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

VARMAX-based closed-loop subspace model identification

Ivo Houtzager, Jan-Willem van Wingerden, and Michel Verhaegen

Abstract— **In this paper a predictor-based subspace model identification method is presented that relaxes the requirement that the past window has to be large for asymptotical consistent estimates. By utilizing a VARMAX model, a finite description of the input-output relation is formulated. An extended least squares recursion is used to estimate the Markov parameters in the VARMAX model set. Using the Markov parameters the state sequence can be estimated and consequently the system matrices can be recovered. The effectiveness of the proposed method in comparison with an existing method is emphasized with a simulation study on a wind turbine model operating in closed loop.**

I. INTRODUCTION

Subspace Model Identification (SMI) methods are efficient methods to identify Linear Time-Invariant (LTI) state-space models from Multi-Input and Multi-Output (MIMO) measurements of a dynamic system and are described in detail in [13], [15]. These methods store input and output data in structured block Hankel matrices, such that it is possible to retrieve certain subspaces that are related to the system matrices. The key linear algebra steps, which are a RQ factorization, an SVD, and the solution of a linear least-squares problem, gives SMI methods their simplicity and numerical stability. However, it is known that SMI methods give biased results when the system to be modelled operates in closed-loop, because the future inputs are correlated with the past noises, due to the feedback controller.

An alternative to SMI is the Observer/Kalman-filter IDentification (OKID) method, see [10], [5]. This method obtains first Markov parameters, by estimating an Vector Auto Regressive with eXogenous inputs (VARX) model. VARX models, with high order, can provide asymptotical consistent estimates even on closed-loop data if there is sufficient excitation from an external signal or a controller of sufficiently high order, and that the system to be modelled or the feedback controller does not contain direct feedthrough. Then from the Markov parameters, a Hankel matrix is constructed. Finally, the system matrices of the state space model are directly estimated from the Hankel matrix by using the Eigensystem Realization Algorithm (ERA). To obtain a good estimate of the system matrices, a large number of Markov parameters are needed to construct a sufficient large Hankel matrix, making OKID less efficient then SMI.

Recently in [7], [4], [2], [1], a number of significant advances have been presented to identify LTI state-space models from measurements of a dynamic system operating in closed loop. The differences between these closed-loop methods are given in detail in [2], [1]. Similar to the OKID method, these recent developments are extensively utilizing a VARX model. Both are using a VARX model to obtain the Markov parameters, but instead the SSNEW ([7]), SSARX ([4]) and Predictor-Based Subspace IDentification (PBSID) method ([1], [2]) are using the Markov parameters to construct a Toeplitz matrix, where from multiplication with past input and output data, and an SVD, an estimation of the state sequence can be obtained. With the state sequence, it is straightforward to recover the system matrices.

In this paper, we present an improvement on the first part of these methods; the estimation of the Markov parameters. Inspired by [9], an Vector AutoRegressive Moving Average with eXogenous inputs (VARMAX) model is utilized, which enables consistent estimates of the model parameters of the finite-order input-output model. This relaxes the asymptotic consistency results in the case of VARX models, resulting in a very large past window in practice. A large past window has a number of disadvantages. First is that the increasing number of parameters to be fit gives an increase in the variance of the estimate. Second is that the matrix with stacked input and output data becomes more ill-conditioned, because normally in practice the input signals do not persistently excite the system enough, which means that the estimate becomes very sensitive to perturbations on the measurement data. Third is that in recursive implementations of SMI the computation cost increases considerably, see [3].

The estimation of a VARMAX model is a non-linear problem; however it sometimes can be estimated quite efficiently. Instead of the residual whitening operations in [9], we propose the use of the Extended Least Squares (ELS) recursion decribed in [6], [8] to solve the model estimation problem. We do not recover the system matrices using ERA, which requires a large number of Markov parameters, but use instead the $\text{PBSID}_{\text{opt}}$ method in [1]. The avoidance of the explicit use of Markov parameters (like in $\text{PBSID}_{\text{opt}}$ through the state) generally results in more efficient calculations and an improvement of accuracy of the estimated model.

The outline is as follows. In Section II, the VARX and VARMAX model sets are defined and discussed, the theoretical framework for SMI of LTI systems operating in closed-loop is presented, and a batch-wise solution is given. In Section III, the effectiveness of the proposed VARMAX-based method in comparison with the existing VARX-based method are emphasized with a simulation study on a wind turbine model operating in closed loop. In the Section IV, we present the conclusion.

This research is supported under the WE@SEA program of SenterNovem, an agency of the Dutch Ministry of Economic Affairs to promote sustainable development and innovation.

Delft Center of Systems and Control, Delft University of Technology, Delft, 2628 CD The Netherlands (Tel: +3115 27 86707; e-mail: I.Houtzager@TUDelft.nl)

II. VARMAX-BASED SUBSPACE MODEL IDENTIFICATION FOR LTI SYSTEMS OPERATING IN CLOSED-LOOP

In this section a SMI for LTI systems operating in closed-loop is presented that does not require the past window to be very large. First in Section II-A, we describe a general problem formulation and in Section II-B the notations and assumptions are explained. In Section II-C, we define and discuss the VARX/VARMAX model sets. Section II-D will describe the data equation and the relation to the state-sequence. In Section II-E, the main estimation problem is solved for batches of input and output sequences and it is given how to obtain the system matrices.

