{bmatrix} (p_k \otimes u_k) + K v_k \end{aligned}$$
$$(4.33)$$
$$y_k = C x_k + D u_k + v_k \qquad (4.34)$$

where $p_k \in \mathbb{R}^s$ contains the additional varying parameters that are measured and $\otimes$ denotes the Kronecker product. In this case the Hankel matrices are more complex and include the Kronecker product of the varying parameters p_k and the inputs u_k and the Kronecker product of p_k and the outputs y_k. The states are obtained using a similar procedure, but the proof that the state sequence can be obtained directly from these matrices, is more involved in this case as can be expected. One thing to note here that the definition of persistent excitation now extends to p as well. In other words, the persistency of excitation condition involves combinations of the inputs u and p, instead of the inputs in the case of the linear time-invariant case.

The basic subspace identification method for LPV systems is cumbersome in practice. The number of rows in the Hankel matrices grow exponentially with the order of the system, hence the amount of memory required to compute the QR factorization becomes high. However [37, 39] present techniques for modifying the algorithm to make it numerically more efficient.

4.5. Bilinear Systems

Bilinear models are a special case of LPV systems in which the varying parameters are the inputs u_k itself. Although this case includes an additional term containing the Kronecker product of u_k with itself, the coefficient of this term can be set to zero. Subspace methods for bilinear systems were developed in [7, 8, 36, 38, 40] and

later generalized to LPV systems [39]. In [36, 40], MIMO bilinear systems of the form

$$x_{k+1} = Ax_k + F(u_k \otimes x_k) + Bu_k + w_k, \quad (4.35)$$
$$y_k = Cx_k + Du_k + v_k, \quad (4.36)$$

are considered. The matrix F is first partitioned as $F = \begin{bmatrix} F_1 & F_2 & \cdots & F_m \end{bmatrix}$, where $F_i \in \mathbb{R}^{n \times n}$. Then each F_i is written as

$$F_i = F_i^{(1)}C + F_i^{(2)}\tilde{C} \quad (4.37)$$

with $\tilde{C} \in \mathbb{R}^{(n-l) \times n}$ such that $\begin{bmatrix} C^{\mathrm{T}} & \tilde{C}^{\mathrm{T}} \end{bmatrix}$. Using the above decomposition (4.35) is written as

$$x_{k+1} = Ax_k + \begin{bmatrix} B & F^{(1)} & -F_D^{(1)} \end{bmatrix} \begin{bmatrix} u_k \\ u_k \otimes y_k \\ u_k \times u_k \end{bmatrix}$$
$$+ F_{\tilde{C}}^{(2)}(u_k \otimes x_k) - F^{(1)}(u_k \otimes v_k) + w_k (4.38)$$

where

$$F^{(1)} = \begin{bmatrix} F_1^{(1)} & F_2^{(1)} & \cdots & F_m^{(1)} \end{bmatrix}$$
$$F_D^{(1)} = \begin{bmatrix} F_1^{(1)}D & F_2^{(1)}D & \cdots & F_m^{(1)}D \end{bmatrix}$$
$$F_{\tilde{C}}^{(2)} = \begin{bmatrix} F_1^{(2)}\tilde{C} & F_2^{(2)}\tilde{C} & \cdots & F_m^{(2)}\tilde{C} \end{bmatrix}$$
$$(4.39)$$

The term due to $(u_k \otimes x_k)$ is then neglected and the error due it shown to have zero mean and for the special case $F_{\tilde{C}}^{(2)} = 0$, the error is zero.

Now, using the augmented input vector $\begin{bmatrix} u_k \\ u_k \otimes y_k \\ u_k \otimes u_k \end{bmatrix}$, the identification algorithm consists of the following three steps:

1) Estimating the system order and the matrices A and C. For this step, the augmented input vector is used and past input and output values are used as instrument variables to eliminate some remaining nonlinear terms and estimate the extended observability matrix using a QR factorization and a singular value decomposition similar to section 2.3. From this matrix, consistent estimates of A and C are obtained.
2) Estimating matrices B and D. A least squares problem which is linear in B and D can then be formulated and solved to obtain consistent estimates of B and D.
3) Estimating the matrix F. This is done by solving a nonlinear least squares problem. A good initial estimate is obtained from the rewritten system in step one.

