Subspace Identification using predictor estimation via Gaussian regression

A. Chiuso, G. Pillonetto, G. De Nicolao

Abstract— In this paper we propose a new nonparametric approach to identification of linear time invariant systems using subspace methods. The nonparametric paradigm to prediction of stationary stochastic processes, developed in a companion paper, is integrated into a recently proposed subspace method. Simulation results show that this approach significantly improves over standard subspace methods when using small sample sizes. In particular, the new approach facilitates significantly the order selection step.

Index Terms— Subspace Methods; kernel-based methods; Bayesian estimation;regularization; Gaussian processes

I. INTRODUCTION

Subspace methods for identification of Multi-Input Multi-Output (MIMO) linear time invariant systems have received a considerable attention in the last decades. While many subspace algorithms have been developed, and still the picture is not completely clear as to relative performance, it has been recently shown [12], [14], [8], [9], [6] that a whole class of subspace methods are based on preliminary estimation of a linear predictor. Among these methods, which rely on the so called "state sequence" approach [2], it is fair to say that the PBSID_{opt} algorithm, introduced in [7] and further studied in [8], plays a special role. In fact, it has been shown in [7], that this method is optimal within a class of methods using the so called "CCA" weight. Moreover, as demonstrated in [8], it can be implemented with a lower computational complexity than most "classical" methods; this algorithm hinges on the estimation of a Vector AutoRegressive with eXogenous inputs (VARX) model; using the coefficients of this VARX model, the state is recovered using the standard circle of ideas from stochastic realization theory [16], [17], [10] subspace methods are based on.

While the asymptotic properties of these methods are by now fairly well understood, see [12], [2], [11], [8], [7], less clear is the situation when the sample size is small. In particular, since the VARX models estimated in a first step may require a large number of coefficients (see e.g. [12], [2], [8]), some care needs to be taken for small sample size.

In our opinion, it is quite natural to ask whether regularization techniques [26], [5], which have played an important role in statistics to deal with either ill-conditioned problems or model estimation with small sample sizes, could help also in the area of subspace identification.

A. Chiuso is with the Department of Management and Engineering, University of Padova, Vicenza, Italy, `chiuso@dei.unipd.it`

G. Pillonetto is with the Department of Information Engineering, University of Padova, Italy `giapi@dei.unipd.it`

G. De Nicolao is with the Dipartimento di Informatica e Sistemistica, University of Pavia, Italy `giuseppe.denicolao@unipv.it`

In a companion paper [22], we have proposed estimation methodologies based on Gaussian regression [24], [19], [21], first introduced in [20], to the purpose of constructing predictors for stationary stochastic processes. These models, see formula (2), can be thought of as non-parametric extensions of infinite order auto-regressions and, as such, are prone to a variety of applications.

In this paper we shall discuss the use of these models in subspace methods. In particular the predictor based on a VARX model, estimated in the first step of the PBSID_{opt} algorithm, is replaced by the nonparametric predictor. We shall report simulation results in a number of examples including a system operating in closed loop.

The simulation results suggests that, for small sample sizes, the parametric predictor outperforms the standard subspace methods in most cases.

The structure of the paper is as follows: Section II summarizes the main facts about the nonparametric predictor. In Section III we describe the PBSID_{opt} method, as well as we quote from the literature two order estimation criteria. Section IV shows how the nonparametric predictor is used in the PBSID_{opt} algorithm, while Section V contains the simulation results. Some conclusions are drawn in Section VI.

II. PREDICTOR IDENTIFICATION VIA GAUSSIAN REGRESSION

In a companion paper [22] we have seen that, given two scalar stationary stochastic processes $\{y_t\}$, $\{u_t\}$, the predictor

$$\hat{y}_t = \sum_{k=1}^{\infty} F_k y_{t-k} + \sum_{k=1}^{\infty} G_k u_{t-k} \tag{1}$$

can be conveniently estimated via regularization techniques by imposing that the impulse responses $\{F_k\}$ and $\{G_k\}$ live in suitable infinite dimensional spaces. In particular, being ζ a finite dimensional parameter to be estimated from data, the predictor is taken to be of the form

$$\hat{y}_t(\zeta) = \sum_{k=1}^{\infty} F_k(\zeta) y_{t-k} + \sum_{k=1}^{\infty} G_k(\zeta) u_{t-k} \tag{2}$$

$$F_t(\zeta) = \sum_{k=1}^{\infty} a_k(\zeta) f_{t-k} \qquad G_t(\zeta) = \sum_{k=1}^{\infty} b_k(\zeta) g_{t-k} \tag{3}$$

where the infinite sequences $\{f_t\}$ and $\{g_t\}$ are modeled as zero mean Gaussian processes of given covariance and $a(\zeta)$ and $b(\zeta)$ represent finite-dimensional components of the model. The choice of the covariance of the Gaussian process, the structure of $a(\zeta)$ and $b(\zeta)$, the estimation of the

sequences $\{f_t\}$ and $\{g_t\}$ and the vector of parameters ζ given two finite sequence of data points $\{y_1,...,y_N\}$, $\{u_1,...,u_N\}$ is discussed in the companion paper [22].

