The asymptotic variance of the PBSID$_{\text{opt}}$ algorithm

J.W. van Wingerden *

* *Delft University of Technology, Delft, 2628 CD The Netherlands (Tel: +3115 27 81720; e-mail: J.W.vanWingerden@tudelft.nl).*

Abstract: It is well-known that system identification is a valuable technique to obtain compact models for controller design and prediction. Subspace identification methods are of interest since they solely rely on tools from linear algebra and the relative ease to work with data generated by systems with multiple inputs and outputs. With respect to other methods (*e.g.* Prediction Error (PE)), subspace methods do not have straightforward asymptotic variance expressions. However, recently some papers appeared with asymptotic variance expressions for a certain class of identification algorithms that merge ideas from PE and subspace methods. In this paper we derive the asymptotic variance expression for the PBSID$_{\text{opt}}$ algorithm and come up with manageable expressions under some reasonable assumptions. We conclude the paper with two simulation examples where we show the strength of the proposed method.

1. INTRODUCTION

A model for modern model based controller design is a mathematical model normally governed by (preferably linear) differential equations. For controller synthesis this model should only contain the relevant dynamics between the input, the output, and the disturbances and should be accurate around the bandwidth of the controller. In experimental modeling, also referred to as system identification, actual input and output data of the system is used to obtain such a mathematical description of the system. System identification for Linear Time-Invariant (LTI) is a well-established methodology to obtain a model for control [Ljung, 1987]. Since the obtained models are obtained from noise corrupted data the model is an estimate of the true system. It is of key importance to acknowledge this uncertainty during the controller design phase.

For system identification a distinction can be made between state-space and input-output model representations, for which the most common identification procedures are Subspace Model Identification (SMI) and Prediction Error (PE) identification, respectively. The main differences between the two approaches are:

- In the PE setting, the first step is to do model structure selection: to select the structure (*e.g.* OE, ARX) and the corresponding orders. In the state-space setting, the only degree of freedom is the state order. However, if we consider the subspace identification scheme, an estimation of the order is a part of the algorithm. The model structures for the input-output setting and their corresponding algorithms are hard to translate to the MIMO setting, while in the state-space setting this occurs naturally.
- PE methods have well-established expressions for the uncertainty of the estimates (*e.g.* bias and variance)[Ljung, 1987], which provides a quality tag which can be used for controller design (*e.g.* robust control). Since SMI methods are two step procedures it is hard to find these expression. However, recently some papers reported asymptotic variance expression under some mild conditions. Typically these expression are rather cumbersome.

The asymptotic variance is an important property of an estimate to quantify the uncertainty. In this paper we will derive an asymptotic variance expression for the optimized Predictor Based Subspace IDentification (PBSID$_{\text{opt}}$) method of which no asymptotic variance result is published. [1] This particular algorithm combines ideas from the PE and the subspace setting (we will illustrate this in this paper). We are interested in this particular identification algorithm since it can deal with data generated under closed loop conditions and the excellent performance of this algorithm when dealing with real data [van Wingerden et al., 2010]. With respect to the traditional subspace methods (*e.g.* MOESP [Verhaegen, 1994] and N4SID [Van Overschee and De Moor, 1996]) the new methods (*e.g.* PBSID Chiuso [2007] and SSARX [Jansson, 2005]) have a close relationship with PEM methods since the first step is to solve a high order Vectorized-ARX problem while the second step can be viewed at as a model reduction step.

In Chiuso [2006] asymptotic variance expressions are derived for the original PBSID algorithm [2] and the algorithm derived by Qin and Ljung [2003]. At the end of that particular paper the assumption is made that the true system is of the ARX-type and the true system is within the model set. To make the notation transparent we will directly make these assumptions and we stay close to the notation used in the traditional subspace papers which will result in manageable formulaes for the asymptotic variance. Combined with the fact that we derive expres-

[1] Although in Chiuso [2007] it is shown that the asymptotic variance of the optimized version is equal or smaller than the original PBSID algorithm.

[2] In this particular paper labelled as whitening filter.

sions for the PBSID$_{\text{opt}}$ algorithm highlight the three main contributions of this paper.

The outline of this paper is as follows; we start in Section 2 with the problem formulation and assumptions. In Section 3 the basic idea behind the PBSID$_{\text{opt}}$ identification scheme is presented. In Section 4 expressions are derived for the asymptotic variance. In Section 5 a simulation example is presented. We end this paper with our conclusions and state a number of directions for future research.

