ELSEVIER

Journal of Process Control 10 (2000) 149–155

JOURNAL OF PROCESS CONTROL

www.elsevier.com/locate/jprocont

Subspace state space system identification for industrial processes

Wouter Favoreel*, Bart De Moor, Peter Van Overschee

Katholieke Universiteit Leuven, Department of Electrical Engineering-ESAT/SISTA, Kardinaal Mercierlaan 94, 3001 Leuven, Belgium

Abstract

We give a general overview of the state-of-the-art in subspace system identification methods. We have restricted ourselves to the most important ideas and developments since the methods appeared in the late eighties. First, the basics of linear subspace identification are summarized. Different algorithms one finds in literature (such as N4SID, IV-4SID, MOESP, CVA) are discussed and put into a unifying framework. Further, a comparison between subspace identification and prediction error methods is made on the basis of computational complexity and precision of the methods by applying them on 10 industrial data sets.

Keywords: Subspace methods; System identification; State-space models; Multivariable systems; Linear algebra

1. Introduction

The beginning of the 1990s witnessed the birth of a new type of linear system identification algorithms, called *subspace methods*. Subspace methods originate in a happy *menage-à-trois* between system theory, geometry and numerical linear algebra. Previous papers and books emphasizing different aspects of subspace system identification and signal processing and in which one can find large sets of references to the literature are [1–4]. We should also mention some special issues of the journals IFAC Automatica (''Special Issue on Statistical Signal Processing and Control'', Jan. 94; ''Special Issue on System Identification'', Dec 95;) and Signal Processing (''Special Issue on Subspace Methods for System Identification'', July 1996), which contain contributions on subspace identification, as well as the Proceedings of the 11th IFAC Symposium on System Identification (Kitakyushu, Japan, July 1997).

Linear subspace identification methods are concerned with systems and models of the form[1]

$$x_{k+1} = Ax_k + Bu_k + w_k, \tag{1}$$

$$y_k = Cx_k + Du_k + v_k, \tag{2}$$

with

$$\mathbf{E}\left[\begin{pmatrix} w_p \\ v_p \end{pmatrix}\begin{pmatrix} {w_q}^T & {v_q}^T \end{pmatrix}\right] = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix}\delta_{pq} \geqslant 0. \tag{3}$$

The vectors $u_k \in \mathbb{R}^{m\times 1}$ and $y_k \in \mathbb{R}^{l\times 1}$ are the measurements at time instant k of, respectively, the m inputs and l outputs of the process. The vector x_k is the state vector of the process at discrete time instant k, $v_k \in \mathbb{R}^{l\times 1}$ and $w_k \in \mathbb{R}^{n\times 1}$ are unobserved vector signals, v_k is called the measurement noise and w_k is called the process noise. It is assumed that they are zero mean, stationary white noise vector sequences and uncorrelated with the inputs u_k. $A \in \mathbb{R}^{n\times n}$ is the system matrix, $B \in \mathbb{R}^{n\times m}$ is the input matrix, $C \in \mathbb{R}^{l\times n}$ is the output matrix while $D \in \mathbb{R}^{l\times m}$ is the direct feed-through matrix. The matrices $Q \in \mathbb{R}^{n\times n}$, $S \in \mathbb{R}^{n\times l}$ and $R \in \mathbb{R}^{l\times l}$ are the covariance matrices of the noise sequences w_k and v_k.

In subspace identification it is typically assumed that the number of available data points goes to infinity, and that the data is ergodic. We are now ready to state the main problem treated:

Given a large number of measurements of the input u_k and the output y_k generated by the unknown system (1)–(3). **Determine** the order n of the unknown system, the system matrices A, B, C, D up to within a similarity transformation and an estimate of the matrices Q, S, R.

In this paper we briefly recapitulate the main concepts and algorithms of linear subspace system identification (Section 2). Different methods of the literature

* Corresponding author. Tel.: +32-16-321-907; fax: +32-16-321-970; http://www.esat.kuleuven.ac.be/sista

E-mail address: wouter.favoreel@esat.kuleuven.ac.be (W. Favoreel).

E-mail addresses: bart.demoor@esat.kuleuven.ac.be (B. De Moor); peter.vanoverschee@esat.kuleuven.ac.be (P. Van Overschee).

[1] **E** denotes the expected value operator and δ_{pq} the Kronecker delta.