A. Problem formulation

Consider that the dynamics of the system to be modelled can be written in the following minimal state-space model in the innovation form:

$$\mathcal{S}\left\{\begin{array}{rl} x_{k+1} & = Ax_k + Bu_k + Ke_k, \\ y_k & = Cx_k + Du_k + e_k, \end{array}\right. \tag{1}$$

where $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^r$, $y_k \in \mathbb{R}^\ell$, are the state, input and output vectors, and $e_k \in \mathbb{R}^\ell$ denotes the zero-mean white innovation process noise. The state-space matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times r}$, $C \in \mathbb{R}^{\ell\times n}$, $D \in \mathbb{R}^{\ell\times r}$, and $K \in \mathbb{R}^{n\times \ell}$ are also called the system, input, output, direct feedthrough, and Kalman gain matrix, respectively. It is well-known that an invertible linear transformation of the state does not change the input-output behavior of a state-space system. Therefore, we can only determine the system matrices up to a similarity transformation $T \in \mathbb{R}^{n\times n}$: $T^{-1}AT$, $T^{-1}B$, $T^{-1}K$, CT, and D. The identification problem can now be formulated as:

Problem Decription 2.1 (LTI system identification):
Given the input sequence u_k, output sequence y_k over a time $k = \{0, \ldots, N-1\}$; find all, if they exist, system matrices A, B, C, D, and K up to a global similarity transformation.

The main assumptions are that the system to be modelled $\mathcal{S}$ is considered observable, the noise sequence e_k is white, the input sequence u_k has sufficient excitation, and the feedback loop does not have direct feedthrough. Further, the problem formulation does not require any other assumptions on the correlation between the input and noise sequence, which opens the possibility to apply the algorithm in closed loop.

B. Notations

The minimal state-space model denoted by (1) corresponds to the following pair of transfer functions:

$$y_k = G(z)\, u_k + H(z)\, e_k, \tag{2}$$

where z denotes the forward shift operator in the following descriptions of the proper transfer functions:

$$G(z) = D + C\left(zI^n - A\right)^{-1} B,$$
$$H(z) = I^\ell + C\left(zI^n - A\right)^{-1} K.$$

Note that I^n is used to represent a n-by-n identity matrix; and $O^{m\times n}$ a m-by-n zero matrix. The state-space description in (1) and the transfer function description in (2) have another equivalent description as a power series of the form:

$$y_k = G(z,\Xi)\, u_k + H(z,\Xi)\, e_k, \tag{3}$$

where the following relationship involving a formal power series holds:

$$\begin{aligned} G(z,\Xi) &= \Xi_0^u + z^{-1}\Xi_1^u + z^{-2}\Xi_2^u + z^{-3}\Xi_3^u + \ldots \\ &= \sum_{i=0}^{\infty} \Xi_i^u z^{-i}, \quad \Xi_i^u = \left\{\begin{array}{ll} D, & \text{if } i = 0, \\ CA^{i-1}B, & \text{if } i > 0, \end{array}\right. \end{aligned} \tag{4}$$

$$\begin{aligned} H(z,\Xi) &= I^\ell + z^{-1}\Xi_1^e + z^{-2}\Xi_2^e + z^{-3}\Xi_3^e + \ldots \\ &= I^\ell + \sum_{i=1}^{\infty} \Xi_i^e z^{-i}, \quad \Xi_i^e = CA^{i-1}K, \end{aligned} \tag{5}$$

where Ξ denotes the set of Markov parameters and $\Xi_i^u \in \mathbb{R}^{\ell\times r}$ and $\Xi_i^e \in \mathbb{R}^{\ell\times\ell}$ are the Markov parameters in the power series of the corresponding transfer functions. Each of these representations has its advantages and disadvantages, therefore we will frequently change the considered representation in the remainder of this paper.

We define a past window denoted by $p \in \mathbb{N}^+$ and a future window denoted by $f \in \mathbb{N}^+$, where $f \leq p$. These windows are used to define the following stacked vectors:

$$\bar{y}_{k-p} = \begin{bmatrix} y_{k-p} \\ y_{k-p+1} \\ \vdots \\ y_{k-1} \end{bmatrix}, \quad \bar{y}_k = \begin{bmatrix} y_k \\ y_{k+1} \\ \vdots \\ y_{k+f-1} \end{bmatrix},$$

In a similar way we can obtain the stacked vectors $\bar{u}_{k-p}$, $\bar{u}_k$, $\bar{e}_{k-p}$, and $\bar{e}_k$. Also, we define the stacked matrix Y:

$$Y = \begin{bmatrix} y_p, & \cdots, & y_{N-1} \end{bmatrix}.$$

In a similar way we can obtain the stacked vectors U, X. Further, we define the stacked matrix $\bar{Z}$:

$$\bar{Z} = \begin{bmatrix} \bar{z}_0, & \cdots, & \bar{z}_{N-p} \end{bmatrix}.$$