For this algorithm, the set of assumptions involve a few additional assumptions. The assumptions are

- The signals u_k, v_k and w_K are zero mean Gaussian white noise sequences.
- The sequence u_k is statically independent of the sequences v_k and w_k.

- The eigenvalues of the matrix A and the matrix $A \otimes A + \sum_{i=1}^{m} E[[u_k]_i^2]F_i \otimes F_i$ have magnitudes less than one. This is a sufficient condition for the state to be wide sense stationary.
- The pair A, C is observable for the linear system A, B, C, D constructed from the bilinear system. This implies that the bilinear system is also observable.
- The matrix C has full row rank.

5. CLOSED-LOOP SYSTEMS

In practice it is often necessary to perform identification experiments on systems operating in closed loop. This is especially true when open loop experiments are not possible due to an unstable plant, undesirable open-loop behavior, or bad open-loop performance. These considerations are important for an expensive industrial process or machinery. In [22], open loop algorithms are shown to yield biased estimates. This is due to the fact that the noise terms v_k and w_k are generally correlated in the closed-loop case. System identification from closed loop data is thus a relevant topic.

There have been several attempts to apply open loop subspace methods as it is to closed-loop systems. Some algorithms such as the CVA [19] is claimed to work well even in a closed-loop environment.

Explicit closed-loop identification algorithms have also been developed [22, 32, 41, 48]. In [32] a closed-loop framework shown in Figure 6 is considered. A limited number of Markov parameters of the controller $\mathcal{C}$ are assumed known and the measurements of u_k and y_k are assumed available. In [48] an extension of [32] is developed. The scenario considered here is shown in Figure 7. Although the controller $\mathcal{C}$ is assumed to be unknown, measurements of signals r_k, u_k and y_k are assumed to be available. In all of the above methods, the plant is a linear time-invariant system. However, [49] considers closed-loop identification of a Wiener system.
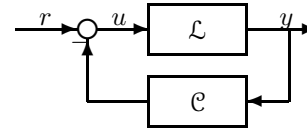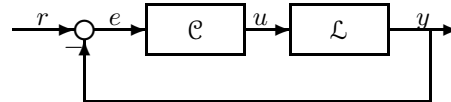

Fig. 6. Closed-Loop Framework I

Fig. 7. Closed-Loop Framework II

6. EXAMPLES

6.1. Example 1: Linear Model of Dryer System Data with Pulsed-Width-Modulated Input

This example is based on 850 time steps of SISO data for a dryer system with pulse-width-modulated input. The data

set was obtained from the web site [6]. For this example linear subspace identification using the `N4SID` command in Matlab was used. Figure 8 shows the input data. The data fit is shown in Figure 9.

6.2. Example 2: Periodic Linear System

In this example taken from [14], a system is constructed as described by (3.1) and (3.2) with period $p = 3$ and with system matrices and with system matrices

$$\left[\begin{array}{c|c} A_1 & B_1 \\ \hline C_1 & D_1 \end{array}\right] = \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 2 & 1 \\ \hline 1 & 0 & 0 \end{array}\right] \tag{6.1}$$

$$\left[\begin{array}{c|c} A_2 & B_2 \\ \hline C_2 & D_2 \end{array}\right] = \left[\begin{array}{cc|c} \frac{1}{5} & 1 & 0 \\ 0 & \frac{2}{5} & 1 \\ \hline 2 & 0 & 0 \end{array}\right] \tag{6.2}$$

$$\left[\begin{array}{c|c} A_3 & B_3 \\ \hline C_3 & D_3 \end{array}\right] = \left[\begin{array}{cc|c} 3 & 1 & 1 \\ 0 & 1 & 2 \\ \hline 1 & 1 & 0 \end{array}\right] \tag{6.3}$$

The inputs and noise were chosen to be similar to the ones in the original paper. A plot of the the output y and the output of the identified model $\hat{y}$ is shown in Figure 10. The dynamics of any regular periodic linear system may be adequately described bye the eigenvalues of its periodic map $\Pi = A_p \cdots A_2 A_1$. The eigenvalues of the period map of the above system are $\Lambda(\Pi) = \{\frac{4}{5}, \frac{3}{5}\}$ The eigenvalues of the period map $\hat{\Pi} = \hat{A}_3 \hat{A}_2 \hat{A}_1$ of the identified system are $\Lambda(\hat{\Pi}) = \{0.8004, 0.5994\}$.