0959-1524/00/$ - see front matter
PII: S0959-1524(99)00030-X

are presented and put into a unifying framework. We comment on the comparison between *prediction error methods* (PEM) and *subspace identification methods*. It should be emphasized that these two identification approaches are by no means competing. Instead, they are "*... a most useful complement to traditional maximum-likelihood based methods*", as emphasized in [5].

2. An overview of the theory

In this section we describe the general concepts in subspace identification. Further, the two basic steps all subspace methods consist of are described. Finally, the different algorithms existing in the literature are analyzed in a unifying framework.

Subspace identification algorithms always consist of two steps. The first step makes a projection of certain subspaces generated from the data, to find an estimate of the extended observability matrix and/or an estimate of the states of the unknown system. The second step then retrieves the system matrices from either this extended observability matrix or the estimated states. We will come back to these two steps in Section 2.2.2, where we describe different subspace identification methods and fit them into a unifying framework.

2.1. The subspace structure of linear systems

The following input–output matrix equation [6], played a very important role in the development of subspace identification:

$$Y_f = \Gamma_i X_i + H_i^d U_f + H_i^s M_f + N_f. \tag{4}$$

The different terms in this equation are:

- The extended observability matrix Γ_i:

$$\Gamma_i \stackrel{\text{def}}{=} \begin{pmatrix} C \\ CA \\ CA^2 \\ \dots \\ CA^{i-1} \end{pmatrix} \tag{5}$$

- The deterministic lower block triangular Toeplitz matrix H_i^d:

$$H_i^d \stackrel{\text{def}}{=} \begin{pmatrix} D & 0 & 0 & \dots & 0 \\ CB & D & 0 & \dots & 0 \\ CAB & CB & D & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ CA^{i-2}B & CA^{i-3}B & CA^{i-4}B & \dots & D \end{pmatrix} \tag{6}$$

- The stochastic lower block triangular Toeplitz matrix H_i^s:

$$H_i^s \stackrel{\text{def}}{=} \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ C & 0 & 0 & \dots & 0 \\ CA & C & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ CA^{i-2} & CA^{i-3} & CA^{i-4} & \dots & 0 \end{pmatrix} \tag{7}$$

- The input and output block Hankel matrices are defined as:

$$U_{0|i-1} \stackrel{\text{def}}{=} \begin{pmatrix} u_0 & u_1 & \dots & u_{j-1} \\ u_1 & u_2 & \dots & u_j \\ \dots & \dots & \dots & \dots \\ u_{i-1} & u_i & \dots & u_{i+j-2} \end{pmatrix}, \tag{8}$$

$$Y_{0|i-1} \stackrel{\text{def}}{=} \begin{pmatrix} y_0 & y_1 & \dots & y_{j-1} \\ y_1 & y_2 & \dots & y_j \\ \dots & \dots & \dots & \dots \\ y_{i-1} & y_i & \dots & y_{i+j-2} \end{pmatrix}, \tag{9}$$

where we assume for stochastic reasons that $j \to \infty$ throughout the paper. For convenience and short hand notation, we call:

$$U_p \stackrel{\text{def}}{=} U_{0|i-1}, \qquad U_f \stackrel{\text{def}}{=} U_{i|2i-1},$$

$$Y_p \stackrel{\text{def}}{=} Y_{0|i-1}, \qquad Y_f \stackrel{\text{def}}{=} Y_{i|2i-1},$$

where the subscript p and f denote, respectively, the past and the future. The matrix containing the inputs U_p and outputs Y_p will be called W_p:

$$W_p \stackrel{\text{def}}{=} \begin{pmatrix} Y_p \\ U_p \end{pmatrix}.$$

The block Hankel matrix formed with the process noise w_k and the measurement noise v_k are defined, respectively, as $M_{0|i-1}$ and $N_{0|i-1}$ in the same way. Once again, we define for short hand notation:

$$M_p \stackrel{\text{def}}{=} M_{0|i-1}, \qquad M_f \stackrel{\text{def}}{=} M_{i|2i-1},$$

$$N_p \stackrel{\text{def}}{=} N_{0|i-1}, \qquad N_f \stackrel{\text{def}}{=} N_{i|2i-1}.$$

We finally denote the state sequence X_i as:

$$X_i \stackrel{\text{def}}{=} \begin{pmatrix} x_i & x_{i+1} & x_{i+2} & \dots & x_{i+j-1} \end{pmatrix}. \tag{10}$$

In what follows, we will use the matrices $\mathcal{A} \in \mathbb{R}^{p \times j}$ and $\mathcal{B} \in \mathbb{R}^{q \times j}$.