We also introduce the lifted matrices:

$$\mathcal{H} = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ CA^{f-2}B & CA^{f-3}B & \cdots & D \end{bmatrix}, \quad \Gamma = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{f-1} \end{bmatrix},$$
$$\mathcal{K} = \begin{bmatrix} A^{p-1}B & \cdots & AB & B \end{bmatrix},$$

where $\mathcal{H} \in \mathbb{R}^{f\ell\times fr}$ is the impulse matrix with a lower block triangular Toeplitz structure, $\Gamma \in \mathbb{R}^{f\ell\times n}$ is the extended observability matrix, and $\mathcal{K} \in \mathbb{R}^{n\times pr}$ is the extended controllability matrix.

C. VARX and VARMAX model sets

First, we will describe the well known VARX model set. Let the state-space model in (1) be rewritten in the Kalman predictor form as:

$$\left\{\begin{array}{rl} x_{k+1} & = \tilde{A}x_k + \tilde{B}u_k + Ky_k, \\ y_k & = Cx_k + Du_k + e_k, \end{array}\right. \tag{6}$$

with $\tilde{A} = A - KC$, and $\tilde{B} = B - KD$. Using the power series description this state-space model becomes:

$$y_k = \tilde{G}(z,\Xi)\,u_k + \left(I^\ell - \tilde{H}(z,\Xi)\right) y_k + e_k. \quad (7)$$

Considering a finite representation up to a past window p, then the approximated one-step-ahead predictor is:

$$\hat{y}_{k|k-1} = \tilde{G}_p(z,\Xi)\,u_k + \left(I^\ell - \tilde{H}_p(z,\Xi)\right) y_k, \quad (8)$$

where

$$\tilde{G}_p(z,\Xi) = \sum_{i=0}^{p} \tilde{\Xi}_i^u z^{-i}, \quad \tilde{H}_p(z,\Xi) = I^\ell - \sum_{i=1}^{p} \tilde{\Xi}_i^y z^{-i},$$

are the finite polynomial matrices up to a past window of the transfer functions $\tilde{G}(z) = D + C(zI^n - \tilde{A})^{-1}\tilde{B}$ and $\tilde{H}(z) = I^\ell - C(zI^n - \tilde{A})^{-1}K$, respectively. Hence, the so-called VARX model set is defined as:

Definition 2.2 (VARX model set): A VARX model set is determined by the polynomial matrices in (8) with:

$$\Xi \triangleq \begin{bmatrix} \tilde{\Xi}_p^u & \tilde{\Xi}_p^y & \cdots & \tilde{\Xi}_1^u & \tilde{\Xi}_1^y & \tilde{\Xi}_0^u \end{bmatrix}, \quad (9)$$

such that

$$\mathcal{M}_{\text{VARX}} \triangleq \left\{ \tilde{G}_p(z,\Xi), \tilde{H}_p(z,\Xi) \,|\Xi \in \mathbb{R}^{\ell \times pm+r} \right\}.$$

The one-step-ahead predictor has the property that is a linear function in the Markov parameters, which is very attractive from a computational point of view. Further, the one-step-ahead prediction error $\epsilon_{k|k-1} = y_k - \hat{y}_{k|k-1}$ consists of a truncation error term and a noise term. If the past window is not chosen large enough, the truncation error of the Markov parameters, can result to a biased estimation of the state sequence. The Kalman predictor, the optimal solution to the noise term, is only found when $p \to \infty$, because then $\tilde{G}_p \to \tilde{G}$ and $\tilde{H}_p \to \tilde{H}$. In the case that there is no noise present, the least-squares criterion will minimize the truncation error term only. Then (6) will become a deadbeat predictor, the optimal solution to the truncation error term, because the estimated K becomes a so-called deadbeat gain matrix of degree p, such that $\tilde{A}^i = 0$ for all $i \geq p$ and with $p \geq n$.

The observation of nilpotency has led to the introduction of the VARMAX model structure. As similar as in [9], we introduce another oberver matrix which creates additional freedom for the optimizer. By considering a deadbeat gain matrix M of degree p, which exists because from the measurement data only an observable part of the system can be identified, we can rewrite (1) in the deadbeat/Kalman predictor form as:

$$\begin{cases} x_{k+1} &= \bar{A}x_k + \bar{B}u_k + My_k + \bar{K}e_k, \\ y_k &= Cx_k + Du_k + e_k, \end{cases} \quad (10)$$

with $\bar{A} = A - MC$, $\bar{B} = B - MD$, and $\bar{K} = K - M$. Using the power series description and the deadbeat property that $\bar{A}^p = 0$ for all $p \geq n$, then the state-space model in (10) can be rewritten as:

$$y_k = \bar{G}(z,\Xi)\,u_k + \left(I^\ell - \bar{F}(z,\Xi)\right) y_k + \bar{H}(z,\Xi)\,e_k. \quad (11)$$