6.3. Example 3: Linear Parameter-Varying System

Here we consider a system of the form (4.33) and (4.34) with $s = 1$ and with the following values of the state-space matrices

$$\left[\begin{array}{c|c} A_0 & B_0 \\ \hline A_1 & B_1 \\ \hline C & D \end{array}\right] = \left[\begin{array}{cc|c} 0.1 & -0.4 & 0 \\ 0.7 & 0.2 & 0.22 \\ \hline 0.1 & 0 & 2 \\ 0 & -0.35 & 0.1 \\ \hline 1 & -2 & 0 \end{array}\right]. \tag{6.4}$$

The signals p_k, u_k and v_k were all chosen to be qhite noise sequences. The eigenvalues of A_0 are $(0.15 \pm 0.5268i)$, while the identified eigenvalues are $(0.1500 \pm 0.5271i)$.

6.4. Example 4: Hammerstein System with Scalar Input

For this example we consider a Hammerstein system whose linear dynamics are given by the discrete-time simple harmonic oscillator

$$A = \left[\begin{array}{cc} 0 & 1 \\ -(\omega^2 T_s^2 + 1) & 2 \end{array}\right], \quad B = \left[\begin{array}{c} 0 \\ T_s^2 \end{array}\right],$$

$$C = \left[\begin{array}{cc} 1 & 0 \end{array}\right], \quad D = 0,$$

where $\omega = 0.7$ and $T_s = 0.1$, with input nonlinearity $\mathcal{N}(u) = u^2$. A total of 1000 data points are used for the identification, and both algorithms were used.

For the selective refinement algorithm, we choose 11 radial basis functions to initialize the algorithm and include 10 random radial basis functions at each subsequent iteration. A total of 2000 iterations are performed, of which 12 are accepted as determined by the data fit decrement. The dominant nonlinearity, involving 131 radial basis functions, is shown in Figure 11.

The basis functions optimization algorithm is employed with sines and cosines and 15 of each are used. The linear system order is specified as 2, and a single dominant nonlinearity is retained at each iteration. The data fit is shown in Figure 12 and the dominant identified nonlinearity is shown in Figure 13.

6.5. Example 5: Hammerstein System with Scalar Input and Rank-2 Nonlinearity

This example is a Hammerstein system based on the discrete-time simple harmonic oscillator of Example 4 cascaded with the low pass filter $\frac{1+z^{-1}}{230-228z^{-1}}$. The scalar input is taken to be a white noise signal with input nonlinearities

$$f(u) = \left[\begin{array}{cc} u^3 & e^{-u} \end{array}\right].$$

The 3rd-order system has the realization

$$A = \left[\begin{array}{ccc} 0.9913 & 1 & 0 \\ 0 & 0 & 0.9901 \\ 0 & -0.995 & 1.9802 \end{array}\right], \quad B = \left[\begin{array}{cc} 0 & 0 \\ 0 & 1 \\ 0.01 & 0 \end{array}\right],$$

$$C = \left[\begin{array}{ccc} 0.0087 & 0.0043 & 0 \end{array}\right], \quad D = \left[\begin{array}{cc} 0 & 0 \end{array}\right].$$

A total of 1000 data points are used for the identification, and the basis function optimization algorithm is implemented with radial basis functions. 15 radial basis functions are used to initialize the algorithm and a total of 10 iterations are performed. The subspace identification algorithm identified a 3rd-order system.