2. PROBLEM FORMULATION

In this section we present the problem formulation and the assumptions we make. Furthermore, we introduce the notation which stays close to the traditional subspace papers.

2.1 Problem formulation

For the derivation of the algorithm we consider the following LTI system:

$$x_{k+1} = Ax_k + Bu_k + Ke_k, \tag{1}$$
$$y_k = Cx_k + e_k, \tag{2}$$

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. The vector $e_k \in \mathbb{R}^\ell$ denotes the zero mean white innovation process with covariance matrix Λ. The matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times r}$, $C \in \mathbb{R}^{\ell\times n}$, $K \in \mathbb{R}^{n\times \ell}$ are the local system, input, output, and the observer matrices. We can rewrite (1)-(2) in the predictor form as:

$$x_{k+1} = \tilde{A}x_k + Bu_k + Ky_k, \tag{3}$$
$$y_k = Cx_k + e_k, \tag{4}$$

with $\tilde{A} = A - KC$. 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 $T \in \mathbb{R}^{n\times n}$: $T^{-1}AT$, $T^{-1}B$, $T^{-1}K$, and CT.

The identification problem can now be formulated as: given the input sequence u_k and the output sequence y_k over a time $k = \{0, \ldots, N-1\}$; find all, if they exist, the system matrices A, B, K, C, and D up to a global similarity transformation. Moreover, find the asymptotic variance in a data-driven basis,[3] so:[4]

$$\text{plim}\Delta\Theta\Delta\Theta^{\mathrm{T}}, \tag{5}$$

with:

$$\Delta\Theta = [\text{vec}(\hat{\mathrm{A}} - \mathrm{T}^{-1}\mathrm{AT})^{\mathrm{T}}, \text{vec}(\hat{\mathrm{B}} - \mathrm{T}^{-1}\mathrm{B})^{\mathrm{T}}, \ldots \\ \text{vec}(\hat{\mathrm{C}} - \mathrm{CT})^{\mathrm{T}}, \text{vec}(\hat{\mathrm{K}} - \mathrm{T}^{-1}\mathrm{K})^{\mathrm{T}}]^{\mathrm{T}}. \tag{6}$$

[3] This data-driven basis will disappear when we look at input output related invariants such as transfer functions, or poles/zeros of the system.

[4] State space models typically overparameterize the system and this gives rise to the problem that the following matrix will be positive semidefinite (in other words will not have full rank).

2.2 Assumptions

The first assumption that we are going to make is that the matrix $\tilde{A}$ is nilpotent, so $||\tilde{A}^p|| = 0$ where we will define p as the past window. The second assumption is that we exactly know the model order of the true system.

2.3 Notation

Similar as in Jansson [2005], Chiuso [2007] we define a past window denoted by p. This window is used to define the following stacked vector:

$$\overline{z}_k^p = \begin{bmatrix} z_k \\ z_{k+1} \\ \vdots \\ z_{k+p-1} \end{bmatrix}, \qquad \text{with} \qquad \mathrm{z_k} = \begin{bmatrix} u_k \\ y_k \end{bmatrix}. \tag{7}$$

This notation allows us the introduce the following matrices:

$$Z = [\overline{z}_1^p, \cdots, \overline{z}_{N-p}^p], \tag{8}$$
$$\overline{Z} = [\overline{z}_2^p, \cdots, \overline{z}_{N-p}^p], \tag{9}$$
$$\underline{Z} = [\overline{z}_1^p, \cdots, \overline{z}_{N-p-1}^p]. \tag{10}$$

We also define:

$$X = [x_p, \cdots, x_N], \tag{11}$$
$$\overline{X} = [x_{p+1}, \cdots, x_N], \tag{12}$$
$$\underline{X} = [x_p, \cdots, x_{N-1}], \tag{13}$$

and in a similar way we have Y, E $\overline{E}$, and $\underline{E}$. Furthermore, we define a number of projections:

$$Z^\dagger = Z^T \left(ZZ^T\right)^{-1}, \tag{14}$$
$$\Pi_Z = Z^T \left(ZZ^T\right)^{-1} Z, \tag{15}$$
$$\Pi_Z^\perp = I - Z^T \left(ZZ^T\right)^{-1} Z. \tag{16}$$

At this point we also partition the following matrix:

$$Z^\dagger = \left[Z_{z_1}^\dagger, \cdots \; Z_{z_p}^\dagger\right]. \tag{17}$$

We assume that the state sequence X, $\overline{X}$ and $\underline{X}$ have full row rank and the extended observability matrix given by:

$$\Gamma^f = \begin{bmatrix} C \\ C\tilde{A} \\ \vdots \\ C\tilde{A}^{f-1} \end{bmatrix}, \tag{18}$$

has full column rank and f represents the future window. We also define the extended controllability matrix:

$$\mathcal{K}^p = \left[\tilde{A}^{p-1}\overline{B}, \; \tilde{A}^{p-2}\overline{B}, \; \cdots, \; \overline{B}\right], \tag{19}$$

with $\overline{B} = [B, K]$.