Considering a finite representation up to a past window p, then the corresponding one-step-ahead predictor is:

$$\begin{aligned} \hat{y}_{k|k-1} = \bar{G}_p(z,\Xi)\,u_k + \left(I^\ell - \bar{F}_p(z,\Xi)\right) y_k \\ + \left(\bar{H}_p(z,\Xi) - I^\ell\right) e_k, \end{aligned} \quad (12)$$

and also by substituting $e_k = y_k - \hat{y}_{k|k+1}$ in the former gives an expression which resembles the well-known ARMAX model structure as:

$$\begin{aligned} \bar{H}_p(z,\Xi)\,\hat{y}_{k|k-1} = \bar{G}_p(z,\Xi)\,u_k \\ + \left(\bar{H}_p(z,\Xi) - \bar{F}_p(z,\Xi)\right) y_k, \end{aligned} \quad (13)$$

where

$$\bar{G}_p(z,\Xi) = \sum_{i=0}^{p} \bar{\Xi}_i^u z^{-i}, \quad \bar{H}_p(z,\Xi) = I^\ell + \sum_{i=1}^{p} \bar{\Xi}_i^e z^{-i},$$

$$\bar{F}_p(z,\Xi) = I^\ell - \sum_{i=1}^{p} \bar{\Xi}_i^y z^{-i},$$

are the finite polynomial matrices up to a past window of the transfer functions $\bar{G}(z) = D + C(zI^n - \bar{A})^{-1}\bar{B}$, $\bar{H}(z) = I^\ell + C(zI^n - \bar{A})^{-1}\bar{K}$, and $\bar{F}(z) = I^\ell - C(zI^n - \bar{A})^{-1}M$ respectively. Hence, the so-called VARMAX model set is defined as:

Definition 2.3 (VARMAX model set): A VARMAX model set is determined by the polynomial matrices in (13) with:

$$\Xi \triangleq \begin{bmatrix} \bar{\Xi}_p^u & \bar{\Xi}_p^y & \bar{\Xi}_p^e & \cdots & \bar{\Xi}_1^u & \bar{\Xi}_1^y & \bar{\Xi}_1^e & \bar{\Xi}_0^u \end{bmatrix}, \quad (14)$$

such that

$$\mathcal{M}_{\text{VARMAX}} \triangleq \left\{ \bar{G}_p(z,\Xi), \bar{H}_p(z,\Xi), \bar{F}_p(z,\Xi) \,|\Xi \in \mathbb{R}^{\ell \times pm+r} \right\}.$$

Altough the one-step-ahead predictor does not have the property that it is a linear function in the Markov parameters, the computation of the solution can still be done efficiently using the extended least squares recursion, see also Section II-E. Now, the one-step-ahead prediction error consists only of a noise term, and therefore if $p \geq n$, then $\bar{G}_p = \bar{G}$ and $\bar{H}_p = \bar{H}$, thus no approximation is needed.

	$\mathcal{M}_{\text{VARX}}$	$\mathcal{M}_{\text{VARMAX}}$
$m =$	$r + \ell$	$r + 2\ell$
$z_k =$	$\begin{bmatrix} u_k^T & y_k^T \end{bmatrix}^T$	$\begin{bmatrix} u_k^T & y_k^T & e_k^T \end{bmatrix}^T$
$\widehat{A} =$	$\tilde{A}$	$\bar{A}$
$\widehat{B} =$	$\begin{bmatrix} \tilde{B} & K \end{bmatrix}$	$\begin{bmatrix} \bar{B} & M & \bar{K} \end{bmatrix}$
$\widehat{D} =$	$\begin{bmatrix} D & O^{\ell \times \ell} \end{bmatrix}$	$\begin{bmatrix} D & O^{\ell \times 2\ell} \end{bmatrix}$

TABLE I

GENERAL NOTATION FOR VARX AND VARMAX MODEL SETS

D. The data equation and the relation with the state

To make the notations more transparant, we introduce in Table I a general notation for all the model sets defined in the previous section. The input-output behavior of the model in (1) is now given by the data equation:

$$\bar{y}_k = \widehat{\Gamma} x_k + \widehat{\mathcal{H}} \bar{z}_k + \bar{e}_k. \quad (15)$$