In [22] we have also shown that this methodology outperforms standard PEM as far the predictive performance of the estimated models is concerned.

This approach can quite naturally be extended to prediction of multivariate time series. We shall not enter into the details, which are not relevant to this paper, but we need at least to set up notation for future use.

Let $y_t \in \mathbb{R}^p$ and $u_t \in \mathbb{R}^m$; we shall write the model (2), in the form:

$$\hat{y}_t(\xi) = \sum_{k=1}^{\infty} \psi_k(\xi) y_{t-k} + \sum_{k=1}^{\infty} \phi_k(\xi) u_{t-k} \qquad (4)$$

for matrix coefficients $\psi_k(\xi)$ and $\phi_k(\xi)$ of suitable dimension. To be more specific, let us introduce the following notation: given a matrix A we shall denote by $[A]_i$ the $i-th$ row of A and by $[A]_{ij}$ the element in position i,j. The $i-th$ row, $i = 1,..,p$, of (4) can be written as

$$[\hat{y}_t(\xi)]_i \quad = \quad \sum_{k=1}^{\infty} [\psi_k(\xi)]_i y_{t-k} + \sum_{k=1}^{\infty} [\phi_k(\xi)]_i u_{t-k} \qquad (5)$$

where

$$[\psi_t(\xi)]_{ij} = \sum_{k=1}^{\infty} a_k^{ij}(\xi)[f_{t-k}]_{ij} \qquad [\phi_t(\xi)]_{i\ell} = \sum_{k=1}^{\infty} b_k^{i\ell}(\xi)[g_{t-k}]_{i\ell}$$

for $j = 1,..,p$ and $\ell = 1,..,m$.

The processes f_k and g_k are independent, matrix valued zero-mean Gaussian processes with suitable covariance functions. A simple and reasonable choice is to assume that $\text{vec}(f_k)$ and $\text{vec}(g_k)$ have independent components each of them having covariance function of the form $\lambda^2 K(\cdot,\cdot;\beta)$ as in [22] where the hyperparameters λ and β are possibly different for each component.

Several simplifications are possible to the purpose of reducing the number of hyperparameters ξ. For instance it is reasonable to assume that $a^{ij} = b^{i\ell}$ for all i, j, ℓ. Of course this is not necessary and distinct parametric parts could be estimated for each output component and for each component of the regressors $[y_k]_j$ and $[u_k]_\ell$.

III. SUBSPACE IDENTIFICATION BASED ON PREDICTOR IDENTIFICATION: THE PBSID$_{opt}$ ALGORITHM

The purpose here is to estimate the parameters (A,B,C,D,K) of a state space model

$$\begin{cases} x_{t+1} &= Ax_t + Bu_t + Ke_t \\ y_t &= Cx_t + e_t \end{cases} \qquad (6)$$

starting from a finite set of data points $\{y_1,..,y_{\bar{N}}\}$, $\{u_1,..,u_{\bar{N}}\}$.

Since this algorithm is designed to be consistent also in the presence of feedback, $D = 0$ is imposed for identifiability reasons. Of course the algorithm can be modified if D is to be estimated, provided there is knowledge that a delay is present somewhere else in the loop.

We shall also denote with $\Lambda := E[e(t)e^\top(t)]$ the variance of the innovation, $F(z) := C(zI-A)^{-1}B$ the transfer function from u to y while $W(z) := C(zI-A)^{-1}K\Lambda^{1/2} + \Lambda^{1/2}$.

In subspace identification the two integers p and f (respectively "past" and "future") are the number of block rows in certain block Hankel data matrices we shall encounter later on. The choice of these parameters is not an entirely trivial fact, and in fact is subject of current research [3], [4], [2], [23], [8], [9], [6].

Uppercase letters (e.g. Y_t^M) denote tails of length M formed with data points starting at time t, i.e.

$$Y_t^M := [y_t, y_{t+1}, .., y_{t+M-1}].$$

Most tails which we shall use are of length $N := \bar{N} - p - f$. When $M = N$ we shall drop the superscript M, i.e. $Y_t := Y_t^N$. The symbol $Y_{[t_1,t_2]}$ denotes a block Hankel matrix as follows:

$$Y_{[t_1,t_2]} := \begin{bmatrix} Y_{t_1} \\ Y_{t_1+1} \\ \vdots \\ Y_{t_2} \end{bmatrix} = \begin{bmatrix} y_{t_1} & y_{t_1+1} & \cdots & y_{t_1+N-1} \\ y_{t_1+1} & y_{t_1+2} & \cdots & y_{t_1+N} \\ \vdots & \vdots & \cdots & \vdots \\ y_{t_2} & y_{t_2+1} & \cdots & y_{t_2+N-1} \end{bmatrix}.$$

We shall define the predictor matrix $\bar{A} := A - KC$ and the observability matrix $\bar{\Gamma}_{f-1} := [C^\top, \bar{A}^\top C^\top, ..., (\bar{A}^{f-1})^\top C^\top]^\top$.