3. PREDICTOR BASED SUBSPACE IDENTIFICATION

In this section we give a brief summary of the identification algorithm (PBSID$_{\text{opt}}$) and some related equations will be introduced for the computation of the asymptotic variance.

3.1 Predictor

The first objective of the predictor-based algorithms is to reconstruct the state sequence up to a similarity transformation. The state x_{k+p} is given by:

$$x_{k+p} = \tilde{A}^p x_k + \underbrace{\left[\, \tilde{A}^{p-1}\overline{B},\ \tilde{A}^{p-2}\overline{B},\ \cdots,\ \overline{B}\,\right]}_{\mathcal{K}^p} \overline{z}_k^p. \quad (20)$$

At this point we use the assumption that $||\tilde{A}^p = 0||$. With this assumption the state x_{k+p} is given by:

$$x_{k+p} = \mathcal{K}^p \overline{z}_k^p. \quad (21)$$

The input-output behavior is now given by:

$$y_{k+p} = C\mathcal{K}^p \overline{z}_k^p + e_{k+p}\,. \quad (22)$$

If the matrix Z has full row rank the matrix $C\mathcal{K}^p$ can be estimated by solving the following linear regression problem: [5]

$$\min_{C\mathcal{K}^p} ||Y - C\mathcal{K}^p Z||_F^2, \quad (23)$$

where $||\cdots||_F$ represents the Frobenius norm [Golub and Loan, 1996]. The solution, error, and residual of this problem are given below:

$$\begin{aligned} \widehat{C\mathcal{K}} &= YZ^\dagger, \\ \Delta\widehat{C\mathcal{K}} &= EZ^\dagger, \\ \widehat{E} &= Y\Pi_Z^\perp, \end{aligned}$$

where $\widehat{(.)}$ is the estimated variable and Δ is defined such that the following equality holds: $\Delta\widehat{(.)} = \widehat{(.)} - (.)$.

3.2 Extended observability times controllability

With the standing assumptions, the product $\mathcal{K}^p Z$ represents by definition the state sequence, X, can not be estimated directly. In the predictor-based identification algorithms $C\mathcal{K}^p$ is used to construct the extended observability matrix times the extended controllability matrix. The following upper block triangular matrix is used in the $\text{PBSID}_{\text{opt}}$ algorithm (for $f = p$) [6]:

$$\Gamma^f\mathcal{K}^p = \begin{bmatrix} C\tilde{A}^{p-1}\overline{B} & C\tilde{A}^{p-2}\overline{B} & \cdots & C\overline{B} \\ 0 & C\tilde{A}^{p-1}\overline{B} & \cdots & C\tilde{A}\overline{B} \\ & & \ddots & \vdots \\ 0 & & & C\tilde{A}^{p-1}\overline{B} \end{bmatrix}. \quad (24)$$

Observe that from the linear regression problems formulated in (23) we can construct this matrix (for $f = p$):

$$\widehat{\Gamma^f\mathcal{K}^p} = \begin{bmatrix} YZ_{z_1}^\dagger & YZ_{z_2}^\dagger & \cdots & YZ_{z_p}^\dagger \\ 0 & YZ_{z_1}^\dagger & \cdots & YZ_{z_{p-1}}^\dagger \\ & & \ddots & \vdots \\ 0 & & & YZ_{z_1}^\dagger \end{bmatrix}. \quad (25)$$

And in a similar way we have:

$$\Delta\widehat{\Gamma^f\mathcal{K}^p} = \begin{bmatrix} EZ_{z_1}^\dagger & EZ_{z_2}^\dagger & \cdots & EZ_{z_p}^\dagger \\ 0 & EZ_{z_1}^\dagger & \cdots & EZ_{z_{p-1}}^\dagger \\ & & \ddots & \vdots \\ 0 & & & EZ_{z_1}^\dagger \end{bmatrix}, \quad (26)$$

[5] This is an vectorized ARX problem and well-known residual tests can be use to find p.