Now we are going to introduce in this procedure an approximation for the state. The state x_k is given by:

$$x_k = \widehat{A}^p x_{k-p} + \widehat{\mathcal{K}} \bar{z}_{k-p}, \quad (16)$$

where $\widehat{A}^p$ is the transition matrix. The main assumption in this section is that we assume that for all model sets the matrix A is deadbeat with degree p, thus $\widehat{A}^i = 0$ for all $i \geq p$. In the case of the VARMAX model set this is automatically satisfied, because a deadbeat observer matrix M exists due to the observability of the system to be modelled. For the VARX model set it can be shown that if the model in (6) is uniformly exponential stable, the truncation error can be made neglectible small by making p large, see also [1], [2]. With the assumption of nilpotency, the state x_k is given by:

$$x_k = \widehat{\mathcal{K}}\bar{z}_{k-p}. \tag{17}$$

In a number of closed-loop SMI methods it is well known to make this relation, see [4], [1]. The output behavior is now approximately given by:

$$y_k = C\widehat{\mathcal{K}}\bar{z}_{k-p} + Du_k + e_k. \tag{18}$$

With the approximation given in (17) we rewrite (15) as:

$$\bar{y}_k = \widetilde{\Gamma\mathcal{K}}\bar{z}_{k-p} + \widehat{\mathcal{H}}\bar{z}_k + \bar{e}_k. \tag{19}$$

The product between the observability and the controllability matrix results in $\widetilde{\Gamma\mathcal{K}} \in \mathbb{R}^{f\ell\times pm}$ and is given by:

$$\widetilde{\Gamma\mathcal{K}} = \begin{bmatrix} C\widehat{A}^{p-1}\widehat{B} & C\widehat{A}^{p-2}\widehat{B} & \cdots & C\widehat{B} \\ 0 & C\widehat{A}^{p-1}\widehat{B} & \ddots & C\widehat{A}\widehat{B} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & C\widehat{A}^{f-1}\widehat{B} \end{bmatrix}. \tag{20}$$

This is an upper block triangular matrix, because the introduced zeros come from the deadbeat assumption[1]. It is clearly visible that the first block row in (20) can be used to construct all the other block rows. Observe now that the product between the state and the observability matrix is given by:

$$\widetilde{\Gamma\mathcal{K}}\bar{z}_{k-p} = \widehat{\Gamma}x_k, \tag{21}$$

and this observation is the key idea behind the PBSID$_{\text{opt}}$ method. This implies that we have to find an estimate of $C\widehat{\mathcal{K}}$ to construct $\widetilde{\Gamma\mathcal{K}}$. In Section II-C, the Markov parameter sets are described such that $\Xi \equiv [C\widehat{\mathcal{K}}\ D]$ and consequently can be used to construct $\widetilde{\Gamma\mathcal{K}}$. To summarize, after the construction of the matrix $\widetilde{\Gamma\mathcal{K}}$ we obtain a product between the observability and the state sequence. The approximation of the matrix $\widetilde{\Gamma\mathcal{K}}$ described in (20), which can be fully constructed by the Markov parameter set Ξ and is given by[2]:

$$\widetilde{\Gamma\mathcal{K}} = \begin{bmatrix} \Xi_{(:,1:pm)} \\ \begin{bmatrix} O^{\ell\times m}, & \Xi_{(:,1:(p-1)m)} \end{bmatrix} \\ \begin{bmatrix} O^{\ell\times 2m}, & \Xi_{(:,1:(p-2)m)} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} O^{\ell\times(f-1)m}, & \Xi_{(:,1:(p-f+1)m)} \end{bmatrix} \end{bmatrix}. \tag{22}$$

[1]Remark: For the VARX-based PBSID$_{\text{opt}}$ method, the zeros are introduced by considering an approximation of the matrix $\widetilde{\Gamma\mathcal{K}}$. ([1], [2])

[2]For simplicity MATLAB notation is used.

E. Batch-wise solution of the model identification problem

The batch-wise solution of the presented closed-loop SMI scheme is divided in the following three steps:

1) The estimation of the Markov parameters: Using the VARX model structure and if the matrix $\Psi = \begin{bmatrix} Z^T & U^T \end{bmatrix}^T$ has full row rank, the Markov parameter set Ξ can be estimated by solving the following linear problem:

$$\min_{\Xi} \|Y - \Xi\Psi\|_F^2. \tag{23}$$

For finite p the solution of this linear problem will be biased due to the approximation made in (17). If $p \to \infty$ the bias disappears. However, the matrix Ψ can become ill-conditioned for large past window sizes, if the input signals do not persistently excite the system enough. For example, in identification experiments it is most of the time needed to design the excitation signal such that it does not exceed the load specifications and ensures that the system to be modelled operates around a particular operation point.

For the VARMAX model set, the estimation problem can not be solved directly using ordinary least squares, because the noise elements in Z are not known in advance. In fact this becomes a non-linear problem, which can be solved for example by applying the residual whitening iterations ([9]) or the recursive Extended Least Squares (ELS) method ([8]), where the recursive scheme is prefered by experience due to its simplicity, computational complexity, and better converging properties. Now that the solution Ξ is computed recursively, the costs of (23) is replaced by:

$$\min_{\Xi} \left[\sum_{k=0}^{N} \lambda^{N-k} \|y_k - \Xi\bar{z}_{k-p}\|_F^2 + \lambda^{N-1}\rho^2 \Xi\Pi\Xi^T \right], \tag{24}$$