The data fit is shown in Figure 14, and the corresponding dominant nonlinearities involving 15 radial basis functions are shown Figure 15 and Figure 16. These nonlinearities provide estimates of the input nonlinearities u^3 and e^{-u}, respectively.

6.6. Example 6: Nonlinear Feedback System with Scalar Input

For this system we consider the nonlinear feedback system

$$y(k+1) = \tfrac{1}{2}\mathrm{sat}(y_k) + u_k,$$

with u_k chosen to be a white noise signal for $k = 1, \ldots, 1000$. Using 11 radial basis functions to initialize the identification and with 5 additional radial basis functions introduced at each iteration, 100 selective refinement iterations are performed of which 5 are accepted based on fit error decrement. The final estimated nonlinearity in Figure 17 is thus a linear combination of 36 radial basis functions.

6.7. Example 7: Hammerstein Modeling of Cylinder Wake System

The data set for this example is associated with the Cylinder Wake Benchmark Experiment developed at the U.S. Air Force Academy [4]. The system input for the experiment is the commanded vertical displacement of the cylinder, while the output is the measured velocities in the x and y directions at a specified downstream location. Data are available for a total of 600 time steps.

The basis function optimization algorithm is used with a set of 11 radial basis functions, and a total of 15 iterations are performed. The subspace algorithm identifies a 5th-order linear system. The data fit for the y velocity components is shown in Figure 18.

6.8. Example 8: Hammerstein-Wiener Model of Space Weather System

This example is based on data used for space weather prediction. The input data set was measured by the NASA Advanced Composition Explorer (ACE) spacecraft and includes the three components of the magnetic field vector, the solar wind speed, solar wind proton density, and temperature. The system output is ground-based magnetometer data from Thule, Greenland.

The Hammerstein-Wiener identification algorithm described [29] is used for this example. The linear inputs to the model are density, B_y, B_z, and B_t, whereas the inputs entering through the Hammerstein nonlinearly are $B_t V_x \sin^4(\theta/2)$ and the sinusoid with a period of one day. Here B_y and B_z are the y and z components of magnetic field measured by ACE satellite respectively, $B_t = \sqrt{B_y^2 + B_z^2}$, V_x is the x-component of the solar wind speed and $\theta \triangleq a \cos B_z / B_t$. Figure 19 shows the output of the identified Hammerstein-Wiener model. Here only data to the left of the black vertical line is used for identification.

7. ACKNOWLEDGMENTS

REFERENCES

[1] E.-W. Bai. An Optimal Two-Stage Identification Algorithm for Hammerstein-Wiener Nonlinear Systems. *Automatica*, 34(3):333–338, 1998.

[2] D. S. Bernstein. *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory*. Princeton University Press, 2005.

[3] C. H. Chen and S. D. Fassois. Maximum Likelihood Identification of Stochastic Wiener-Hammerstein-Type Non-Linear Systems. *Mech. Sys. Sig. Proc.*, 6(2):135–153, 1992.

[4] K. Cohen, S. Siegel, and T. McLaughlin. Orthogonal decomposition modeling of a controlled ginzburg-landau cylinder wake model. In *41st AIAA Aerospace Sciences Meeting*, Reno, NV, January 2003. AIAA Paper 2003-1292.

[5] K. De Cock and B. De Moor. Subspace identification. in Contribution to section 5.5, Control systems robotics and automation of EOLSS, UNESCO Encyclopedia of Life Support Systems, (Unbehauen H.D.), vol. 1 of 3, Eolss Publishers Co., Ltd. (Oxford, UK http://www.eolss.net/), pp. 933-979, 2003. ftp://ftp.esat.kuleuven.ac.be/pub/SISTA/goethals/eolss_6_43_9_3.ps.

[6] B. De Moor et al. DAISY: Database for the Identification of Systems. http://www.esat.kuleuven.ac.be/sista/daisy/.

[7] W. Favoreel. *Subspace Methods for Identification and Control of Linear and Bilinear Systems*. PhD thesis, Leuven Catholic University, Leuven, Belgium, Nov 1999.

[8] W. Favoreel, B De Moor, and P Van Overschee. Subspace identification of bilinear systems subject to white noise. 44:1157–1165, June 1999.