First of all we recall that reader that subspace methods for identification of linear time-invariant (LTI) systems in state space form can be thought of as two-step procedures based on the following:

1) Estimate bases for the state space X_p and X_{p+1}. These will be matrices with n rows (the number of state components, which will be determined as part of the estimation procedure, see subsection III-A) and N columns (N is of the order of the number of data available)

2) From the estimated state sequences X_p and X_{p+1}, estimate the system matrices A, B, K, C by solving

$$\begin{cases} \hat{X}_{p+1} &\simeq A\hat{X}_p + BU_p + K\hat{E}_p \\ Y_p &\simeq C\hat{X}_p \end{cases} \qquad (7)$$

in the least squares sense, where $\hat{E}_p := Y_p - \hat{C}\hat{X}_p$.

The first step in the optimized version of PBSID introduced in [7] (PBSID$_{opt}$) is based on the estimation of the coefficients of the VARX model

$$y_t = \sum_{i=1}^{\infty} \Xi_i z_{t-i} + e_t \qquad (8)$$

where $\Xi_i = [\psi_i \ \phi_i]$ and $z_k := [y_k^\top \ u_k^\top]^\top$. These coefficients are estimated in the least squares sense, i.e. solving[1]

$$\arg \min_{\Phi_i} \|Y_p^{f-1+N} - \sum_{i=1}^{p} \Xi_i Z_{p-i}^{f-1+N}\|_F; \qquad (9)$$

this may include estimation of the appropriate p using standard criteria (e.g AIC) for VARX order estimation.

---

[1]The subscript $_F$ denotes Frobenius norm.

Let us now define the multi-step predictors

$$\hat{Y}_{[p,p+f)} := \begin{bmatrix} \hat{\Xi}_p & \dots & \hat{\Xi}_{p+f-1} & \dots & \hat{\Xi}_1 \\ 0 & \hat{\Xi}_p & \dots & \dots & \hat{\Xi}_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \hat{\Xi}_p & \dots & \hat{\Xi}_f \end{bmatrix} Z_{[0,p)} \quad (10)$$

and

$$\hat{Y}_{[p+1,p+f+1)} := \begin{bmatrix} \hat{\Xi}_p & \dots & \hat{\Xi}_{p+f-1} & \dots & \hat{\Xi}_1 \\ 0 & \hat{\Xi}_p & \dots & \dots & \hat{\Xi}_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \hat{\Xi}_p & \dots & \hat{\Xi}_f \end{bmatrix} Z_{[1,p+1)} \quad (11)$$

The observability matrix and the state sequences are then estimated as follows:

- Given the weighting matrix[2] W_p, compute the Singular Value Decomposition (SVD)

$$W_p^{-1} \hat{Y}_{[p,p+f)} = PDQ^\top = [P_n \tilde{P}_n] \begin{bmatrix} D_n & 0 \\ 0 & \tilde{D}_n \end{bmatrix} \begin{bmatrix} Q_n^\top \\ \tilde{Q}_n^\top \end{bmatrix} \quad (12)$$

- Form an estimate of the observability matrix $\bar{\Gamma}_{f-1}$ discarding the "less significant" singular values (i.e. pretending $\tilde{D}_n \simeq 0$) from

$$\hat{\bar{\Gamma}}_{f-1} = W_p P_n D_n^{1/2}.$$

- Estimate a basis for the state space from

$$\hat{X}_p := \hat{\bar{\Gamma}}_{f-1}^{-L} \hat{Y}_{[p,p+f)} \quad (13)$$

and

$$\hat{X}_{p+1} := \hat{\bar{\Gamma}}_{f-1}^{-L} \hat{Y}_{[p+1,p+f+1)} \quad (14)$$

- An estimator of the innovation sequence E_p can be found by

$$\hat{E}_p := Y_p - Y_p \hat{X}_p^\top \left(\hat{X}_p \hat{X}_p^\top \right)^{-1} \hat{X}_p. \quad (15)$$

A. Order Estimation

Estimating a model from data includes selecting the complexity of the model. When considering linear state space models, the complexity is given by the state space dimension, denoted in this paper by n. In particular, when using subspace methods, order estimation amounts to deciding a suitable value of n such that in the SVD (12) one can approximate $\tilde{D}_n \simeq 0$. Similarly to standard criteria for model order estimation such as AIC, MDL, BIC (see e.g. [18], [25]) also in subspace methods one trades between "goodness of fit" (measured by a suitable criterion) and the model complexity $d(n)$, measured by the number of estimated parameters. This trade off, of course, depends upon the number of data available N.

