Rank of finite-sample covariance matrices for model order estimation

Cristian Guajardo and Ricardo Rojas

***Abstract*— Several approaches for structure estimation have been developed in recent years, which are mostly based on exhaustive singularity tests of covariance matrices. This work intends to reduce the computational cost of the previous methods, by knowing the rank of (instrumental) finite-sample covariance matrices.**

I. INTRODUCTION

Fitting a model to measured data is a common problem, which is normally accomplished by estimating the parameters of a previously selected model structure. This selection is an important step before parameter estimation, for example, methods such as [1] requiere the previous knowledge of at least the system order, and also the consistence of well-known identification methods, *least squares* (LS) and *intrumental variables* (IV), depends on the proper choice of the model order [2], [3].

Recently several approaches for structure estimation have been developed, mostly based on singularity tests of covariance matrices, which search exhaustively for the system structure among the possible models [4], [5], [6] and [7]. This kind of search would assume an important computational effort if the set of candidate models is big, since a covariance matrix should be built and tested for every candidate model, in order to identify the system structure. When considering a noisy environment, some noise reductions techniques are applied in advance: Instrumental variables [5], [8], [9]; rescaling the input by an input-output power ratio [10], [11]; data fitting techniques [4].

This work intends to reduce the computational cost of the structure determination, so that the rank of finite-sample covariance matrices is obtained for the noiseless case in section 2. Section 3 takes advantage of this rank providing a non-exhaustive method for structure identification. In section 4, the rank of instrumental finite-sample covariance matrices is obtained, so that the algorithm proposed in section 3 can be applied in presence of output noise of unknown probability distribution. Finally, conclusions and comments are given in section 5.

II. PRODUCT MATRIX PROPERTIES

Consider a system described by $x(k) = G(q)u(k)$

$$G(q) = \frac{B(q)}{A(q)} = \frac{b_0 + b_1 q^{-1} + \cdots + b_{n_b} q^{-n_b}}{1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a}} \tag{1}$$

b_{n_b} and a_{n_a}	are not zero
$B(q)$ and $A(q)$	are coprime
$G(q)$	is a stable filter

which can also be written as a linear regression through

$$\begin{aligned} x(k) &= \omega_{ux}^{n_b,n_a}(k)^T\beta \\ \omega_{ux}^{n_b,n_a}(k) &= \begin{bmatrix} u(k) \\ \vdots \\ u(k-n_b) \\ x(k-1) \\ \vdots \\ x(k-n_a) \end{bmatrix}, \quad \beta = \begin{bmatrix} b_0 \\ \vdots \\ b_{n_b} \\ -a_1 \\ \vdots \\ -a_{n_a} \end{bmatrix} \end{aligned} \tag{2}$$

The main problem now consists in obtaining the values n_b and n_a of the system in (1), from the measured data

$$Z_{ux}^{L,M} = \begin{bmatrix} u(-M-L) & \cdots & u(M) \\ x(-M-L) & \cdots & x(M) \end{bmatrix} \tag{3}$$

This can be achieved by arranging the measured data $Z_{ux}^{L,M}$ in the following $(2M+1)\times(\hat{n}_b+1+\hat{n}_a)$ matrix

$$\Omega_{ux}^M(\hat{n}_b,\hat{n}_a) := \begin{bmatrix} \omega_{ux}^{\hat{n}_b,\hat{n}_a}(-M)^T \\ \vdots \\ \omega_{ux}^{\hat{n}_b,\hat{n}_a}(M)^T \end{bmatrix} \tag{4}$$

where $L \geq \max(\hat{n}_b,\hat{n}_a)$ and the number of measured data pairs is $N = 2M+1+L$.

Assume now that M is large enough. Clearly, for the case $i = \min(\hat{n}_b - n_b, \hat{n}_a - n_a) > 0$, the matrix $\Omega_{ux}^M(\hat{n}_b,\hat{n}_a)$ is always rank-deficient, because it has at least i linearly dependent (l.d.) columns due to the system equation

$$x(k-l) = \omega_{ux}^{n_b,n_a}(k-l)^T\beta, \quad l = 1,\ldots,i$$

Then, the matrix $\Omega_{ux}^M(\hat{n}_b,\hat{n}_a)$ has a rank given by

$$\begin{aligned} &\text{if } i \leq 0, && \operatorname{rank}\Omega_{ux}^M(\hat{n}_b,\hat{n}_a) \leq \hat{n}_b+1+\hat{n}_a \\ &\text{if } i > 0, && \operatorname{rank}\Omega_{ux}^M(\hat{n}_b,\hat{n}_a) \leq \hat{n}_b+1+\hat{n}_a - i \end{aligned}$$

where $\hat{n}_b+1+\hat{n}_a-i = \max(\hat{n}_b+n_a, n_b+\hat{n}_a)+1$.