Definition (*Orthogonal projections*)

The orthogonal projection of the row space of $\mathcal{A}$ into the row space of $\mathcal{B}$ is denoted by $\mathcal{A}/\mathcal{B}$ and defined as[2]:

[2] $\bullet^{\dagger}$denotes the Moore–Penrose pseudo-inverse. This generalized inverse could be replaced by less "restricted" generalized inverses, but this will not be pursued here.

$\mathcal{A}/\mathcal{B} = \mathcal{A}\mathcal{B}^{\dagger}\mathcal{B}$.

$\mathcal{A}/\mathcal{B}^{\perp}$ is the projection of the row space of $\mathcal{A}$ into $\mathcal{B}^{\perp}$, the orthogonal complement of the row space of $\mathcal{B}$, for which we have $\mathcal{A}/\mathcal{B}^{\perp} = A - \mathcal{A}/\mathcal{B}$.

2.2. The two basic steps in subspace identification

In this section we will explore the two main steps that all subspace algorithms consist of (see Fig. 1). The first step always performs a weighted projection of the row space of the previously defined data Hankel matrices. From this projection, the observability matrix Γ_i and/or an estimate $\tilde{X}_i$ of the state sequence X_i can be retrieved. In the second step, the system matrices A, B, C, D and Q, S, R are determined. As shown in Fig. 1, a clear distinction can be made between the algorithms that use the extended observability matrix Γ_i to obtain the state space matrices, and those using the estimated state sequence $\tilde{X}_i$.

2.2.1. First step: finding the state sequence and/or the extended observability matrix

In this section, we show how an orthogonal projection with data block Hankel matrices forms one of the key elements in subspace system identification algorithms.

All subspace methods start from the previously presented matrix input–output equation (4). It states that the block Hankel matrix containing the future outputs Y_f is related in a linear way to the future input block Hankel matrix U_f and the future state sequence X_i. The basic idea of subspace identification now is to recover the $\Gamma_i X_i$-term of this equation. This is a particularly interesting term since either the knowledge of Γ_i or X_i leads to the system parameters (see Section 2.2.2. Moreover $\Gamma_i X_i$ is a rank deficient term (of rank n, i.e. the system order!) which means that once $\Gamma_i X_i$ is known, Γ_i, X_i and the order n can be simply found from a SVD.

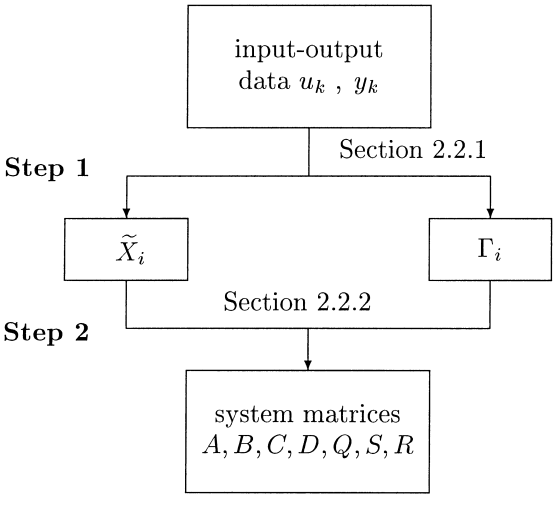

Fig. 1. Subspace algorithms always consist of two main steps. In Section 2.2.1 it is explained how from the first step the extended observability matrix Γ_i and an estimate $\tilde{X}_i$ of the state sequence X_i are determined. In the second step the system matrices A, B, C, D and the noise covariance matrices Q, R, S are calculated using one of the algorithms described in Section 2.2.2.

How can an estimate of $\Gamma_i X_i$ be extracted from the above equation? For this we need the previously defined notion of orthogonal projection. By projecting the row space of Y_f into the orthogonal complement $U_f^{\perp}$ of the row space of U_f we find:

$$Y_f/\boldsymbol{U_f^{\perp}} = \Gamma_i X_i/\boldsymbol{U_f^{\perp}} + H_i^s M_f/\boldsymbol{U_f^{\perp}} + N_f/\boldsymbol{U_f^{\perp}}.$$

Since it is assumed that the noise is uncorrelated with the inputs we have that:

$$M_f/\boldsymbol{U_f^{\perp}} = M_f, \qquad N_f/\boldsymbol{U_f^{\perp}} = N_f.$$

Therefore:

$$Y_f\boldsymbol{U_f^{\perp}} = \Gamma_i X_i/\boldsymbol{U_f^{\perp}} + H_i^s M_f + N_f.$$