To date, there is no established procedure for estimating the order of subspace methods. According to the recent surveys [1], [2], we considered the following two criteria:

1) Singular Value Criterion (SVC)

$$SVC(n) := \sigma_{n+1}^2 + \frac{\log(N)}{N} d(n) \quad (16)$$

where σ_i^2 is the $i-th$ singular value in the SVD step, i.e. the $i-th$ eigenvalue of the matrix D;

2) Innovation Variance Criterion (IVC)

$$IVC(n) := \log(\det(\hat{\Lambda}_n)) + \frac{\log(N)}{N} d(n) \quad (17)$$

where $\hat{\Lambda}_n$ is the variance of the prediction error using the $n-th$ order model.

The order is estimated by computing the value of n which minimizes either SVC or IVC. The simulation results included in the paper have been obtained using SVC since we observed that IVC tended to overestimate the order. Of course, this might simply depend upon the specific example and it is outside the scope of this paper to comment further on this issue.

IV. USING THE NON-PARAMETRIC PREDICTOR IN SUBSPACE IDENTIFICATION

In this paper we propose using the non-parametric predictors introduced in Section II in conjunction with the PBSID$_{opt}$ algorithm in Section III. In fact, as discussed above, the first step in PBSID$_{opt}$ consists in estimating a linear predictor of y_t, see equations (8) and (9). As discussed in the papers [2], [7], [8], the length p of this VARX model needs to be large, so that $(A-KC)^p$ becomes negligible. In practice a suitable value of p can be chosen by standard order estimation methods such as AIC.

It can be shown that (see [12], Remark 3.6, [15], page 93, end of Section 2 and [13]) if $\hat{p}$ is the minimizer of the AIC criterion, $p = M\hat{p}$, with $M > 1$, is a possible choice.

Unfortunately having $(A-KC)^p$ small might be difficult to achieve, especially when either the eigenvalues of $A-KC$ are close to the unit circle or the number of data available N is small.

In order to circumvent the difficulties in estimating these predictors, for small sample sizes, we suggest substituting the estimated coefficients $\hat{\Xi}_i$ in (9) with the nonparametric predictor coefficients[3] $\hat{\Xi}_i^{NP} := [\hat{\psi}_i(\xi) \; \hat{\phi}_k(\xi)]$. These coefficients are then used in lieu of $\hat{\Xi}_i$ in formulas (10) and (11).

Unfortunately it seems rather hard to provide a statistical analysis of this procedure and therefore, in this paper, we shall only report simulation results in a number of interesting examples.

V. EXAMPLES

In order to test the proposed approach we consider the following three examples. Data have been generated according to the model
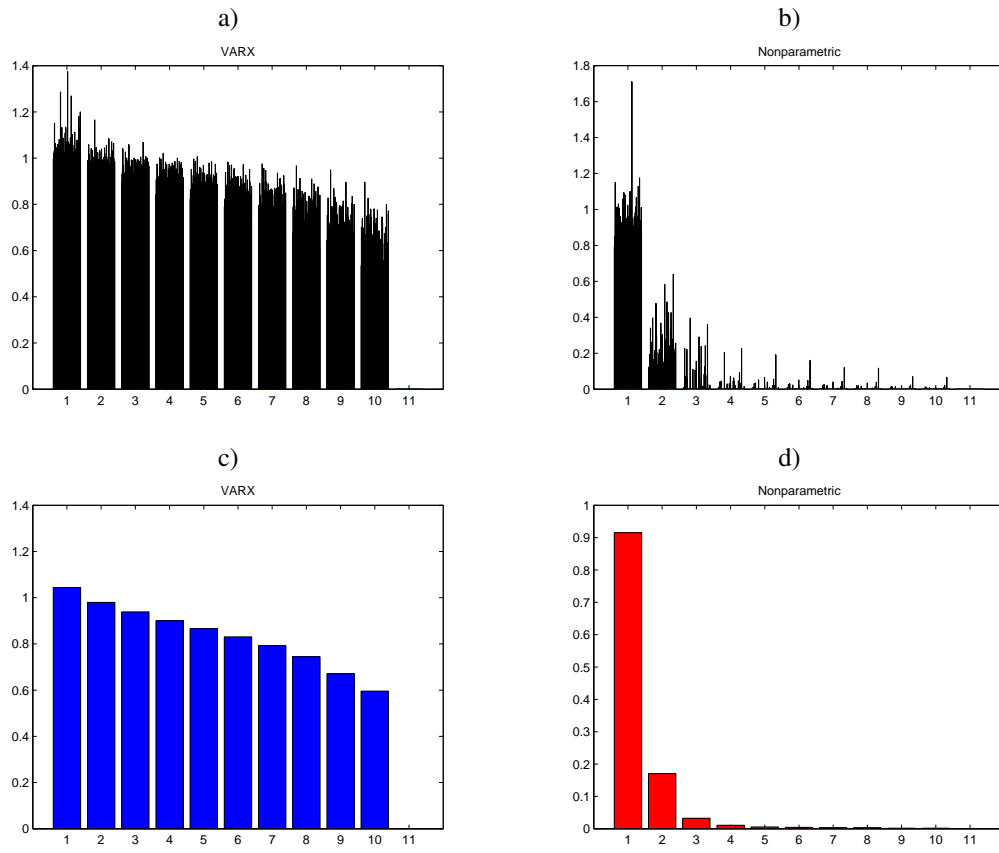
$$y_t = F(z)u_t + W(z)e_t$$