[6] This condition is only given to make the expressions suitable for presentation purposes. Typically $f < p$ and the corresponding matrix will be the first f block rows

with the following equality:

$$\widehat{\Gamma^f\mathcal{K}^p} = \Gamma^f\mathcal{K}^p + \Delta\widehat{\Gamma^f\mathcal{K}^p}. \quad (27)$$

From the constructed matrix $\widehat{\Gamma^f\mathcal{K}^p}$ we can compute $\widehat{\Gamma^f\mathcal{K}^p}Z$ which equals by definition the extended observability matrix times the state sequence, $\widehat{\Gamma^f X}$. By computing a Singular Value Decomposition (SVD) of this estimate we can estimate the state sequence and the order of the system. We will use the following SVD:

$$W\widehat{\Gamma^f\mathcal{K}^p}Z = [\,\mathcal{U}\ \mathcal{U}_\perp\,] \begin{bmatrix} \Sigma_n & 0 \\ 0 & \Sigma \end{bmatrix} \begin{bmatrix} \mathcal{V} \\ \mathcal{V}_\perp \end{bmatrix}, \quad (28)$$

where Σ_n is the diagonal matrix containing the n largest singular values and $\mathcal{V}$ is the corresponding row space. The matrix W represents a given weighting matrix. In this article we assume that we know the order of the underlying true system and consequently the state can be estimated as:

$$\widehat{X} = \underbrace{\left(W^{-1}\mathcal{U}\right)^\dagger}_{S} \widehat{\Gamma^f\mathcal{K}^p}Z. \quad (29)$$

It is well-known that this state contains a similarity transformation. So, the following equality holds:

$$\widehat{X} = TX + \Delta\widehat{X}. \quad (30)$$

This observation directly implies that the system matrices are estimated up to an unknown similarity matrix T.

3.3 Recovery of the system matrices

In the stacked version we have the following relation:

$$\widehat{\overline{X}} - \Delta\overline{X} = A\left(\underline{\widehat{X}} - \Delta\underline{X}\right) + B\underline{U} + K\left(\underline{\widehat{E}} - \Delta\underline{E}\right), \quad (31)$$

$$Y = C\left(\hat{X} - \Delta X\right) + \hat{E} - \Delta E. \quad (32)$$

Assuming we have a consistent estimate (for conditions see Appendix A) the solution is now given by:

$$\left[\widehat{A},\ \widehat{B},\ \widehat{K}\right] = \widehat{\overline{X}} \begin{bmatrix} \underline{\widehat{X}} \\ \underline{U}, \\ \underline{\widehat{E}} \end{bmatrix}^\dagger, \quad (33)$$

$$\left[\widehat{C}\right] = (Y - \hat{E})\widehat{X}^\dagger. \quad (34)$$

This basically ends the description of the identification algorithm but before we end this section an important equation is the uncertainty of the estimates and this is given by:

$$\left[\Delta\widehat{A},\ \Delta\widehat{B},\ \Delta\widehat{K}\right] = (\Delta\widehat{\overline{X}} - A\Delta\underline{\hat{X}} - K\Delta\underline{\hat{E}}) \begin{bmatrix} \underline{\widehat{X}} \\ \underline{U} \\ \underline{\widehat{E}} \end{bmatrix}^\dagger, \quad (35)$$

$$\left[\widehat{\Delta C}\right] = (-C\Delta\widehat{X} - \Delta\hat{E})\widehat{X}^\dagger. \quad (36)$$

4. ASYMPTOTIC VARIANCE

In the previous section we gave a rather straightforward explanation of the identification algorithm. At some points we already looked at the effect of the noise on the estimates. In this section we will continue the analysis to come up with an expression for the asymptotic variance.

4.1 Estimate of $\Delta\widehat{X}$

In (29) we already found an expression for the estimated state. Using (27) and (30) and the error on top of this state sequence if given by:

$$\Delta\widehat{X} = S\widehat{\Delta\Gamma^f\mathcal{K}^p}Z \tag{37}$$

Where S is a matrix of appropriate dimensions $S \in \mathbb{R}^{n\times\ell f}$. Note that in the previous section this matrix S is for instance given by $S = \left(W^{-1}\mathcal{U}\right)^\dagger$ but the result will hold for different full rank matrices S. Furthermore, we define the following matrices (for $f = p$):

$$Q = \begin{bmatrix} Z_{z_1}^\dagger & Z_{z_2}^\dagger & \cdots & Z_{z_p}^\dagger \\ 0 & Z_{z_1}^\dagger & \cdots & Z_{z_{p-1}}^\dagger \\ & & \ddots & \vdots \\ 0 & & & Z_{z_1}^\dagger \end{bmatrix} Z. \tag{38}$$