The following step consists in weighting this projection to the left and to the right with some matrices W_1 and W_2:

$$W_1.Y_f/\boldsymbol{U_f^{\perp}}.W_2 = \underbrace{W_1.\Gamma_i}_{1.}\underbrace{X_i/\boldsymbol{U_f^{\perp}}.W_2}_{2.} + \underbrace{W_1.\left(H_i s^s M_f + N_f\right).W_2}_{3.}.$$

Of course, the inputs U_f and the weighting matrices W_1 and W_2 can not be chosen arbitrarily but they should satisfy the following three conditions:

1. $\text{rank}\ (W_1.\Gamma_i) = \text{rank}\ \Gamma_i$ (11)

2. $\text{rank}\ \left(X_i/\boldsymbol{U_f^{\perp}}.W_2\right) = \text{rank}\ X_i$ (12)

3. $W_1.\left(H_i^s M_f + N_f\right).W_2 = 0$ (13)

The first two conditions guarantee that the rank-n property of $\Gamma_i X_i$ is preserved after projection onto $U_f^{\perp}$ and weighting by W_1 and W_2. The third condition expresses that W_2 should be uncorrelated with the noise sequences w_k and v_k. If these three conditions are satisfied, we have that:

$$\mathcal{O}_i \overset{\text{def}}{=} W_1.Y_f/\boldsymbol{U_f^{\perp}}.W_2 \tag{14}$$

$$= W_1.\Gamma_i X_i/\boldsymbol{U_f^{\perp}}.W_2$$

with SVD:

$$\mathcal{O}_i = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix}.$$

The following important properties can now be stated:

$$\text{rank} \quad \mathcal{O}_i = n,$$

$$W_1.\Gamma_i \quad = U_1 S_1^{1/2},$$

$$X_i/\boldsymbol{U_f^{\perp}}.W_2 = S^{1/2} V_2^T.$$

Obviously, the singular value decomposition of the matrix $W_1.Y_f/\boldsymbol{U_f^{\perp}}.W_2$ delivers the order n of the system. Moreover, from the left singular vectors corresponding to the non-zero singular values the extended observability matrix Γ_i can be found (up to a similarity transformation) where as the right singular vectors contain information about the states X_i. For an appropriate choice of the weighting matrix W_2, the matrix

$$\tilde{X}_i \overset{\text{def}}{=} X_i/\boldsymbol{U_f}.W_2 \tag{15}$$

can indeed be considered as an estimate of the state sequence X_i. It was shown [4] that, for a particular choice of W_2, $\tilde{X}_i$ is a Kalman filter estimate of X_i. One might wonder about the effect of choosing the weights W_1 and W_2 in (14). Without going into details here, it suffices to say that, by choosing appropriate weighting matrices W_1 and W_2, all subspace algorithms for LTI systems can be interpreted in the above framework, including N4SID [7], MOESP [8], CVA [9], *basic*-4SID and IV-4SID [10]. It should be noted that for the *basic*-4SID algorithm, condition (13) is not satisfied which implies that in general this method is not consistent. We present in Fig. 2 the acronyms of these algorithms and the weights to be plugged into (14). We refer to [4,11,12] for proofs and details.

2.2.2. Second step: finding the state space model

What have we done so far? We have found how an estimate $\tilde{X}_i$ of the state sequence X_i and the extended observability matrix Γ_i can be retrieved from the weighted projection (14) of the future outputs Y_f into the orthogonal complement of the future inputs U_f. At this point, we can clearly distinguish two classes of subspace algorithms. The first class uses the state estimates $\tilde{X}_i$ (the right singular vectors) to find the state space model. Algorithms that follow this approach are N4SID [7] and CVA [9]. The second class of algorithms uses the extended observability matrix Γ_i (i.e. the left singular vectors) to find the model parameters. Examples in the literature are MOESP [8] and IV-4SID, *basic*-4SID [10,13,14].