In order to turn the previous rank inequalities into equalities, the following $(\hat{n}_b+1+\hat{n}_a)\times(\hat{n}_b+1+\hat{n}_a)$ product is defined

$$\begin{aligned} \Lambda_{zuy}^M(\hat{n}_b,\hat{n}_a) &:= \frac{\Omega_{uz}^M(\hat{n}_b,\hat{n}_a)^T\Omega_{uy}^M(\hat{n}_b,\hat{n}_a)}{2M+1} \\ \Lambda_{zuy}^M(\hat{n}_b,\hat{n}_a) &= \frac{1}{2M+1}\sum_{k=-M}^{M}\omega_{uz}^{\hat{n}_b,\hat{n}_a}(k)\omega_{uy}^{\hat{n}_b,\hat{n}_a}(k)^T \end{aligned} \tag{5}$$

C. Guajardo is with Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile, cgware@elo.utfsm.cl

R. Rojas is with the Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile, ricardo.rojas@usm.cl

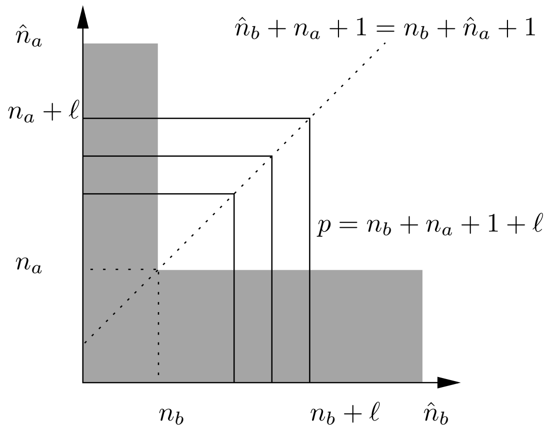

Fig. 1. Gray region: Λ_{xux}^M is full-rank. White region: Λ_{xux}^M is rank-deficient.

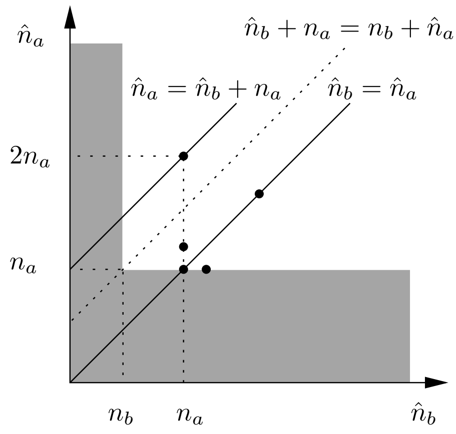

Fig. 2. Search for (n_b, n_a) in the case $n_b < n_a$.

where each element of this matrix is defined by the sample-covariance for $\tau_1, \tau_2 \in [0, L]$

$$\varphi_{\alpha\beta}^M(\tau_2 - \tau_1) := \frac{1}{2M+1} \sum_{k=-M}^{M} \alpha(k - \tau_1)\beta(k - \tau_2) \quad (6)$$

Then, the rank of the matrices

$$\operatorname{rank} \Lambda_{xux}^M(\hat{n}_b, \hat{n}_a) = \operatorname{rank} \Omega_{ux}^M(\hat{n}_b, \hat{n}_a)$$

can be totally determined with the following lemma

Lemma 2.1: Let the data $Z_{ux}^{L,M}$ be measured from the system in (1). If $s(k) = A(q)^{-1}u(k)$ is locally persistent exciting (l.p.e.)[1] of order $p \ \forall k \in [-M, M]$, then the matrix $\Lambda_{xux}^M(\hat{n}_b, \hat{n}_a)$ has the following rank

$$\operatorname{rank} \Lambda_{xux}^M(\hat{n}_b, \hat{n}_a) = \begin{cases} \hat{n}_b + 1 + \hat{n}_a, & i \leq 0 \\ \max(\hat{n}_b + n_a, n_b + \hat{n}_a) + 1, & i > 0 \end{cases}$$

$$\begin{aligned} p &= \max(\hat{n}_b + n_a, n_b + \hat{n}_a) + 1, \quad 2M + 1 \geq p \\ i &= \min(\hat{n}_b - n_b, \hat{n}_a - n_a), \quad L \geq \max(\hat{n}_b, \hat{n}_a) \end{aligned}$$