In a similar way the matrix $\overline{Q}$ and $\underline{Q}$ are defined using $\overline{Z}$ and $\underline{Z}$ at the end of the expression, respectively. So, now we can write:

$$\Delta\widehat{X} = S\underbrace{(I_f \otimes E)}_{E_{I_f}}Q, \tag{39}$$

where $\otimes$ represent the Kronecker product and I_f and identity matrix of the size f. Furthermore, you can proof that $\Delta\hat{E} = E\Pi_Z$ and we define $\underline{\hat{E}} = EZ^\dagger\underline{Z}$. With these observations and definitions we can rewrite (35)-(36) as:

$$\left[\Delta\widehat{A},\ \Delta\widehat{B},\ \Delta\widehat{K}\right] = (SE_{I_f}\overline{Q} - ASE_{I_f}\underline{Q} - KEZ^\dagger\underline{Z}) \begin{bmatrix}\widehat{X}\\ \underline{U}\\ \underline{\widehat{E}}\end{bmatrix}^\dagger,$$

$$\left[\Delta\widehat{C}\right] = (-CSE_{I_f}Q - EZ^\dagger Z)\widehat{X}^\dagger.$$

This equation is the starting point for further analysis. We start by simply computing the asymptotic variance.

4.2 Asymptotic variance

The asymptotic variance was already defined in (5) and with the previous expression we now have:

$$\begin{aligned}\mathrm{vec}(\left[\Delta\widehat{A},\ \Delta\widehat{B},\ \Delta\widehat{K}\right]) &= \alpha_1\mathrm{vec}(\mathrm{E_{I_f}}) + \beta_1\mathrm{vec}\,(\mathrm{E})\,,\\ \mathrm{vec}(\left[\Delta\widehat{C}\right]) &= \alpha_2\mathrm{vec}(\mathrm{E_{I_f}}) + \beta_2\mathrm{vec}\,(\mathrm{E})\,.\end{aligned}$$

The matrices α and β can be computed (but are very large) and can be found in Appendix B.

Now we define a matrix P such that $\mathrm{vec}(\mathrm{E_{I_f}}) = \mathrm{P}\mathrm{vec}(\mathrm{E})$. With this definition we have:

$$\mathrm{vec}(\left[\Delta\widehat{A},\ \Delta\widehat{B},\ \Delta\widehat{K},\ \Delta\widehat{C}\right]) = \begin{bmatrix}\alpha_1P + \beta_1\\ \alpha_2P + \beta_2\end{bmatrix}\mathrm{vec}(\mathrm{E}). \tag{40}$$

The asymptotic variance in now given by:

$$\mathrm{plim}\Delta\Theta\Delta\Theta^* = \begin{bmatrix}\alpha_1P + \beta_1\\ \alpha_2P + \beta_2\end{bmatrix}(I\otimes\Lambda)\begin{bmatrix}\alpha_1P + \beta_1\\ \alpha_2P + \beta_2\end{bmatrix}^T. \tag{41}$$

This expression depends on the true system matrices and covariance of the innovation sequence. Under the consistency property these quantities can be replaced by their estimates.

The matrix P if of the size $(N-p)p^2\times(N-p)$ and can give rise to memory issues while computing. In Appendix C the structure in P is exploited to derive manageable expressions

5. EXAMPLE

5.1 Example I

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

$$\begin{aligned}x_{k+1} &= Ax_k + Bu_k + Ke_k,\\ y_k &= Cx_k + e_k,\\ u_k &= F_by_k + r_k,\end{aligned}$$

where F_b is the feedback gain and r_k the reference signal. The system matrices are given in Table 5.1. As reference signal we take a zero-mean Gaussian white noise signal with $cov(r_k) = I_r$ and for e_k we take a Gaussian white noise with the following covariance $cov(e_k) = I_\ell$. For the identification experiment we used $N = 2000$ and $p = f = 100$ (this corresponds with $||\tilde{A}^p|| = 1\times10^{-23}$. The collected data ($u_k$ and y_k) is used to identify an LTI model and the corresponding asymptotic variance. The performance of the identified system is evaluated by looking at the bode magnitude plot.