2.2.2.1. Algorithms using an estimate $\tilde{X}_i$ of the state sequence.

Without going into the details [4,7], we mention that if the weights W_1 and W_2 correspond to those of the N4SID algorithm (See Fig. 2), the estimated state sequence $\tilde{X}_i$ can be interpreted as the solution of a bank of Kalman filters, working in parallel on each of the columns of the matrix W_p. Besides $\tilde{X}_i$, we also need the state sequence $\tilde{X}_{i+1}$. This sequence can be obtained from a projection and *new* weights $\overline{W}_1$, $\overline{W}_2$ in (14) based on $W_{0|i}$, $Y_{i+1|2i-1}$ and $U_{i+1|2i-1}$ (see Section 2.1 for notations). This leads to the sequence $\mathcal{O}_{i+1}$ and the Kalman filter states $\tilde{X}_{i+1}$:

$$\mathcal{O}_{i+1} = \overline{W}_1.Y_{i+1|2i-1}/\boldsymbol{U^{\perp}_{i+1|2i-1}}.\overline{W}_2$$

$$= \overline{W}_1.\Gamma_{i-1}\tilde{X}_{i+1}.\overline{W}_2.$$

System model: The state space matrices A, B, C and D can now be found by solving a simple set of over determined equations in a least squares sense [4,7]:

$$\begin{pmatrix} \tilde{X}_{i+1} \\ Y_{i|i} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \tilde{X}_i \\ U_{i|i} \end{pmatrix} + \begin{pmatrix} \rho_w \\ \rho_v \end{pmatrix}, \tag{16}$$

with obvious definitions for ρ_w and ρ_v as residual matrices. This reduces to

$$\min_{A,B,C,D} \left\| \begin{pmatrix} \tilde{X}_{i+1} \\ Y_{i|i} \end{pmatrix} - \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \tilde{X}_i \\ U_{i|i} \end{pmatrix} \right\|_F^2$$

Noise model: The noise covariances Q, S and R can be estimated from the residuals ρ_w and ρ_v as:

$$\begin{pmatrix} Q & S \\ S^T & R \end{pmatrix}_i = \frac{1}{j}\left[\begin{pmatrix} \rho_w \\ \rho_v \end{pmatrix} \cdot \begin{pmatrix} \rho_w^T \rho_s^T \end{pmatrix} \right] \geqslant 0,$$

where the index i denotes a bias induced for finite i, which disappears as $i \to \infty$ (see further). As is obvious by construction, this matrix is guaranteed to be positive semi-definite. This is an important feature since only positive definite covariances can lead to a physically realizable noise model.

There is an important observation to be made here: Corresponding columns of $\tilde{X}_i$ and of $\tilde{X}_{i+1}$ are Kalman filter state estimates of X_i and X_{i+1}. It would lead us beyond the scope of the present paper to give all the mathematical details, but it should be mentioned that although $\tilde{X}_i$ and $\tilde{X}_{i+1}$ are both Kalman filter estimates of the states of the system, the state sequences needed to initialize these Kalman filters are different. As a consequence, the set of equations (16) is not theoretically consistent which means that the estimates of the system matrices A, B, C, D are slightly biased. One should refer to [4] for a more thorough explanation on this topic, and for more involved algorithms that provide consistent estimates of A, B, C, D and slightly biased estimates of

Q, R, S that are consistent (if $i \to \infty$). These algorithms tackle the origin of the bias (i.e. the difference in initial state for the Kalman filter sequences $\tilde{X}_i$ and $\tilde{X}_{i+1}$) to find an unbiased version of the algorithm presented in this paper. The algorithms in [4] have moreover been optimized with respect to numerical robustness, bias and noise sensitivity. Also Matlab-code is provided for these algorithms. Since the aim of the present paper is only to give an overview of the existing methods, we restricted ourselves here to a simple, but slightly biased version of more sophisticated N4SID algorithms.

2.2.2.2. Algorithms using the extended observability matrix Γ_i. Contrarily to the previous class of algorithms, here the system matrices are determined in two separate steps: first, A and C are determined from Γ_i while in a second step B and D are computed.

Determination of A and C: The matrices A and C can be determined from the extended observability matrix in different ways. All the methods, make use of the shift invariance property of the matrix Γ_i, which implies that [15] (following Matlab notations):

$$A = \underline{\Gamma}_i^{\dagger}.\overline{\Gamma}_i, \qquad C = \Gamma_i(1:l,:).$$

Determination of B and D: After the determination of A and C, the system matrices B and D have to be computed. Here we will only sketch one possible way to do so. From the input–output equation (4), we find that:

$$\underbrace{\Gamma_i^{\perp}.Y_f U_j^{\dagger}}_{\in\mathbb{R}^{(li-n)\times mi}} = \underbrace{\Gamma_i^{\perp}}_{\in\mathbb{R}^{(li-n)\times li}} \cdot \underbrace{H_i^d}_{\in\mathbb{R}^{li\times mi}}$$

where $\Gamma_i^{\perp} \in \mathbb{R}^{(li-n)\times li}$ is a full row rank matrix satisfying $\Gamma_i^{\perp}.\Gamma_i = 0$. Here once again the noise is cancelled out due to the assumption that the input u_k is correlated with the noise. Observe that with known matrices $A, C, \Gamma_i^{\perp}, U_f$ and Y_f this equation is linear in B and D.