For the case $M \to \infty$ the result is the same, but l.p.e. should be replaced by p.e[2]. The proof for the $M \to \infty$ case is similar to the finite case, shown below.

Proof: For the case $i \leq 0$, the matrix $\Lambda_{xux}^M(\hat{n}_b, \hat{n}_a)$ is positive definite by theorem 1.3 (see appendix). For the case $i > 0$, the rank is determined by removing all the known l.d. columns of $\Omega_{ux}^M(\hat{n}_b, \hat{n}_a)$ and then by applying the theorem 1.4 (see appendix). ■

A different approach is used in [11] to obtain similar results.

Remark 2.1: Note that the requirements for lemma 2.1 are quite mild, since a finite number of samples is needed ($N \geq 2M + 1 + L$) and also the signal $s(k) = A(q)^{-1}u(k)$ needs just to satisfy the required (local) persistence of excitation. In case of $u(k)$ is previously given, that is, there was no chance to design it, the (l.)p.e. condition can be verified by using the definitions 1.1 or 1.2.

[1] locally persistent exciting, see definition 1.1

[2] persistently exciting, see definition 1.2

Figure 1 shows the regions where $\Lambda_{xux}^M(\hat{n}_b, \hat{n}_a)$ is full-rank and rank-deficient, and also some level curves for the required (local) persistence of excitation p. Now, the structure determination problem consists basically in finding the corner (n_b, n_a), so that singularity tests can be exhaustively performed on $\Lambda_{xux}^M(\hat{n}_b, \hat{n}_a)$, by constructing a table of some *Singularity Index* $\mathrm{SI}(\hat{n}_b, \hat{n}_a)$ for the region $(\hat{n}_b, \hat{n}_a) \in [1, \hat{n}_{\max}] \times [1, \hat{n}_{\max}]$, with $\hat{n}_b$ varying across the rows and $\hat{n}_a$ across the columns of the table. Next, a search is performed for the corner before $\mathrm{SI}(\hat{n}_b, \hat{n}_a)$ drops sharply. Examples of such singularity indexes can be

$$\det \Lambda_{xux}^M(\hat{n}_b, \hat{n}_a), \quad \sigma_{\min}(\Lambda_{xux}^M(\hat{n}_b, \hat{n}_a))$$

where $\sigma_{\min}(A)$ is the smallest singular value of A. This kind of identification procedure is performed in [4], [6], [7], [8] and [12], and it requires the construction of $\hat{n}_{\max}^2$ matrices, which could lead to high computational effort if $\hat{n}_{\max}$ is large.

III. Proposed algorithm

Consider that lemma 2.1 is satisfied and that $(\hat{n}, \hat{n})$ is chosen, with $\hat{n} > \max(n_b, n_a)$, then

$$\operatorname{rank} \Lambda_{xux}^M(\hat{n}, \hat{n}) = \hat{n} + \max(n_b, n_a) + 1$$

This means that the system order can be directly obtained, without iterations, when the rank of $\Lambda_{xux}^M(\hat{n}, \hat{n})$ is experimentally obtained

$$n := \max(n_b, n_a) = \operatorname{rank} \Lambda_{xux}^M(\hat{n}, \hat{n}) - (\hat{n} + 1)$$

Once the system order n is known, then it should be verified whether this order is n_a or n_b, which gives 3 cases

1) Figure 2 shows the case $n_b < n_a$, then this is true if the following matrices satisfy

$$\begin{aligned} \operatorname{rank} \Lambda_{xux}^M(n+1, n) &= \text{full-rank} \\ \operatorname{rank} \Lambda_{xux}^M(n, n+1) &= \text{rank-deficient} \end{aligned}$$

and this combination leads to $n_a = n$. Once n_a is known, n_b can be determined if $\hat{n}_b + n_a \leq n_b + \hat{n}_a$

$$\operatorname{rank} \Lambda_{xux}^M(\hat{n}_b, \hat{n}_a) = n_b + \hat{n}_a + 1$$