[9] W. Favoreel, B. De Moor, and P. Van Overschee. Subspace state space system identification for industrial processes. *Jour. of Process Control*, 10:149–155, 2000.

[10] I. Goethals. PhD thesis, Leuven Catholic University, Leuven, Belgium, in preparation.

[11] I. Goethals, K. Pelckmans, J. A. K. Suykens, and B. De Moor. Identification of mimo hammerstein model using least squares support vector machines. accepted for publication in Automatica. available at ftp://ftp.esat.kuleuven.ac.be/pub/SISTA/goethals/goethals_n4sid_hammerstein.pdf.

[12] W. Greblicki. Nonparametric Approach to Wiener System Identification. *IEEE Trans. Circ. Sys.*, 44(6):538–545, 1997.

[13] Z. Hasiewicz. Identification of a Linear System Observed Through Zero-Memory Non-Linearity. *Int. J. Sys. Sci.*, 18(9):1595–1607, 1987.

[14] J. J. Hench. A technique for the idnetification of linear periodic state-space models. *Int. J. Cont.*, 62(2):289–301, 1995.

[15] S. L. Lacy and D. S. Bernstein. Subspace Identification for Nonlinear Systems That Are Linear in Unmeasured States. In *Proc. Conf. Dec. Contr.*, pages 3518–3523, Orlando, Florida, December 2001.

[16] S. L. Lacy, R. S. Erwin, and D. S. Bernstein. Identification of Wiener Systems with Known Noninvertible Nonlinearities. *J. Dyn. Sys. Meas. Contr.*, 123:566–571, December 2001.

[17] S. L. Lacy, R. S. Erwin, and D. S. Bernstein. Identification of Wiener systems with known noninvertible nonlinearities. In *Proc. Amer. Contr. Conf.*, pages 4890–4895, Arlington VA, June 2001.

[18] W. E. Larimore. Canonical Variate Analysis in Identification, Filtering and Adaptive Control. In *Proc. 29th Conf. Dec. Contr.*, pages 596–604, Honolulu, HI, December 1990.

[19] W. E. Larimore. Optimal Order Selection and Efficiency of Canonical Variate Analysis System Identification. In *Proc. of the World Congress of the International Federation of Automatic Control, IFAC*, volume I, pages 151–156, San Francisco, USA, 1996.

[20] Wallace E. Larimore. *ADAPTx Users Manual*. Adaptics, Inc, McLean VA.

[21] L. Ljung. *System Identification: Theory for the User*. Prentice Hall Information and System Sciences Series. Prentice Hall, 2nd edition, January 1999.

[22] L. Ljung and T. Mckelvey. Subspace Identification from Closed Loop Data. *Signal Processing*, 52:209–215, 1996.

[23] M. Lovera, T. Gustafsson, and M. Verhaegen. Recursive Subspace Identification of Linear and Non-Linear Wiener State-Space Models. *Automatica*, 36:1639–1650, 2000.

[24] MATLAB. *System Identification Toolbox*. Mathworks, Natick, MA.

[25] R. A. Meyer and C. S. Burrus. A unified analysis of multirate and periodically time-varying digital filters. *IEEE Trans. on Ciruits and Systems*, 22:162–168, 1975.

[26] M. Moonen, B. De Moor, L. Vandenberghe, and J. Vandewalle. On-and Off-Line Identification of Linear State-Space Models. *Int. J. Contr.*, 49(1):219–232, 1989.

[27] G. A. Pajunen. Recursive Identification of Wiener Type Nonlinear Systems. In *Proc. Amer. Contr. Conf.*, pages 1365–1370, Boston, MA, June 1985.

[28] H. Palanthandalam-Madapusi, J.B. Hoagg, and D. S. Bernstein. Basis-function optimization for subspace-based nonlinear identification of systems with measured-input nonlinearities. In *Proc. Amer. Contr. Conf.*, pages 4788–4793, Boston, MA, July 2004.