Fig. 1. System 1: singular values. a) and b): 100 Monte Carlo runs. c) and d) Average over 100 Monte Carlo runs.

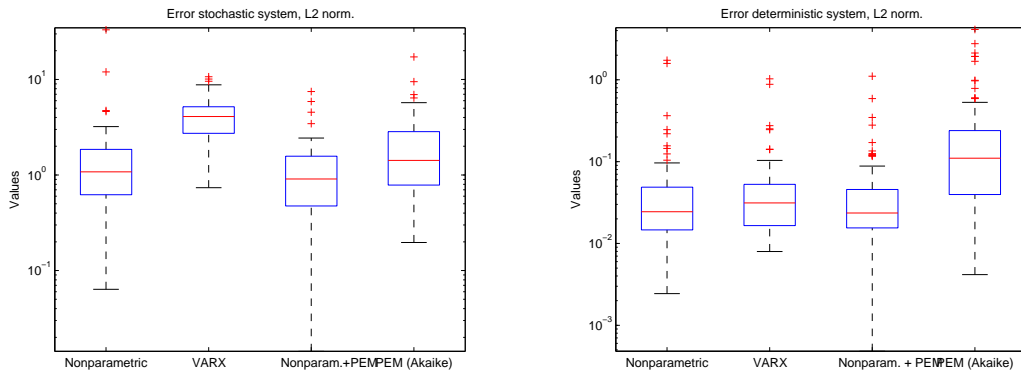Fig. 2. System 1: Mean square error (Average L_2 norm of the transfer function error).

	$F(z)$	$W(z)$
System 1	$\dfrac{0.5578z - 0.242}{z^2 - 0.7z - 0.18}$	$\dfrac{z^2 + 0.4z - 0.21}{z^2 - 0.7z - 0.18}$
System 2	$\dfrac{0.4802z - 1.351}{z^2 - 0.6z + 0.73}$	$\dfrac{z^2 + 0.6z - 0.27}{z^2 - 0.6z + 0.73}$
System 3	$\dfrac{1.067z^3 - 6.824z^2 - 1.39z - 0.8556}{z^4 - 1.1z^3 + 0.95z^2 - 0.523z - 0.153}$	$\dfrac{z^4 + 0.8z^3 + 0.8z^2 + 0.256z - 0.1785}{z^4 - 1.1z^3 + 0.95z^2 - 0.523z - 0.153}$

TABLE I

DETERMINISTIC ($F(z)$) AND STOCHASTIC ($W(z)$) PART OF THE MODELS.

The input u_t is either zero mean unit variance white noise (System 1 and 3) or obtained by a feedback mechanism (System 2)

$$u_t = -y_t + r_t$$

where r_t is zero mean unit variance white noise.

The details about the transfer functions $F(z)$ and $W(z)$ are reported in Table V.

The following identification methods have been compared:

(i) the proposed identification scheme (named *Nonparametric* in the plots). The lenght of the nonparametric predictor has been set to 30, which is the maximum value of p allowed for the VARX model in the PBSID$_{opt}$ method below. The future horizon has been set to 10. The order has been selected using the SVC criterion.

(ii) the PBSID$_{opt}$ method, which is based on the VARX model (named *VARX* in the plots). The length of the past horizon p used here has been selected using the AIC routine in Matlab. The maximum length of p has been set to 30. The future horizon has been set to 10. The order has been selected using the SVC criterion.

(iii) Matlab PEM initialized with the nonparametric model obtained using the proposed approach. The order is determined by the order of the model obtained in (i).

(iv) the "default" Matlab PEM, which uses the `n4sid.m` routine as initialization. AIC has been used to select the order in the range $[1 : 10]$.

All identification experiments have been performed with $N = 100$ for systems 1 and 2, while $N = 150$ data points have been used for system 3; 100 Monte Carlo runs have been performed for all the three experiments.