Different extensions of the present algorithms to other classes of systems such as bilinear systems, continuous-time systems, periodic systems, systems operating in closed-loop, etc. can be found in [16].

3. Subspace identification vs. prediction error methods for some industrial data sets

In this section it is our purpose to make a direct comparison between prediction error methods (PEM) [17] and the currently discussed subspace identification algorithms. First we will analyze some general differences between these two approaches. Further we will apply both methods to the same data sets obtained from real-life applications.

Besides some conceptual novelties, such as re-emphasizing of the state in the field of system identification (see Section 2), subspace methods are characterized by several advantages with respect to PEMs. One of them is the so-called parameterization problem, which is particularly non-trivial for systems with multiple outputs (see references in [17]). In subspace methods on the contrary, the model is parameterized by the full state space model, and the model order is decided upon in the identification procedure. Further, there is no basic complication for subspace algorithms in going from SISO to MIMO systems. Also, a nonzero initial state poses no additional problems in terms of parameterization, which is not the case with input–output based parameterizations, typically used in PEMs. Finally, stable systems are treated exactly the same way as unstable ones. Another main advantage is that subspace methods, when implemented correctly, have better numerical properties than PEMs. For instance, subspace identification methods do not involve nonlinear optimization techniques which means they are fast (since non-iterative) and accurate (since no problems with local minima occur). The price to be paid is that they are suboptimal. In order to demonstrate this trade-off, we have compared two methods on 10 practical examples. The 10 industrial examples are mechanical, from process industry and thermal ([4] pp. 189–196 and the references therein for more details). It should be noted that all the data sets that are discussed here can be downloaded freely from the internet site DAISY[3].

N4SID "Robustified" version of the N4SID algorithm presented above. For more details we refer to [4].

PEM The prediction error algorithm described in [18], which uses a full parameterization of the state space model combined with regularization. The implementation in Matlab of the algorithm was obtained from McKelvey [19]. As an initial starting value of the model we took the result of the above mentioned N4SID subspace identification algorithm.

What is of interest here, is the trade-off for each of the above methods, between two quantities:

- The computational requirements (measured in the number of floating points operations as indicated by Matlab).
- The prediction error on the validation data set, defined as:

$$\varepsilon = 100\frac{1}{l}\sum_{c=1}^{l}\left[\sqrt{\frac{\sum_{k=1}^{s}\left((y_k)_c-\left(y_k^p\right)_c\right)^2}{\sum_{k=1}^{s}\left((y_k)_c\right)^2}}\right]\%.$$

where y_k^p is the one step ahead predicted output.

We can say that, for nine out of the 10 practical examples, the error for the subspace methods is only 15% larger than for prediction error methods. This illustrates the fact that subspace identification methods

[3] http://www.esat.kuleuven.ac.be/sista/daisy

Acronym	W_1	W_2
N4SID	I_{li}	$(W_p/\boldsymbol{U}_{\boldsymbol{f}}^{\perp})^{\dagger}W_p$
CVA	$[(Y_f/\boldsymbol{U}_{\boldsymbol{f}}^{\perp}).$ $(Y_f/\boldsymbol{U}_{\boldsymbol{f}}^{\perp})^T]^{-1/2}$	$(W_p/\boldsymbol{U}_{\boldsymbol{f}}^{\perp})^{\dagger}.$ $(W_p/\boldsymbol{U}_{\boldsymbol{f}}^{\perp})$
MOESP	I_{li}	$(W_p/\boldsymbol{U}_{\boldsymbol{f}}^{\perp})^{\dagger}.$ $(W_p/\boldsymbol{U}_{\boldsymbol{f}}^{\perp})$
Basic-4SID	I_{li}	I_j
IV-4SID	I_{li}	Φ

Fig. 2. This table interprets different existing subspace identification algorithms in a unifying framework. All these algorithms first calculate a weighted projection (14) followed by a SVD. The left and right weighting matrices are W_1 and W_2, respectively. The first two algorithms use the state estimates $\tilde{X}_i$ (the right singular vectors) to find the system matrices while the last three algorithms are based on the extended observability matrix Γ_i (the left singular vectors). The matrix Φ of the IV-4SID method is a matrix containing the instrumental variables.