where the scalar λ is called the forgetting factor and the matrix Π is the regularization at the beginning of the iterations. It is recommended to employ such an exponentially-weighted and regularized cost function, because it ensures that during the initial iterations of the scheme a local solution exists and converges. It is possible to apply the exponentially weighting only during initial calculations and set it later close to one, which means no forgetting. The cost function can easily be solved by an exponentially-weighted and regularized RLS scheme, because $\bar{z}_{k-p}$ does not depend on e_k. Given the regularization value $\rho > 0$, and a forgeting factor $0 \ll \lambda \leq 1$, the solution Ξ can be computed recursively using ELS, by starting with $P_{-1} = \frac{1}{\rho}I^{pm+r}$, and iterate as[2]:

$$\begin{aligned} \bar{z}_{k-p} &= \begin{bmatrix} \bar{z}_{k-p-1}^T{}_{(m+1:pm+r)} & y_{k-1}^T & e_{k-1}^T & u_k^T \end{bmatrix}^T, \\ G_k &= \bar{z}_{k-p}^T P_{k-1} \left(\lambda I + \bar{z}_{k-p}^T P_{k-1} \bar{z}_{k-p}\right)^{-1}, \\ \widehat{\Xi}_k &= \widehat{\Xi}_{k-1} + \left(y_k - \widehat{\Xi}_{k-1}\bar{z}_{k-p}\right) G_k, \\ P_k &= \lambda^{-1}\left(P_{k-1} - P_{k-1}\bar{z}_{k-p}G_k\right), \\ e_k &= y_k - \widehat{\Xi}_k\bar{z}_{k-p}. \end{aligned}$$

For clarification, the iteration shows the conventional implementation of the recursive ELS scheme. For practical use it is recommended to use square-root or array-type of implementations, due to better numerical robustness against round-off

errors, see for more discussion [3]. Another advantage is that array methods can be used to exploit the shift structure in the data, for example a fast-array scheme that exploits the shift structure of the data vector $\bar{z}_{k-p}$ in (24) results in an algorithm with computational load of $\mathcal{O}(pm^2N)$, instead of $\mathcal{O}(p^2m^2N)$.

The practical experience with different models is that the global minimum is usually found without too much problem, especially if the initial Ξ_{-1} obtained from (23) is close to the global minimum. If the solution still has not converged to the minimum yet, then the iteration can be repeated with as initial solution the previous. Altough not common, the method is known to diverge if the following positive real condition:

$$\mathrm{Re}\left\{\bar{H}\left(e^{i\omega}\right)^{-1}\right\} > \frac{1}{2}, \quad \forall\omega \in \mathbb{R}, \tag{25}$$

is not fulfilled, see for proof and possible solutions [6], [8].

2) The estimation of state sequence: After the estimation of the Markov parameters, the matrix $\widetilde{\Gamma\mathcal{K}Z}$ is constructed using (22), which equals by definition the extended observability times the state sequence, $\tilde{\Gamma}X$. By computing a Singular Value Decomposition (SVD) of $\widetilde{\Gamma\mathcal{K}Z}$ as:

$$\widehat{\widetilde{\Gamma\mathcal{K}Z}} \approx \begin{bmatrix} U & U_\perp \end{bmatrix} \begin{bmatrix} \Sigma_n & 0 \\ 0 & \Sigma \end{bmatrix} \begin{bmatrix} V \\ V_\perp \end{bmatrix}, \tag{26}$$

the state sequence X and the order of the system n is retrieved. The diagonal matrix Σ_n contains the n largest singular values and the orthogonal matrix V contains the corresponding row space. Note that we can find the largest singular values by detecting a gap between the singular values. The state sequence is now estimated by:

$$\widehat{X} = \Sigma_n^{1/2} V = \left(U\Sigma_n^{1/2}\right)^{\dagger} \widehat{\widetilde{\Gamma\mathcal{K}Z}}. \tag{27}$$

3) The estimation of the system matrices: It is well known that when the state, input, and output sequence are known, the system matrices A, B, C, D, K can be estimated by solving two linear problems obtained from (1). In [15], it is described to find a guarenteed stabilizing estimate of the observer matrix K using the Riccati equation.

III. SIMULATION STUDY

In the previous section, we presented an identification approach to identify LTI systems operating in open loop and closed loop. In this section we use a wind turbine model of the dynamics to demonstrate the effectiveness of the algorithm with a small past window size.

A. First-principle model of a horizontal-axis wind turbine

In this paper, we consider a seventh order model of a Horizontal-Axis Wind Turbine (HAWT) described in [14] to demonstrate the closed-loop subspace LPV system identification algorithm. The model describes the rotational dynamics of a wind turbine around a particular operating poin and has a constant state matrix while the input and output matrices strongly depend on the azimuth angle. Similar as in [12], [11], the Coleman transformation have been used to transform the model to LTI. After the Coleman transformation, the system to be modelled fits in the model structures ($\mathcal{S} \in \mathcal{M}$).