[29] H. Palanthandalam-Madapusi, A. J. Ridley, and D. S. Bernstein. Identification and prediction of ionospheric dynamics using a hammerstein-wiener model. Proc. Amer. Contr. Conf., Portland, OR, June 2005.

[30] K. M. Pekpe, K. Gasso, G. Mourot, and J. Ragot. Subspace identification of switching model. In *13th IFAC Symposium on System Identification*, pages 1131–1136.

[31] C. D. Schaper, W. E. Larimore, D. E. Sborg, and D. A. Mellichamp. Identification of chemical processes using canonical variate analysis. *Computers and Chemical Engineering*, 18(1):55–69, 1994.

[32] P. Van Overchee and B. De Moor. Closed loop subspace system identification. In *Proc. Conf. Dec. Contr.*, pages 1848–1853, San Diego, California USA, December 1997.

[33] P. Van Overschee and B. De Moor. N4SID: Subspace Algorithms for the Identification of Combined Deterministic-Stochastic Systems. *Automatica*, 30(1):75–93, January 1994.

[34] P. Van Overschee and B. De Moor. *Subspace Identification for Linear Systems: Theory, Implementation, Applications*. Kluwer, 1996.

[35] T. Van Pelt and D. S. Bernstein. Nonlinear System Identification Using Hammerstein and Nonlinear Feedback Models with Piecewise Linear Static Maps. *Int. J. Contr.*, 74:1807–1823, 2001.

[36] V. Verdult. *Nonlinear System Identification: A State-Space Approach*. PhD thesis, University of Twente, Faculty of Applied Physics, Enschede, The Netherlands, 2002.

[37] V. Verdult and M. Verhaegen. Kernel methods for subspace identification of multivariable lpv and bilinear systems. preprint.

[38] V. Verdult and M. Verhaegen. Subspace identification of mimo bilinear systems. In *Proc. of the European Contr. Conf.*, Karlsruhe, Germany, September 1999.

[39] V. Verdult and M. Verhaegen. Identification of multivariable linear parameter-varying systems based on subspace techniques. In *Proc. Conf. Dec. Contr.*, pages 1567–1572, Sydney, Australia, December 2000.

[40] V. Verdult, M. Verhaegen, C. T. Chou, and M. Lovera. Efficient and systematic identification of mimo bilinear state space models. In *Proc. Conf. Dec. Contr.*, pages 1260–1265, Tampa, Florida USA, December 1998.

[41] M. Verhaegen. Application of a subspace model identification algorithms technique to lti systems operating in closed-loop. *Automatica*, 29(4):1027–1040, 1993.

[42] M. Verhaegen. Identification of the deterministic part of MIMO state space models given in innovations form from inputoutput data. *Automatica*, 30(1):61–74, 1994.

[43] M. Verhaegen and D. Westwick. Identifying MIMO Hammerstein Systems in the Context of Subspace Model Identification Methods. *Int. J. Contr.*, 63(2):331–349, 1996.

[44] M. Verhaegen and X. Yu. A class of subspace model identification algorithms to identify periodically and arbitrarily time-varying systems. *Automatica*, 31(2):201–216, 1995.

[45] M. Vilberg. Subspace methods in system identification. In *Proc. of SYSID*, volume 1, pages 1–12, Copenhagen, Denmark, 4-6 July 1994.

[46] D. Westwick and M. Verhaegen. Identifying MIMO Wiener Systems Using Subspace Model Identification Methods. *Signal Processing*, 52:235–258, 1996.

[47] T. Wigren. Convergence Analysis of Recursive Identification Algorithms Based on the Nonlinear Wiener Model. *IEEE Trans. Autom. Contr.*, 39(11):2191–2206, 1994.

[48] Y. Zhao and D. Westwick. Closed-loop system identification using subspace-based methods. In *IEEE CCECE 2003, Canadian Conf.*, volume 3, pages 1727–1730, 4-7 May 2003.