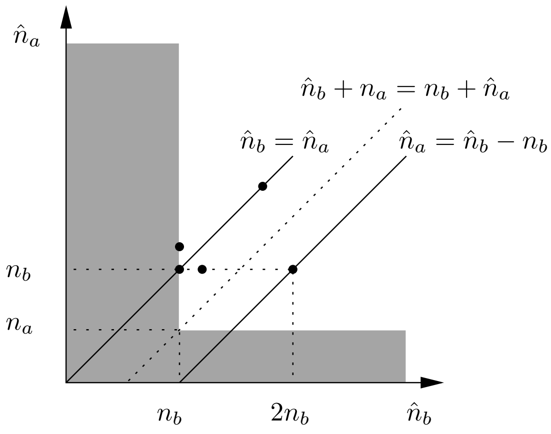

Fig. 3. Search for (n_b, n_a) in the case $n_b > n_a$.

thus, the point $(\hat{n}_b, \hat{n}_a) = (n_a, 2n_a)$ satisfies the last inequality, giving

$$n_b = \operatorname{rank} \Lambda^M_{xux}(n_a, 2n_a) - (2n_a + 1)$$

2) Figure 3 shows the case $n_b > n_a$, then this is true if the following matrices satisfy

$$\begin{aligned} \operatorname{rank} \Lambda^M_{xux}(n+1, n) &= \text{rank-deficient} \\ \operatorname{rank} \Lambda^M_{xux}(n, n+1) &= \text{full-rank} \end{aligned}$$

and this combination leads to $n_b = n$. Once n_b is kown, n_a can be determined if $\hat{n}_b + n_a \geq n_b + \hat{n}_a$

$$\operatorname{rank} \Lambda^M_{xux}(\hat{n}_b, \hat{n}_a) = \hat{n}_b + n_a + 1$$

thus, the point $(\hat{n}_b, \hat{n}_a) = (2n_b, n_b)$ satisfies the last inequality, giving

$$n_a = \operatorname{rank} \Lambda^M_{xux}(2n_b, n_b) - (2n_b + 1)$$

3) The case $n_b = n_a$ is true if the following ranks satisfy

$$\begin{aligned} \operatorname{rank} \Lambda^M_{xux}(n+1, n) &= \text{full-rank} \\ \operatorname{rank} \Lambda^M_{xux}(n, n+1) &= \text{full-rank} \end{aligned}$$

then n_b and n_a equal the system order $n_b = n_a = n$.

Notice that the matrices $\Lambda^M_{xux}(n+1, n)$ and $\Lambda^M_{xux}(n, n+1)$ can not be simultaneously rank-deficient. This procedure for structure estimation needs only the construction of at most 4 matrices, allowing an important reduction of computational effort.

IV. Instrumental product matrix properties

Consider now that a noise-corrupted output is measured

$$y(k) = x(k) + v(k) \tag{7}$$

where $v(k)$ is additive noise, of unknown probability distribution, generated by

$$v(k) = H(q)e(k), \quad H(q) = \sum_{j=0}^{\infty} h(i)q^{-j}$$

Then the perturbed matrix $\Lambda^M_{yuy}(\hat{n}_b, \hat{n}_a)$ can be built, but at first glance, nothing can be said about its rank. In fact, when the output *signal to noise ratio* (SNR) is low enough, this matrix is in most cases full-rank, making the task of structure estimation difficult to achieve.