B. Simulation of the closed-loop wind turbine model

The LTI model of the HAWT is used to obtain the input, and output sequence for the identification algorithm. For this purpose, the equations are converted to discrete time using a zero-order hold discretization method with a sample time of 0.1 s. The wind turbine system is not asymptotically stable, it has an integrator. Therefore, controllers are added in a feedback loop to the system for stabilization. The descriptions of these controllers and additional filters used can be found in [12], [11]. For consistent closed-loop estimation for all frequencies, it is important that there is sufficient excitation from an external excitation signal or a controller of sufficiently high order. As the controllers are not sufficiently high order, we take an additional zero-mean white noise with $\mathrm{var}\left(\theta_{k,i}\right) = 1$ deg, which is added to the control signal of the collective pitch controller. As additional excitation input for the generator torque we take also a zero-mean white noise signal with $\mathrm{var}\left(T_{ge,k}\right) = 1 \cdot 10^6$ Nm. The wind disturbance signal is also zero-mean white noise with $\mathrm{var}\left(v_{k,i}\right) = 1$ m/s, but this signal is assumed to be unknown.

C. Closed-loop subspace model identification results

The collected data of u_k, and y_k from the simulations are used in the identification experiments. For the identification experiments we used $N = 10000$, $p = f = 10$ and $\lambda = 0.99$. To emphasize the difference in performance with small windows, the VARX model estimation is also carried out with the same small window size. To investigate the sensitivity of the identification algorithm with respect the wind disturbances, a Monte-Carlo simulation with 100 runs was carried out. For each of the 100 simulations a different realization of the input u_k and wind disturbance v_k is used. The performance of the identified system is evaluated by looking at the singular values of matrix $\widetilde{\Gamma\mathcal{K}Z}$, and comparing the frequency response functions of the identified models to the real model.

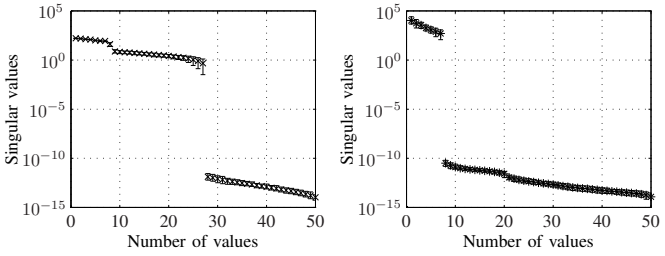

Fig. 1. The singular values of matrix $\widetilde{\Gamma\mathcal{K}Z}$ for 100 experiments using VARX (left) and VARMAX (right).

In Figure 1 the singular values of matrix $\widetilde{\Gamma\mathcal{K}Z}$, including the errorbounds for 100 experiments are illustrated using VARX/VARMAX-based model identification. As expected, the singular values of VARMAX-based identification show a large gap after the first seven largest values, which equals the order of the system. This is not the case with the VARX-based identification, because the past window is close to order of the system to be modelled. Figures 2 and 3

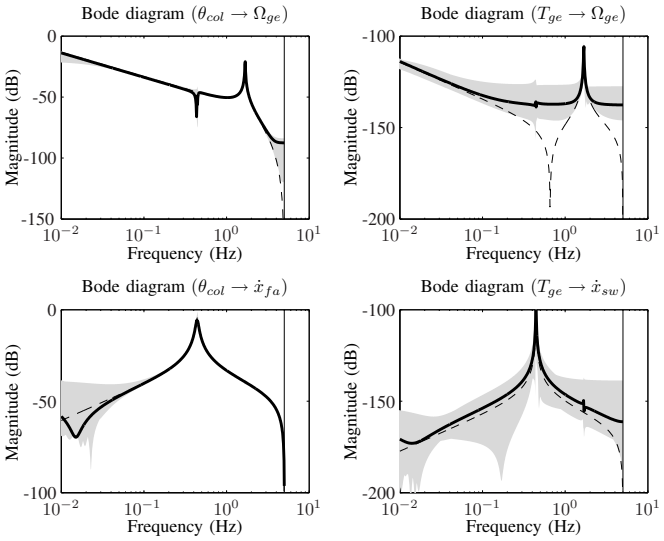

Fig. 2. Bode diagrams of the original transfer functions (dashed) and the identified transfer functions of the experiment with the best fit (bold) using VARX-based identification. The transfer functions of the other 99 experiments are within the gray region.

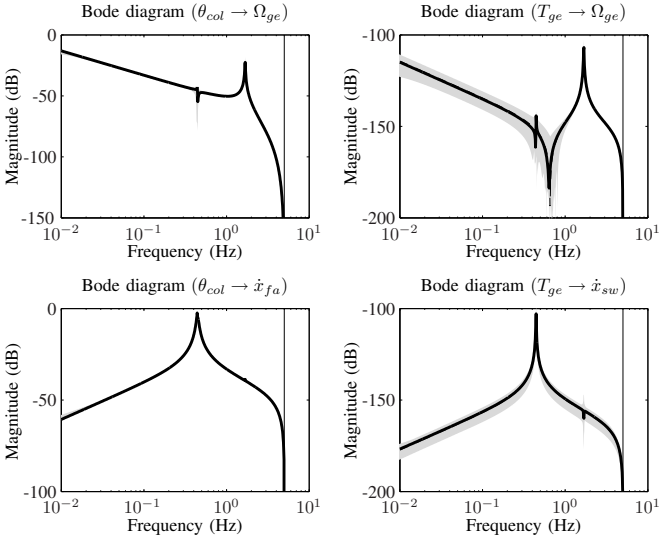

Fig. 3. Bode diagrams of the original transfer functions (dashed) and the identified transfer functions of the experiment with the best fit (bold) using VARMAX-based identification. The transfer functions of the other 99 experiments are within the gray region.