[49] Y. Zhao and D. Westwick. A direct approach to closed-loop identification of wiener models using subspace methods. In *Proc. Amer. Contr. Conf.*, volume 4, pages 3585–3589, 4-6 June 2003.
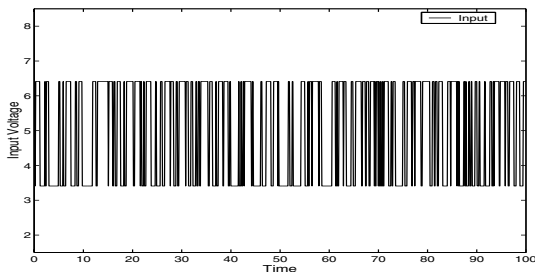
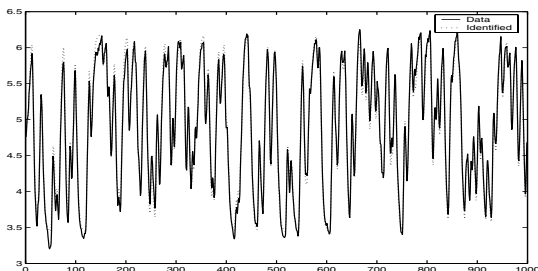Fig. 8. Example 1: Pulse-width-modulated input data for the dryer example.

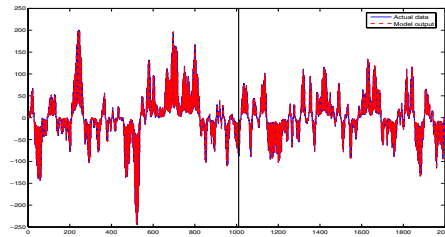

Fig. 9. Example 1: Data fit for the dryer system.

Fig. 10. Example 2: True output y and the output of the identified model $\hat{y}$. Data to the left of the vertical line were used for identification and the right of the vertical line is prediction using the identified model.
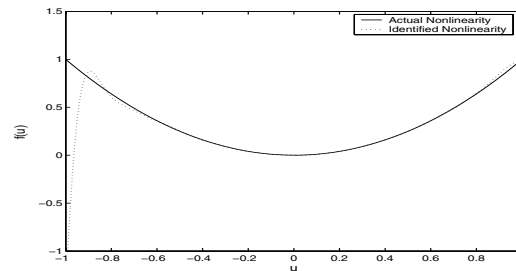

Fig. 11. Example 3: True (u^2) and identified input nonlinearities for second-order Hammerstein system with scalar input and rank 1 nonlinearity using selectively refined radial basis functions.
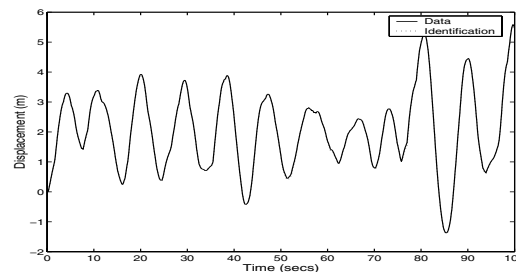

Fig. 12. Example 3: Data fit for second-order Hammerstein system with scalar input and rank 1 nonlinearity using optimized sines and cosines.
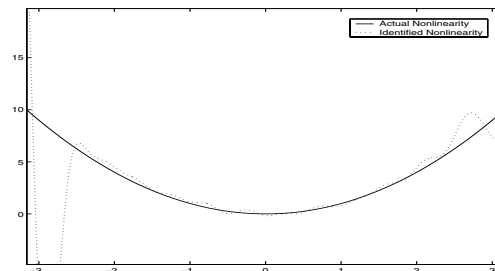