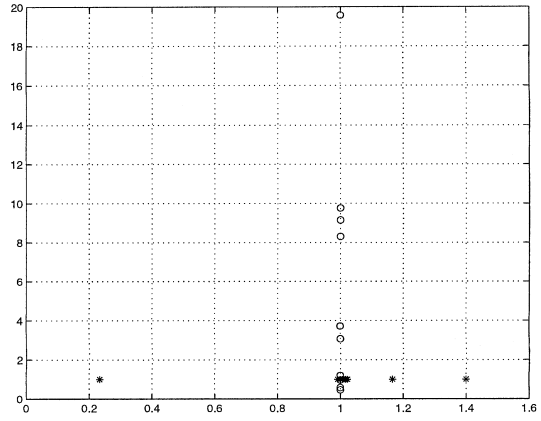

Fig. 3. The above plot shows the prediction errors of Table 1 (x axis) and the computational time of Table 2 (y axis) for PEM (o), N4SID (*). Both methods have been applied to the different industrial data sets. The computational complexity has been normalized to the value obtained by N4SID while the prediction error has been normalized to the value obtained by PEM. One can clearly see that the subspace method is much faster than the prediction error method although a little less accurate.

are suboptimal w.r.t. prediction error methods, who minimize the maximum likelihood criterion. On the other hand, this does not mean that PEM are always better than subspace methods. From the glass oven example for instance, it can be seen that the error for PEM is much larger than for subspace identification. This means that the PEM algorithm only found a local optimum which, in this case, appears to be much worse than the suboptimal solution of N4SID. Furthermore, from a computational point of view, the subspace methods are about 20 times faster (see Fig. 3).

The conclusion of this comparison is that subspace methods present a valid alternative to the "classical" versions of prediction error methods (PEM). They are fast because no iterative nonlinear optimization methods are involved and moreover, they are sufficiently accurate

Table 1
Prediction errors for the validation data[a]

	PEM	N4SID
Glass tubes	13.8	14
Dryer	4.4	4.5
Glass oven	56.4	13.2
Wing flutter	1.2	1.4
Robot arm	1.5	2.1
Evaporator	22.4	22.8
Chemical process	60.7	60.2
CD player arm	16.0	16.1
Ball and beam	36.2	36.5
Wall temperature	16.0	16.0

[a] If no validation data was available, the identification data was used. PEM computes the most accurate models (for almost all cases). We conclude from these tables that the subspace identification algorithms compute accurate models, and that these models (if needed) provide excellent initial starting values for optimization algorithms.

Table 2
Computational complexity of the algorithms, i.e. the number of floating point operations (flops) computed by Matlab[a]

	PEM	N4SID
Glass tubes	515	62
Dryer	2.1	4.5
Glass oven	791	81
Wing flutter	46	15
Robot arm	7	12
Evaporator	942	103
Chemical	3640	186
CD player arm	141	38
Ball and beam	14	14.5
Wall temperature	48	40

[a] The optimization based algorithm (PEM) is a lot slower for multivariable systems (up to a factor 20 for the chemical process example) than the subspace algorithm (N4SID).

in practical applications. From a theoretical point of view, prediction error methods are more accurate than subspace methods, as they clearly optimize an objective function. However, if a good initial estimate of the model parameters is not available, the solution one finds might not be the optimal solution (due to local minima in the optimization problem).

4. Conclusions

In this paper we have given a brief overview of linear subspace system identification methods. We made a clear distinction between methods using the states and methods starting from the observability matrix to recover the system parameters. Further a direct comparison between the N4SID subspace identification method and a PEM identification method on the basis accuracy and computational time is made. Therefore, both methods were applied on a wide variety of real-life

data sets. The conclusion is that subspace methods and prediction error methods are complementary in the sense that a good initial model can be quickly obtained with subspace methods while a further optimization of the parameters (if needed at all) can be done with prediction error methods.