To investigate the sensitivity of the identification algorithm with respect to noise, a Monte-Carlo simulation with 20 runs was carried out. For each of the 20 simulations a different realization of the input and noise is used. For the first identification run the asymptotic variance is computed and the 99% confidence bound is plotted (see Fig 1). One can observe that for this particular example the bounds give realistic values. We also compared the performance of the suggested approach with the existing PE implementations in Matlab (Ljung's toolbox) and similar results are obtained. However, here we have to say that this implementation uses a subspace method to obtain an initial estimate before they start a nonconvex optimization.

5.2 Example II

In this second example we tested the proposed scheme on a challenging 19^{th}-order SISO closed-loop model. The system matrices are given in Verhaegen and Verdult [2007] and are derived from an accoustic duct. As input signal we take a zero-mean Gaussian white noise signal with $cov(r_k) = I_r$ and for e_k we take a Gaussian white noise with the following covariance $cov(e_k) = I_\ell$, which corresponds with a signal-to-noise ratio of approximately 8 dB. For the identification experiment we used $N = 2000$ and $p = f = 100$. The collected data (u_k and y_k) is used to identify an LTI model and the corresponding asymptotic variance. The performance of the identified system is evaluated by looking at the bode plot. As clearly can be seen in Fig. 2 the proposed method is also suitable to work with a realistic complex model.

$$A = \begin{bmatrix} 0.67 & 0.67 & 0 & 0 \\ -0.67 & 0.67 & 0 & 0 \\ 0 & 0 & -0.67 & -0.67 \\ 0 & 0 & 0.67 & -0.67 \end{bmatrix}, \quad B = \begin{bmatrix} 0.6598 & 1.9698 \\ -0.5256 & 0.4845 \\ -0.6968 & 0.1722 \\ 0.1474 & 0.5646 \end{bmatrix}, \quad K = \begin{bmatrix} -0.6968 & -0.1474 \\ 0.1722 & 0.5646 \\ 0.6484 & -0.4660 \\ -0.9400 & 0.1032 \end{bmatrix},$$

$$C = \begin{bmatrix} 0.3749 & 0.0751 & -0.5225 & 0.5830 \\ -0.8977 & 0.7543 & 0.1159 & 0.0982 \end{bmatrix}, \quad F_b = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}.$$