To solve the previous problem, the *instrumental product matrix* (IPM) $\Lambda^M_{zuy}(\hat{n}_b, \hat{n}_a)$ is defined, where the instrumental sequence $z(k) = F(q)u(k)$ is generated by

$$F(q) = \frac{L(q)}{K(q)} = \frac{l_0 + l_1 q^{-1} + \cdots + l_{m_l} q^{-m_l}}{1 + k_1 q^{-1} + \cdots + k_{m_k} q^{-m_k}} \tag{8}$$

l_{m_l} and k_{m_k}	not zero
$L(q)$ and $K(q)$	coprime

Now, in view of the following lemma, the noise can be eliminated from the problem

Lemma 4.1: Consider the data Z^∞_{uy} and Z^∞_{uz}, measured from (7) and (8) respectively, where $u(k)$ and $e(k)$ are uncorrelated, that is $\varphi^\infty_{ue}(\tau) = 0$, then

$$\varphi^\infty_{uy} = \varphi^\infty_{ux}, \qquad \varphi^\infty_{zy} = \varphi^\infty_{zx}$$

Proof: Consider the sample-covariances

$$\varphi^\infty_{uy} = \varphi^\infty_{ux} + \varphi^\infty_{uv}, \qquad \varphi^\infty_{zy} = \varphi^\infty_{zx} + \varphi^\infty_{zv}$$

if the Fourier transforms of φ^∞_{uv} and φ^∞_{zv} are computed

$$\begin{aligned} \Phi^\infty_{uv}(e^{j\theta}) &= \Phi^\infty_{ue}(e^{j\theta})H(e^{-j\theta}) = 0 \\ \Phi^\infty_{zv}(e^{j\theta}) &= F(e^{j\theta})\Phi^\infty_{ue}(e^{j\theta})H(e^{-j\theta}) = 0 \end{aligned}$$

the desired result can be obtained. ■

The result of the previous lemma leads to

$$\Lambda^\infty_{zuy}(\hat{n}_b, \hat{n}_a) = \Lambda^\infty_{zux}(\hat{n}_b, \hat{n}_a)$$

Now, it would be desirable for this noiseless matrix to share the same properties as $\Lambda^M_{xux}(\hat{n}_b, \hat{n}_a)$, in order to use the algorithm in section III.

Theorem 4.1: Let the data $Z^{L,M}_{ux}$ and $Z^{L,M}_{uz}$be measured from (1) and (8) respectively. If $s(k) = A(q)^{-1}u(k)$ is l.p.e. of order p $\forall k \in [-M, M]$, and the order of $F(q)$ satisfies $\max(m_l - n_b, m_k - n_a) \geq 0$, then

$$\operatorname{rank} \Lambda^M_{zux}(\hat{n}_b, \hat{n}_a) = \begin{cases} \hat{n}_b + 1 + \hat{n}_a, & i \leq 0 \text{ y } j \leq 0 \\ \max(\hat{n}_b + n_a, n_b + \hat{n}_a) + 1, & i > 0 \text{ y } j \leq 0 \end{cases}$$

$$\begin{aligned} p &= \max(\hat{n}_b, \hat{n}_a) + \max(m_l, m_k, n_b, n_a) + 1 \\ i &= \min(\hat{n}_b - n_b, \hat{n}_a - n_a) \\ j &= \min(\hat{n}_b - m_l, \hat{n}_a - m_k) \\ L &\geq \max(\hat{n}_b, \hat{n}_a), \quad 2M + 1 \geq p \end{aligned}$$

Proof: Due to theorem 1.6, $\Omega^M_{uz}(\hat{n}_b, \hat{n}_a)$ is needed to be full-rank, in order to have

$$\operatorname{rank} \Lambda^M_{zux} = \operatorname{rank} \left\{ (\Omega^M_{uz})^T \Omega^M_{ux} \right\} = \operatorname{rank} \Omega^M_{ux}$$

For $\Omega^M_{uz}(\hat{n}_b, \hat{n}_a)$ to be full-rank in the region $j \leq 0$, by lemma 2.1, it is sufficient to have $s(k) = A(q)^{-1}u(k)$ l.p.e. of order p $\forall k \in [-M, M]$

$$p \geq \max(\hat{n}_b + m_k, \hat{n}_a + m_l) + 1$$

For $\Omega_{ux}^M(\hat{n}_b, \hat{n}_a)$ to have the rank of lemma 2.1, it is sufficient to have $s(k) = A(q)^{-1}u(k)$ l.p.e. of order $p\ \forall k \in [-M, M]$

$$p \geq \max(\hat{n}_b + n_a, \hat{n}_a + n_b) + 1$$

Besides, the condition $\max(m_l - n_b, m_k - n_a) \geq 0$ is required for $\Omega_{uz}^M(\hat{n}_b, \hat{n}_a)$ to be full-rank, while the rank of $\Omega_{ux}^M(\hat{n}_b, \hat{n}_a)$ changes from *full* to *deficient*, that is, (n_b, n_a) belongs to the region $\min(\hat{n}_b - m_l, \hat{n}_a - m_k) \leq 0$. ■