Fig. 13. Example 3: True (u^2) and identified input nonlinearity for second-order Hammerstein system with scalar input and rank 1 nonlinearity using optimized sines and cosines.
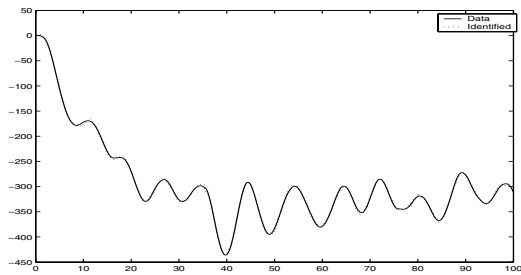
Fig. 14. Example 4: Data fit for third-order Hammerstein system with scalar input and rank 2 nonlinearity using optimized radial basis functions.
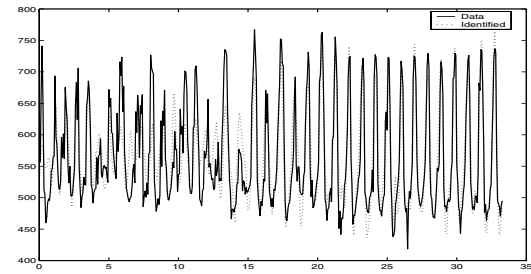

Fig. 18. Example 7: y-direction data fit for Cylinder Wake Benchmark Experiment using 5th-order model and a rank 2 input nonlinearity.
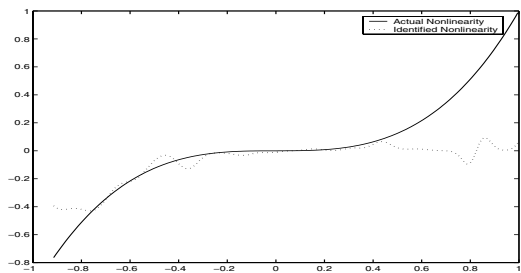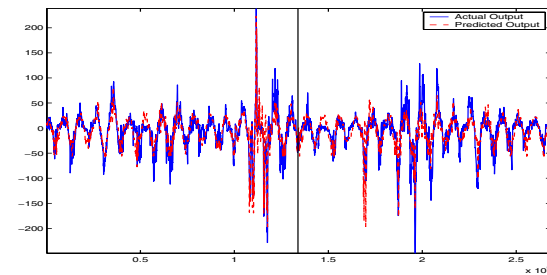

Fig. 15. Example 4: True (u^3) and identified input nonlinearities for third-order Hammerstein system with scalar input and rank 2 nonlinearity using optimized radial basis functions.

Fig. 16. Example 4: True (e^{-u}) and identified input nonlinearities for third-order Hammerstein system with scalar input and rank 2 nonlinearity using optimized radial basis functions.

Fig. 19. Example 8: Measured and predicted data of the Thule magnetometer using a Hammerstein-Wiener model identified in the second step. Data to the left of the vertical line are used for identification, while right part is prediction using the identified model.
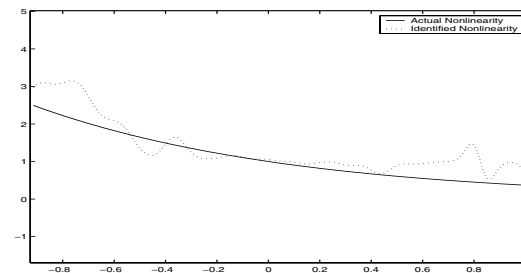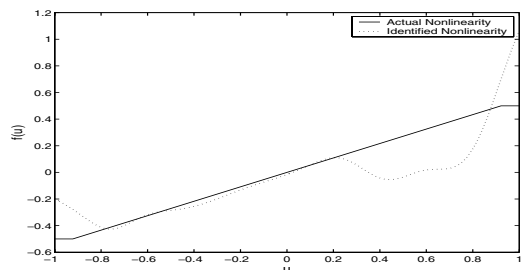

Fig. 17. Example 5: Identified feedback nonlinearity for first-order nonlinear feedback system using selective refinement with radial basis functions.
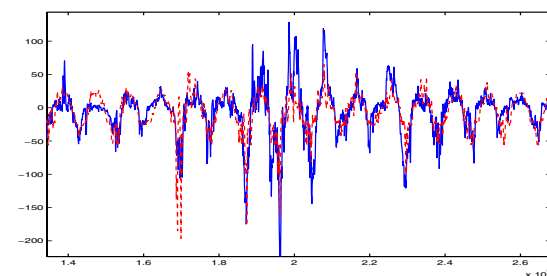

Fig. 20. Magnified view of the prediction region in Figure 19