It is clear from the simulation results (see Figures 1, 3, 5, plots a) and c)) that, with such a small number of data, order estimation is difficult by inspecting the singular values. In fact, for the subspace methods based on the linear (VARX) predictor, a "jump" in the singular values plot is hardly seen. This is reflected by the fact that very rarely the SVC criterion has estimated the correct order.

Instead, rather interestingly, when using the nonparametric predictor there is a quite remarkable jump in the singular values. In fact, in this case the SVC has correctly estimated the order in all runs.

In order to evaluate the quality of the estimated model we have computed, for each Monte Carlo run, the L_2 norm of the estimation error for both the deterministic ($F(z)$) and the "stochastic" ($W(z)$) transfer functions. Figures 2, 4, 6 show the relative boxplots. It is clear that the subspace algorithm which used the nonparametric predictor performs much better in essentially all cases than the subspace algorithm based on VARX. It is also remarkable that not much improvement (if any) is gained running PEM initialized with the nonparametric subspace estimator; moreover PEM in conjunction with AIC performs much worse. Indeed, this latter method tends to strongly overestimate the order.

VI. CONCLUSIONS

In a companion paper we have proposed using nonparametric techniques based Gaussian regression for predictor identification. In this paper we suggested using these predictors to improve subspace methods. In particular we substituted the VARX models, used in the PBSID$_{opt}$ subspace algorithm [8], with the nonparametric predictor. Based on a number of simple examples we envision that the resulting algorithm will help improving standard subspace methods, in particular when using small sample sizes. Our future work will include a more systematic analysis of the proposed method; it will also be of interest to develop dedicated order estimation criteria.

VII. ACKNOWLEDGMENTS

REFERENCES

[1] D. Bauer, "Order estimation for subspace methods," *Automatica*, vol. 37, pp. 1561–1573, 2001.

[2] ——, "Asymptotic properties of subspace estimators," *Automatica*, vol. 41, pp. 359–376, 2005.

[3] D. Bauer and M. Jansson, "Analysis of the asymptotic properties of the MOESP type of subspace algorithms," *Automatica*, vol. 36, pp. 497–509, 2000.

[4] D. Bauer and L. Ljung, "Some facts about the choice of the weighting matrices in Larimore type of subspace algorithm," *Automatica*, vol. 38, pp. 763–773, 2002.

[5] M. Bertero, "Linear inverse and ill-posed problems," *Advances in Electronics and Electron Physics*, vol. 75, pp. 1–120, 1989.

[6] A. Chiuso, "Choosing the future horizon in closed-loop CCA-type subspace algorithms," University of Padova, Tech. Rep., 2007, sumbitted to IEEE Trans. on Aut. Control, available at: http://www.dei.unipd.it/~chiuso.

[7] ——, "On the relation between CCA and predictor-based subspace identification," *IEEE Trans. on Automatic Control*, vol. 52, no. 10, pp. 1795–1812, October 2007.

[8] ——, "The role of Vector AutoRegressive modeling in predictor based subspace identification," *Automatica*, vol. 43, no. 6, pp. 1034–1048, June 2007.

[9] ——, "Some insights on the choice of the future horizon in CCA-type subspace algorithms," in *Proc. of ACC*, 2007.

[10] A. Chiuso and G. Picci, *Geometry of Oblique Splitting, Minimality and Hankel Operators*, ser. Lect. Notes in Control and Information Sciences. Springer, 2003, no. 286, pp. 85–124.

[11] ——, "Consistency analysis of some closed-loop subspace identification methods," *Automatica*, vol. 41, no. 3, pp. 377–391, 2005.

[12] A. Dahlén and W. Scherrer, "The relation of CCA subspace method to a balanced reduction of an autoregressive model," *Journal of Econometrics*, vol. 118, no. 1-2, pp. 293–312, 2004.

[13] E. Hannan and M. Deistler, *The Statistical Theory of Linear Systems*. Wiley, 1988.

[14] M. Jansson and B. Wahlberg, "A linear regression approach to state-space subspace system identification," *Signal Processing*, vol. 52, pp. 103–129, 1996.

[15] G. Kuersteiner, "Automatic inference for infinite order vector autoregressions," *Econometric Theory*, vol. 21, pp. 85–115, 2005.

[16] A. Lindquist and G. Picci, "Realization theory for multivariate stationary gaussian processes," *SIAM J. on control and Optimiz.*, vol. 23, no. 6, pp. 809–857, 1985.

[17] ——, "Canonical correlation analysis, approximate covariance extension and identification of stationary time series," *Automatica*, vol. 32, pp. 709–733, 1996.

[18] L. Ljung, *System Identification - Theory For the User*. Prentice Hall, 1999.

[19] M. Neve, G. D. Nicolao, and L. Marchesi, "Nonparametric identification of population models via gaussian processes," *Automatica*, vol. 43, pp. 1134–1144, 2007.