Remark 4.1: When M is large enough, but finite, the l.p.e. condition in theorem 4.1 can be approximated by p.e. condition and the result should remain the same. Nevertheless, when infinite M is considered, there are some strange cases in which the rank in theorem 4.1 does not hold. These cases are very rare, and are produced by unlucky combinations of $F(q)$, $G(q)$ and the input signal $u(k)$, see [2, example 4.4] and [13, example 8.6]. From a practical point of view, there is no need to worry about these cases, because they appear so seldom, provided some mild conditions, see [2, theorem 4.1] and [13, "generic consistency" in section 8.2].

With lemma 4.1 and theorem 4.1, it is possible to recover the properties for structure identification, by using the matrix $\Lambda_{zuy}^M(\hat{n})$ with large M, even though output noise of unknown probability distribution is present.

Example 4.1: Consider the system in [4, example 4]

$$\begin{aligned} B(q) &= 1 + 0.3452q^{-1} + 0.53q^{-2} \\ &\quad +0.3985q^{-3} + 0.8138q^{-4} \\ A(q) &= 1 + 0.7907q^{-1} + 0.042q^{-2} - 0.5556q^{-3} \\ &\quad -0.0247q^{-4} + 0.3846q^{-5} + 0.3026q^{-6} \end{aligned}$$

whose structure $(n_b, n_a) = (4, 6)$ is to be identified. The system output is perturbed by noise $v(k)$, generated by

$$v(k) = H(q)e(k), \quad H(q) = \frac{0,4q}{q - 0,6}$$

where $e(k)$ is uniform white noise of zero mean and unit variance.

The parameters are chosen so that $M = 2000$ and $L = 15$. The input signal $u(k)$ is designed as a realization of gaussian white noise of zero mean and variance 1, which is p.e. of any order (note that l.p.e condition can be approximated by p.e. since the number of samples is large). Finally the instrumental filter is designed as simple as possible, that is $F(q) = q^{-\hat{n}_b}$. All these conditions satify theorem 4.1, since models in the region $(\hat{n}_b, \hat{n}_a) \in [n_b, 15] \times [n_a, 15]$ are going to be used.

The steady state data $Z_{uy}^{L,M}$ is measured and singular values for different matrices $\Lambda_{zuy}^M(\hat{n}_b, \hat{n}_a)$ are plotted in figure 4, in order to determine their rank. To verify the behaviour of the "null singular values", 50 realizations are made for each matrix.

Following the algorithm of section III, the model $(\hat{n}_b, \hat{n}_a) = (10, 10)$ is selected, since it is supposed to be larger than the unknown system order. Now, from figure 4, $\operatorname{rank} \Lambda_{zuy}^M(10, 10) = 17$ is obtained, thus the system order can be identified

$$n = \operatorname{rank} \Lambda_{xux}^M(10, 10) - (10 + 1) = 6$$

Once the system order $n = 6$ is known, the rank of the following matrices are obtained from figure 4

$$\begin{aligned} \operatorname{rank} \Lambda_{zuy}^M(7, 6) &= \text{full-rank} \\ \operatorname{rank} \Lambda_{zuy}^M(6, 7) &= \text{rank-deficient} \end{aligned}$$

this shows that the order of the system is given by n_a, that is, $n_a = n = 6$. Finally n_b is obtained from the model $(\hat{n}_b, \hat{n}_a) = (n_a, 2n_a)$, thus $\operatorname{rank} \Lambda_{zuy}^M(6, 12) = 17$ is obtained from figure 4, giving

$$n_b = \operatorname{rank} \Lambda_{zuy}^M(6, 12) - (2 \cdot 6 + 1) = 4$$

The identified system structure is thus $(n_b, n_a) = (4, 6)$.