shows the Bode diagrams of a selected number of inputs and outputs. Adding the Moving Average terms, gives more consistent identifications results for the wind turbine system. It shows for both methods that the identified natural frequencies are very close to the true natural frequencies, although a zero, which is clearly visible in the transfer function of $(T_{ge} \rightarrow \theta_{ge})$, is not estimated with the VARX-based identification method.

Further simulations showed that increasing the value of the past window size, which was possible due to the additional white noise excitation signal, improves the model obtained from VARX-based identification considerably, and hardly any improvement was visible in the model obtained from VARMAX-based identification. As expected from the theory in Section II-B, the VARX-based identification method gives for large past window sizes ($p > 30$) almost similar results as the VARMAX-based identification results with $p = 10$.

IV. CONCLUSION

In this paper a PBSID$_{\text{opt}}$ method is presented that relaxes the requirement that the past window has to be large for asymptotical consistent estimates. By utilizing a VARMAX model, a finite description of the input-output relation is formulated. An extended least squares recursion is used to estimate the Markov parameters in the VARMAX model. Using the Markov parameters the state sequence can be estimated and consequently the system matrices can be recovered. The method is beneficial when the past window size is restricted, because the computational time for estimating VARMAX models can be larger then the batch-wise estimation VARX models with larger window sizes. The effectiveness of the proposed method with a small past window is shown by a simulation study on a wind turbine model operating in closed loop.

REFERENCES

[1] A. Chiuso, "The role of vector auto regressive modeling in predictor based subspace identification," *Automatica*, vol. 43, no. 6, pp. 1034–1048, 2007.

[2] A. Chiuso and G. Picci, "Consistency analysis of some closed-loop subspace identification methods," *Automatica*, vol. 41, no. 3, pp. 377–391, 2005.

[3] I. Houtzager, J. W. van Wingerden, and M. Verhaegen, "Fast-array recursive closed-loop subspace model identification," in *15th IFAC Symposium on System Identification*, Saint-Malo, France, 2009.

[4] M. Jansson, "Subspace identification and ARX modeling," in *13th IFAC Symposium on System Identification*, Rotterdam, The Netherlands, 2003.

[5] J.-N. Juang and R. S. Pappa, "An eigensystem realization algorithm for model parameter identification and model reduction," *Journal of Guidance, Control, and Dynamics*, vol. 8, no. 5, pp. 620–627, 1985.

[6] L. Ljung, "On positive real transfer functions and the convergence of some recursive schemes," *IEEE Transactions on Automtic Control*, vol. 22, no. 4, pp. 539–551, 1977.

[7] L. Ljung and T. McKelvey, "Subspace identification from closed loop data," *Signal Processing*, vol. 52, pp. 209–215, 1996.

[8] L. Lung, *System Identification: Theory for the user, 2nd Edition*. Prentice Hall, 1987.

[9] M. Phan, L. G. Horta, J.-N. Juang, and R. W. Longman, "Improvement of observer/Kalman filter identification (OKID) by residual whitening," *Journal of Vibrations and Acoustics*, vol. 117, pp. 232–238, 1995.

[10] M. Phan, J.-N. Juang, and R. W. Longman, "Identification of linear multivariable systems by identification of observers with assigned real eigenvalues," *Journal of the Astronautical Sciences*, vol. 40, no. 2, pp. 261–279, 1992.

[11] K. Selvam, S. Kanev, J. W. van Wingerden, T. G. van Engelen, and V. M., "Feedback-feedforward individual pitch control for wind turbine load reduction," *International Journal of Robust and Nonlinear Control*, vol. 19, no. 1, pp. 72–91, 2009.

[12] T. G. van Engelen, H. Markou, T. Buhl, and B. Marrant, "Morphological study of aeroelastic control concepts for wind turbines," Energy research Centre of the Netherlands, Tech. Rep., Jul. 2007.

[13] P. Van Overschee and B. De Moor, *Subspace Identification for Linear Systems; Theory, Implementation, Applications*. Kluwer Academic Publishers, 1996.

[14] J. W. van Wingerden, I. Houtzager, F. Felici, and M. Verhaegen, "Closed-loop identification of the time-varying dynamics of variable-speed wind turbines," *International Journal of Robust and Nonlinear Control*, vol. 19, no. 1, pp. 4–21, 2009.

[15] M. Verhaegen and V. Verdult, *Filtering and System Identification*. Cambridge University Press, 2007.