Acknowledgements

W. Favoreel is a Research Assistant supported by the I.W.T. (Institute for Science and Technology, Flanders). B. De Moor is a Research Associate supported by the Fund for Scientific Research (F.W.O. Vlaanderen). P. Van Overschee was a Senior Research Assistant of the Fund for Scientific Research (F.W.O. Vlaanderen) and is now working at ISMC (Intelligent System Modeling and Control). Work supported by the Flemish Government (Administration of Science and Innovation (Concerted Research Action MIPS: Model-based Information Processing Systems, Bilateral International Collaboration: Modeling and Identification of nonlinear systems, IWT-Eureka SINOPSYS: Model-based structural monitoring using in-operation system identification), FWO-Vlaanderen: Analysis and design of matrix algorithms for adaptive signal processing, system identification and control, based on concepts from continuous time system theory and differential geometry, Numerical algorithms for subspace system identification: Extension towards specific applications, FWO-Onderzoeksgemeenschappen: Identification and Control of Complex Systems, Advanced Numerical Methods for Mathematical Modeling); Belgian Federal Government (Interuniversity Attraction Pole IUAP IV/02: Modeling, Identification, Simulation and Control of Complex Systems, Interuniversity Attraction Pole IUAP IV/24: IMechS: Intelligent Mechatronic Systems); European Commission: (Human Capital and Mobility: SIMONET: System Identification and Modeling Network, SCIENCE-ERNSI: European Research Network for System Identification.)

References

[1] A. Van Der Veen, E.F. Deprettere, A.L. Swindlehurst, Subspace-based signal analysis using singular value decompositions, Proceedings of the IEEE 81 (9) (1993) 1277–1308.

[2] M. Viberg, Subspace-based methods for the identification of linear time-invariant systems, Automatica 31 (12) (1995) 1835–1852.

[3] B. De Moor, P. Van Overschee, Numerical algorithms for subspace state space system identification, in: A. Isidori (Ed.), Trends in Control. A European Perspective, Springer, Italy, 1995, pp. 385–422.

[4] P. Van Overschee, B. De Moor, Subspace Identification for Linear Systems: Theory, Implementation, Applications. Kluwer Academic, Dordrecht, The Netherlands, 1996.

[5] L. Ljung, Developments for the system identification toolbox for matlab, in: Proc. of the 11th IFAC Symposium on System Identification, SYSID 97, 8–11 July, Kitakyushu, Japan, vol. 2, pp. 969–973, 1997.

[6] B. De Moor, Mathematical concepts and techniques for modeling of static and dynamic systems, PhD thesis, Department of Electrical Engineering, Katholieke Universiteit Leuven, Belgium, 1988.

[7] P. Van Overschee, B. De Moor, N4SID: subspace algorithms for the identification of combined deterministic and stochastic systems, Automatica 30 (1) (1994) 75–93.

[8] M. Verhaegen, P. Dewilde, Subspace identification, part I: the output-error state space model identification class of algorithms, Internat. J. Control 56 (1992) 1187–1210.

[9] W.E. Larimore, Canonical variate analysis in identification, filtering and adaptive control, in: Proc. of the 29th Conference on Decision and Control, CDC 90, HI, pp. 596–604, 1990.

[10] M. Viberg, Subspace methods in system identification, in: Proc. of the 10th IFAC Symposium on System Identification, SYSID 94, 4–6 July, Copenhagen, Denmark, vol. 1, pp. 1–12, 1994.

[11] M. Jansson, B. Wahlberg, On consistency of subspace based system identification methods, in: Proc. of the 13th IFAC World Congress, San Francisco, CA, pp. 181–186, 1996.

[12] P. Van Overschee, B. De Moor, Choice of state space basis in combined deterministic–stochastic subspace identification, Automatica 31 (2) (1995) 1877–1883.

[13] T. Gustafsson, System identification using subspace-based instrumental variable methods, in: Proc. of the 11th IFAC Symposium on System Identification, SYSID 97, 8–11 July, Kitakyushu, Japan, vol. 3, pp. 1119–1124, 1997.

[14] B. Ottersten, M. Viberg, A subspace based instrumental variable method for state space system identification, in: Proc. of the 10th IFAC Symposium on System Identification, SYSID 94, 4–6 July, Copenhagen, Denmark, vol. 2, pp. 139–144, 1994.

[15] S.Y. Kung, A new identification method and model reduction algorithm via singular value decomposition, in: 12th asilomar conf. on Circuits, Systems and Comp., pp. 705–714, 1978.

[16] B. De Moor, P. Van Overschee, W. Favoreel, Numerical Algorithms for Subspace State Space System Identification—An overview, Birkhauser, 1998.

[17] L. Ljung, System Identification—Theory for the User. Prentice Hall, Englewood Cliffs, NJ, 1987.

[18] T. McKelvey, On state-space models in system identification, PhD thesis, Department of Electrical Engineering Linköping University, Sweden, 1994.

[19] T. McKelvey, Ssid—a matlab toolbox for multivariable state-space model identification, Technical report, Dept. of EE, Linköping University, Linköping, Sweden, 1994.