V. Conclusions

In a noiseless environment, and under the conditions of lemma 2.1, the rank of the finite-sample covariance matrix $\Lambda_{xux}^M(\hat{n}_b, \hat{n}_a)$ is completely known, when a signal of sufficiently high (local) persistence of excitation is used. This result could lead to exhaustive identification tests, based on the singularity of $\Lambda_{xux}^M(\hat{n}_b, \hat{n}_a)$, while searching for the system structure among the possible models

$$(\hat{n}_b, \hat{n}_a) \in [1, \hat{n}_{\max}] \times [1, \hat{n}_{\max}]$$

This requires the construction of $\hat{n}_{\max}^2$ matrices, which implies a high computational effort if $\hat{n}_{\max}$ is large.

The proposed algorithm in section III takes advantage of the rank in lemma 2.1, by reducing the number of requiered $\Lambda_{xux}^M(\hat{n}_b, \hat{n}_a)$ matrices (to a maximum of 4), for the process of structure identification.

If the system output is corrupted by noise of unkown probability distribution, the instrumental finite-sample covariance matrix $\Lambda_{zuy}^M(\hat{n}_b, \hat{n}_a)$ can be built, which reduces the effect of output noise by lemma 4.1. Besides, under the conditions of theorem 4.1, the rank of the noiseless instrumental matrix $\Lambda_{zux}^M(\hat{n}_b, \hat{n}_a)$ is also completely known, when a signal of sufficiently high (local) persistence of excitation is used. Thus, with lemma 4.1 and theorem 4.1, it is possible to recover the procedure of section III, by using the matrix $\Lambda_{zuy}^M(\hat{n})$ for large M, even though output noise of unknown probability distribution is present.

The singular value decomposition can be used to compute the rank of these matrices in the noisy case, by plotting the singular values in a decreasing order, whereby the rank of each matrix is determined by neglecting the smallest singular values, which were perturbed by output noise.

Appendix

Definition 1.1: Let the signal $s(k)$ be defined at least $\forall k \in [-M - p + 1, M]$, then $s(k)$ is said to be "locally persistent exciting" (l.p.e.) of order $p\ \forall k \in [-M, M]$ if the matrix