Table 1. Simulation parameters

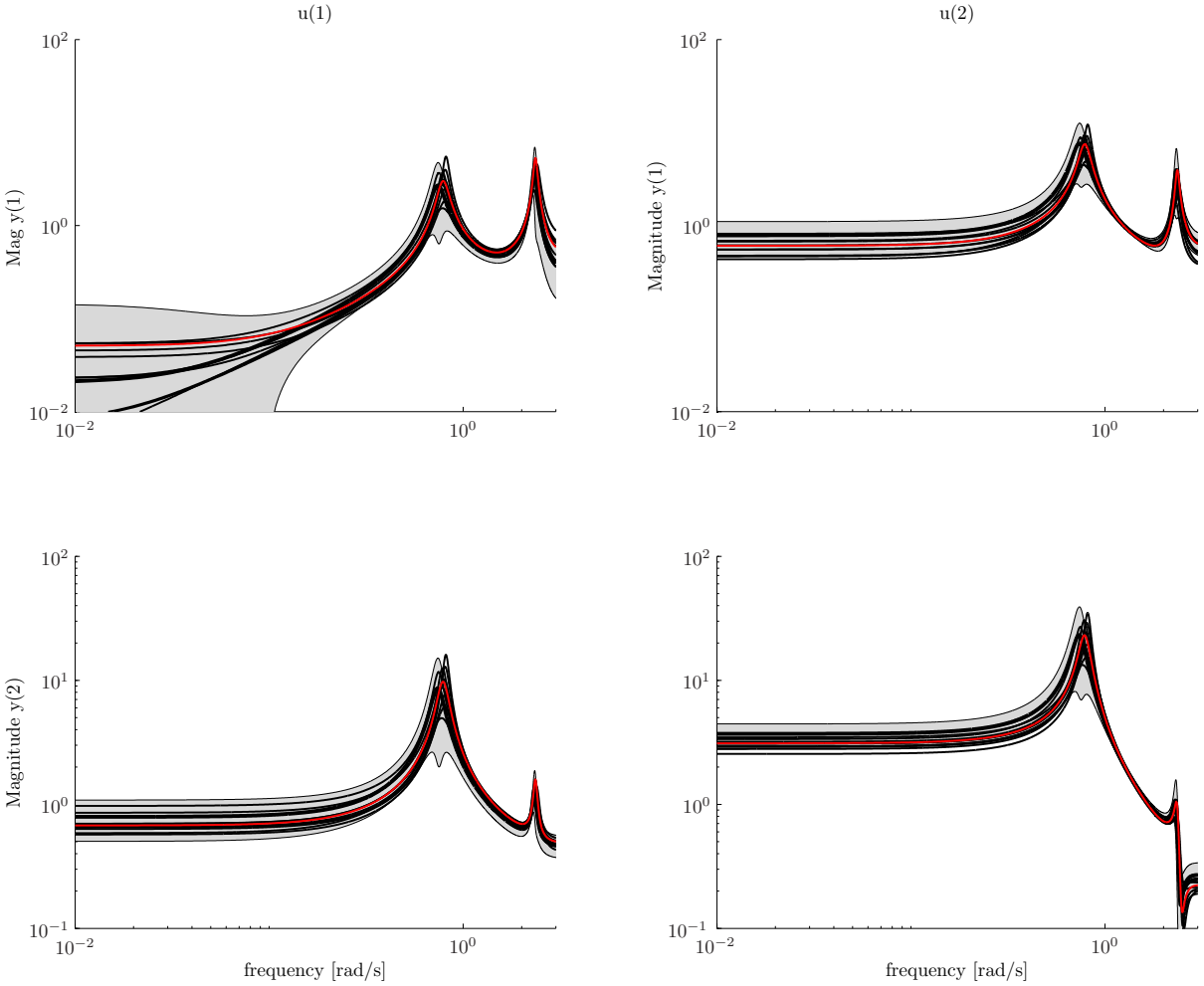

Fig. 1. In this figure the real system is represented by means of its Bode magnitude plot (red line) and the black lines are the identified models based on the 20 Monte Carlo simulations. On top of one of estimates the 99% confidence bounds is give by a gray shade.

6. CONCLUSION

Under the assumption that $\tilde{A}^p = 0$ and the system is within the model set we derived an expression for the asymptotic variance of a predictor based subspace identification method called $\text{PBSID}_{\text{opt}}$ which is a dedicated closed-loop identification algorithm that mixes the advantages properties of traditional PE and SMI methods. We derived manageable expression and we stayed close to the notation used in the traditional subspace papers. We showed the effectiveness of the approach using two simulations examples.

Appendix A

For having a consistent estimate the following two equations have to hold:

$$\lim_{N\to\infty} (\Delta\overline{\widehat{X}} - A\Delta\underline{\hat{X}} - K\Delta\underline{\hat{E}}) \begin{bmatrix} \widehat{X} \\ \underline{U} \\ \underline{\hat{E}} \end{bmatrix}^{\dagger} = 0, \qquad \text{(A.1)}$$

$$\lim_{N\to\infty} (-C\Delta\widehat{X} - \Delta\hat{E})\widehat{X}^{\dagger} = 0, \qquad \text{(A.2)}$$

this is true under the condition that $||\tilde{A}^p|| = 0$ and we have the true model order (so no undermodeling).