[20] G. D. Nicolao and G. Pillonetto, "A new kernel-based approach for system identification," in *Proceedings of 2008 American Control Conf., Seattle, WA, USA*, 2008.

[21] G. Pillonetto and B. Bell, "Bayes and empirical bayes semi-blind deconvolution using eigenfunctions of a prior covariance," *Automatica*, 2007.

[22] G. Pillonetto, A. Chiuso, and G. De Nicolao, "Predictor estimation via Gaussian regression," in *Submitted to IEEE CDC 2008, Available at.....*, 2008.

[23] S. Qin and L. Ljung, "On the role of future horizon in closed-loop subspace identification," in *Proc. of SYSID 2006*, Newcastle, Australia, 2006.

[24] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[25] T. Soderstrom and P. Stoica, *System Identification*. Prentice Hall, 1989.

[26] A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems*. Washington, D.C.: Winston/Wiley, 1977.
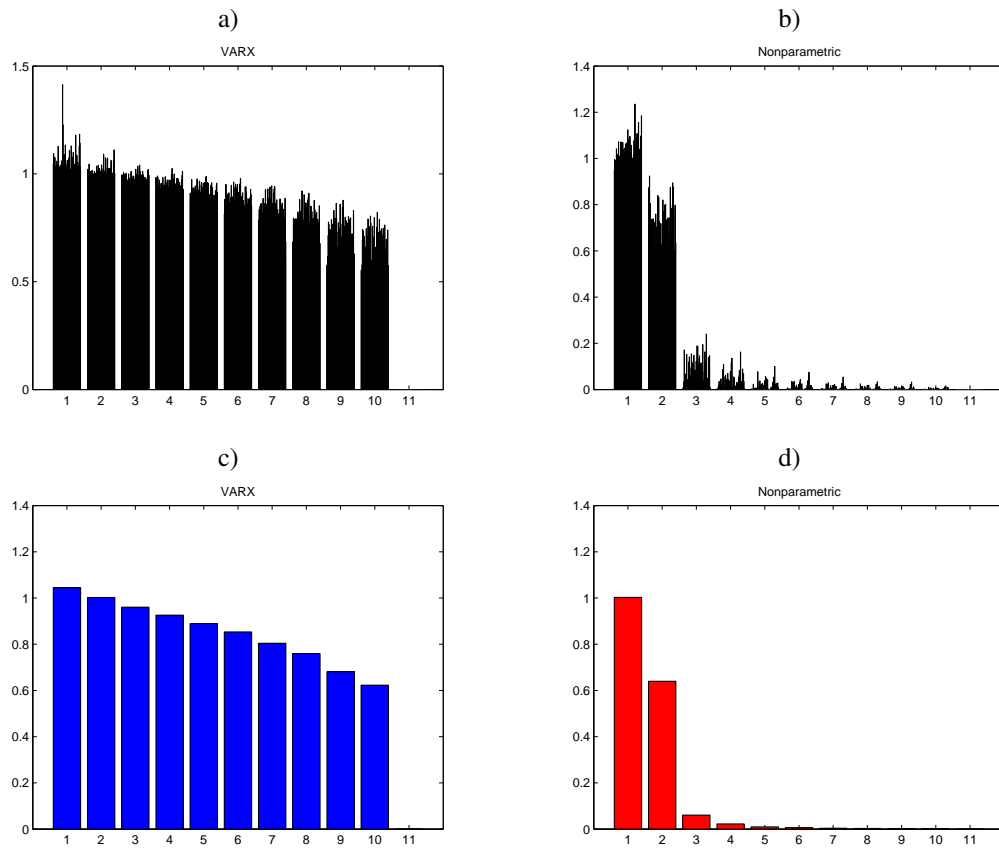
Fig. 3.   System 2: singular values. a) and b): 100 Monte Carlo runs. c) and d) Average over 100 Monte Carlo runs.
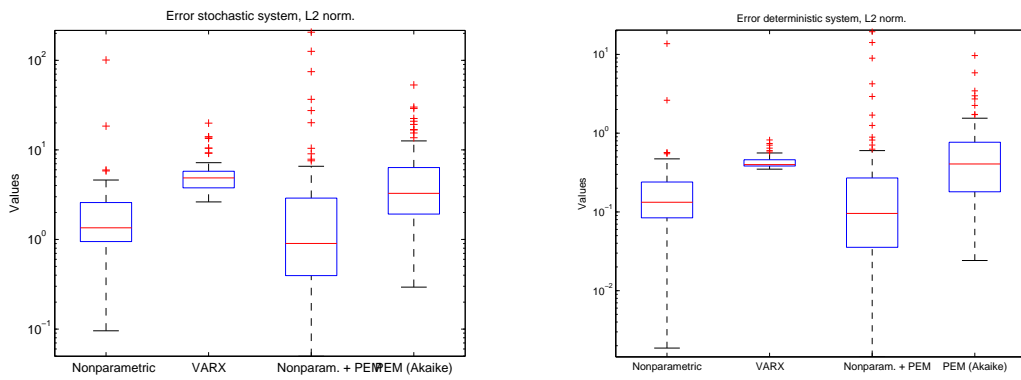

Fig. 4.   System 2: Mean square error (L_2 norm of the transfer function error).
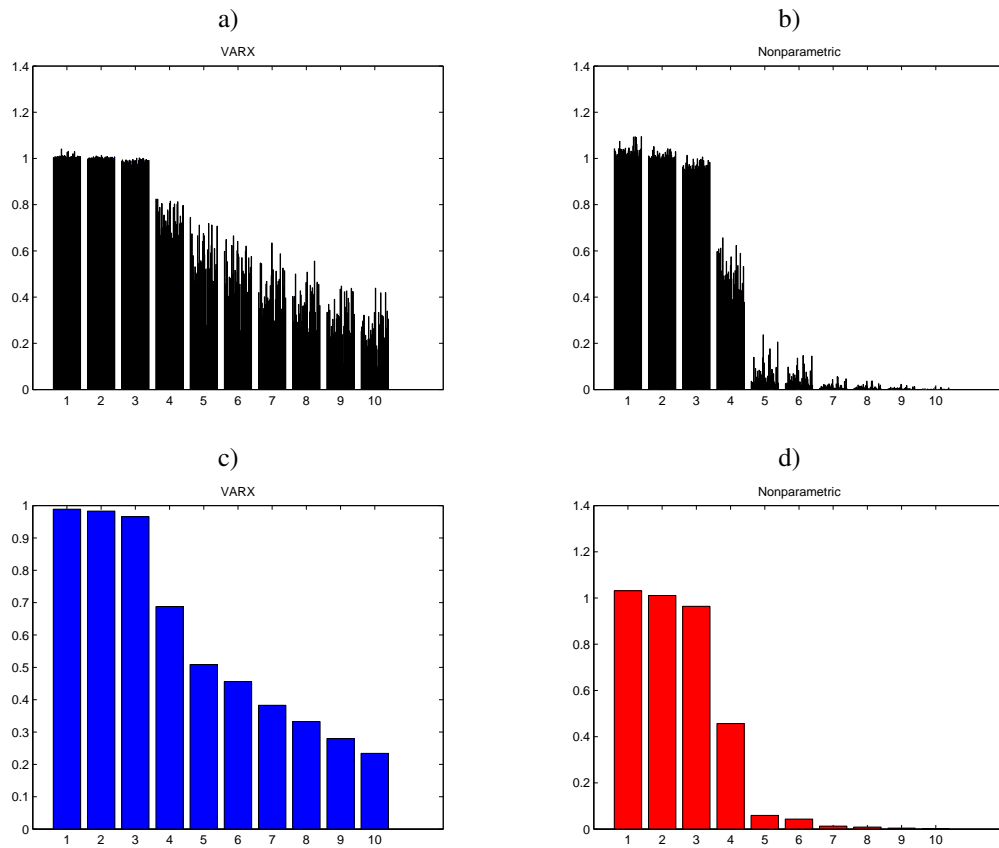
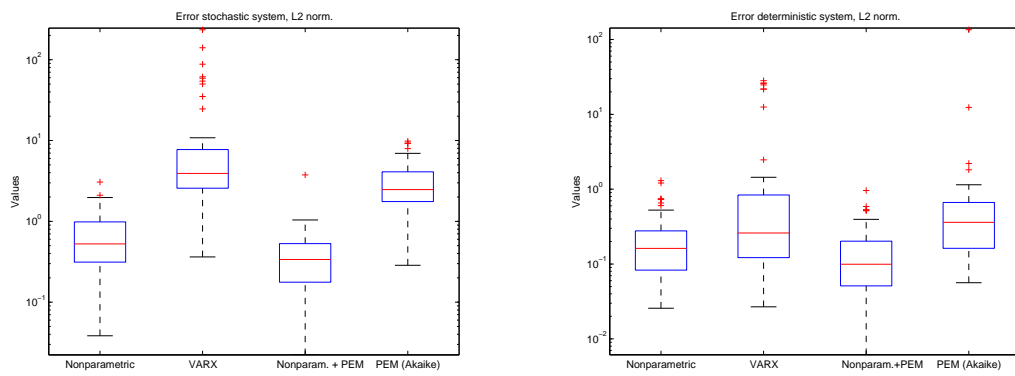Fig. 5.   System 3: singular values. a) and b): 100 Monte Carlo runs. c) and d) Average over 100 Monte Carlo runs.

Fig. 6.   System 3: Mean square error (L_2 norm of the transfer function error).