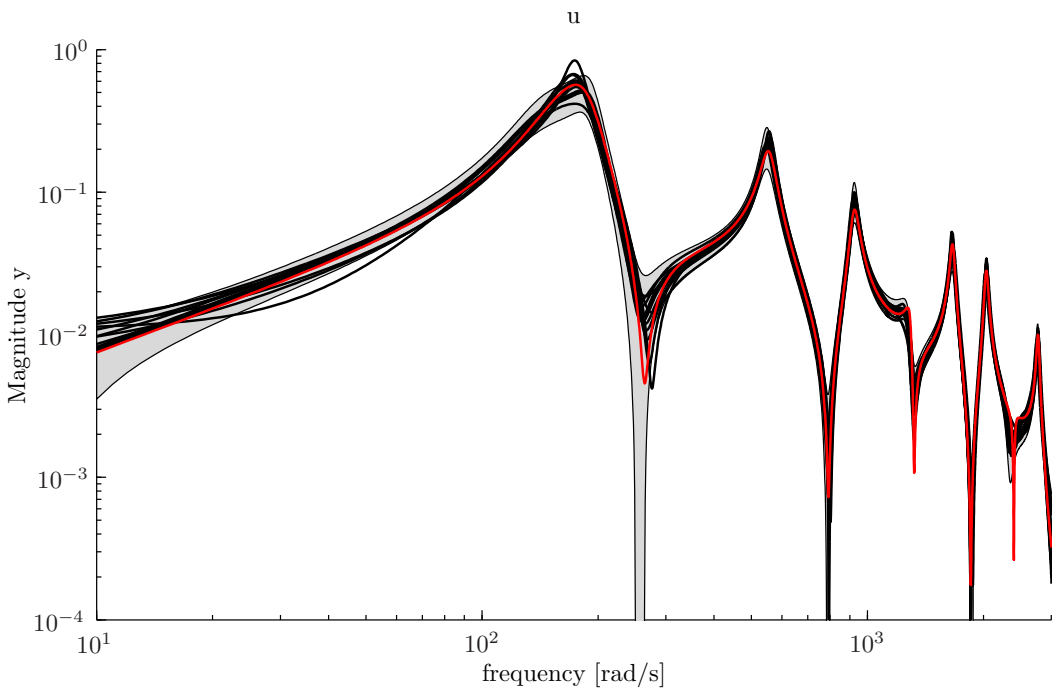

Fig. 2. In this figure the real system is represented by means of its Bode magnitude plot (red line) and the black lines are the identified models based on the 20 Monte Carlo simulations. On top of one of estimates the 99% confidence bounds is give by a gray shade.

Appendix B.

We have the following matrices where we already assume that the unknown true matrices are replaced by their estimated equivalent:

$$\alpha_1 = \left(\overline{Q}\begin{bmatrix}\widehat{\underline{X}}\\ \underline{U}\\ \widehat{\underline{E}}\end{bmatrix}^{\dagger}\right)^T \otimes S - \left(\underline{Q}\begin{bmatrix}\widehat{\underline{X}}\\ \underline{U}\\ \widehat{\underline{E}}\end{bmatrix}^{\dagger}\right)^T \otimes (\hat{A})S \quad \text{(B.1)}$$

$$\beta_1 = -\left(Z^{\dagger}\underline{Z}\begin{bmatrix}\widehat{\underline{X}}\\ \underline{U}\\ \widehat{\underline{E}}\end{bmatrix}^{\dagger}\right)^T \otimes (\hat{K}) \quad \text{(B.2)}$$

$$\alpha_2 = \left(Q\left[\widehat{X}\right]^{\dagger}\right)^T \otimes (-\hat{C})S \quad \text{(B.3)}$$

$$\beta_2 = -\left(\Pi_Z\left[\widehat{X}\right]^{\dagger}\right)^T \otimes I \quad \text{(B.4)}$$

Appendix C.

We define the following matrix

$$[S_1\ S_2\ \cdots\ S_f] = S \quad \text{(C.1)}$$

$$[Q_1^T\ Q_2^T\ \cdots\ Q_f^T] = Q^T \quad \text{(C.2)}$$

with $S_i \in \mathbb{R}^{n\times l}$ and $Q_i \in \mathbb{R}^{N-p\times N-p}$. Similar definitions hold for $\overline{Q}_i$ and $\underline{Q}_i$. We have the following matrices:

$$\alpha_1 P = \sum_{i=1}^{f}\left(\left(\overline{Q}_i\begin{bmatrix}\widehat{\underline{X}}\\ \underline{U}\\ \widehat{\underline{E}}\end{bmatrix}^{\dagger}\right)^T \otimes S_i - \left(\underline{Q}_i\begin{bmatrix}\widehat{\underline{X}}\\ \underline{U}\\ \widehat{\underline{E}}\end{bmatrix}^{\dagger}\right)^T \otimes (\hat{A})S_i\right)$$

$$\alpha_2 P = \sum_{i=1}^{f}\left(\left(Q_i\left[\widehat{X}\right]^{\dagger}\right)^T \otimes (-\hat{C})S_i\right)$$

This equation basically implies that the biggest matrix you have to store is of the size $N-p\times N-p$.

REFERENCES

A. Chiuso. Asymptotic variance of closed-loop subspace identification methods. *IEEE Transactions on Automatic Control*, 51(8):1299–13–14, 2006.

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

G. H. Golub and C. F. van Loan. *Matrix computations*. The John Hopkins University Press, 1996.

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

L. Ljung. *System identification; Theory for the user*. Prentice Hall, 1987.

S. Qin and L. Ljung. Closed-loop subspace identfication with innovation estimate. In *In proceedings of IFAC system identification conference (SYSID)*, Rotterdam, The Netherlands, 2003.

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

J.W. van Wingerden, A.W. Hulskamp, T. Barlas, I. Houtzager, H.E.N. Bersee, G. A. M Van Kuik, and

M. Verhaegen. Two-degree-of-freedom active vibration control of a prototyped 'smart' rotor. *to appear in IEEE Transactions on Control System Technologies*, -:–, 2010.

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

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