$$\Pi_s^M(p) := \frac{1}{2M + 1} \sum_{k=-M}^{M} v_s^p(k) v_s^p(k)^T$$

is positive-definite, where $2M + 1 \geq p$ and

$$v_s^p(k)^T = \begin{bmatrix} s(k) & \cdots & s(k - p + 1) \end{bmatrix}$$

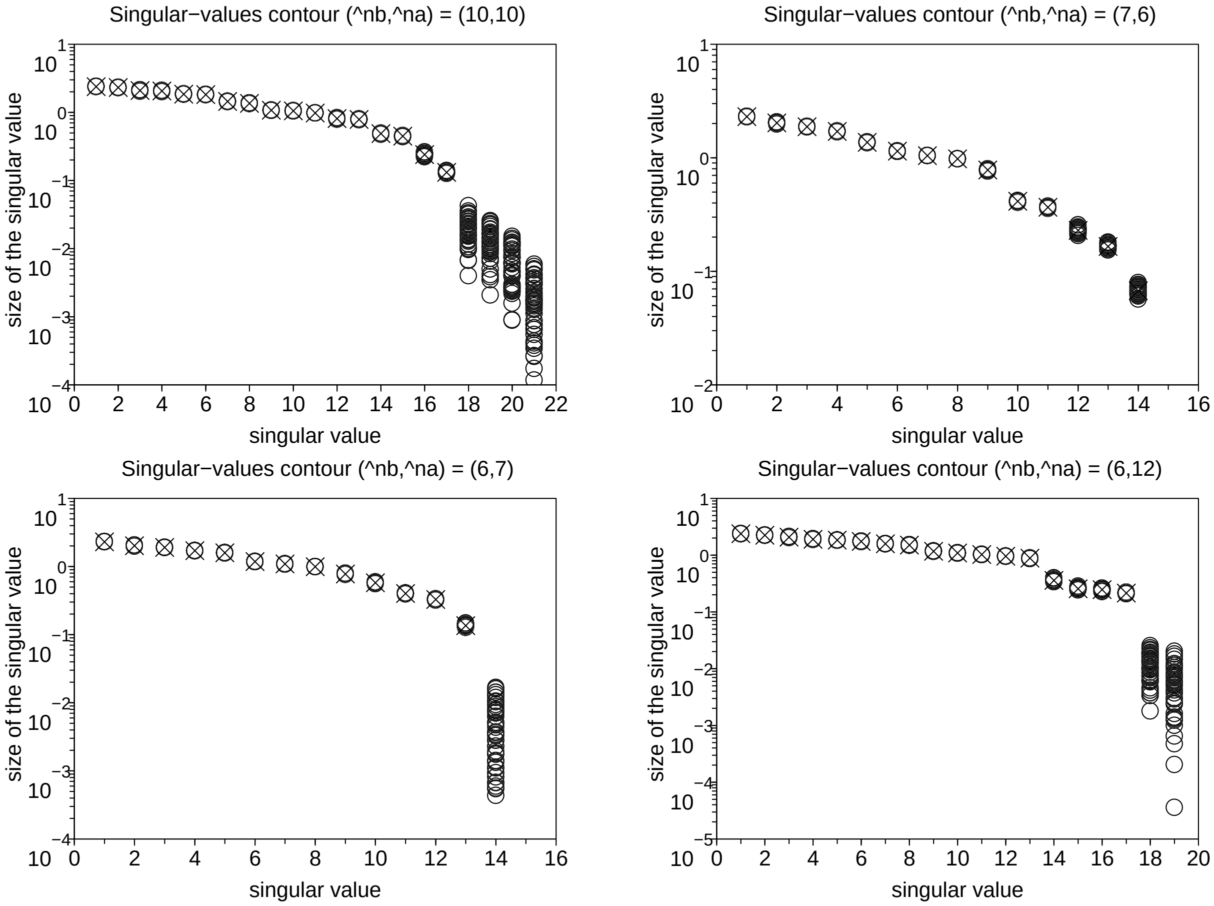

Fig. 4. Singular values of $\Lambda^M_{zuy}(\hat{n}_b, \hat{n}_a)$ are shown as "∘". Singular values of $\Lambda^M_{zux}(\hat{n}_b, \hat{n}_a)$ are shown as "×". For each graph 50 realizations were made with $M = 2000$ and output $\mathrm{SNR} \approx 10[dB]$.

Theorem 1.1: Let the signal $s(k)$ be defined at least $\forall k \in [-M-p+1, M]$, then $s(k)$ is l.p.e. of order p $\forall k \in [-M, M]$ if and only if, for any filter

$$M_p(q) = m_0 + m_1 + \cdots + m_{p-1}q^{-p+1}$$

the following implication holds

$$\Big(\forall k \in [-M, M],\ M_p(q)s(k) = 0\Big) \Rightarrow M_p(q) = 0$$

Proof: It is clear that $m^T \Pi^M_s(p) m \geq 0$ for all $m \in \mathbb{R}^{p-1}$, where $m = \left[\begin{array}{ccc} m_0 & \cdots & m_{p-1} \end{array}\right]$. Furthermore the following relations holds

$$\begin{aligned} m^T \Pi^M_s(p) m = 0 & \\ \Leftrightarrow & \quad \Big(\forall k \in [-M, M],\ m^T v^p_s(k) = 0\Big) \\ \Leftrightarrow & \quad \Big(\forall k \in [-M, M],\ M_p(q)s(k) = 0\Big) \end{aligned}$$

and also $m = 0 \Leftrightarrow M_p(q) = 0$, which completes the proof. ∎

Definition 1.2: The signal $s(k)$ is said to be "persistent exciting" (p.e.) of order p if the matrix $\Pi^\infty_s(p)$ is positive-definite, or equivalently, if for any filter of the form

$$M_p(q) = m_0 + m_1 + \cdots + m_{p-1}q^{-p+1}$$

the following implication holds

$$\forall \theta, \left|M_p(e^{j\theta})\right|^2 \Phi^\infty_{ss}(e^{j\theta}) = 0 \Rightarrow \forall \theta, M_p(e^{j\theta}) = 0$$

where $\Phi^\infty_{ss}(e^{j\theta})$ is the Fourier transform of $\varphi^\infty_{ss}(\tau)$.
For a proof of this equivalences see [3, lemma 13.1].

Theorem 1.2: Let the signal $s(k) = C(q)^{-1}u(k)$, where

$$C(q) = c_0 + c_1 q^{-1} + \cdots + c_{n_c} q^{-n_c}$$

If $u(k)$ is p.e. of order p then $s(k)$ is p.e. of order p .
The proof follows from the definition 1.2.

Theorem 1.3: Let the data $Z^{L,M}_{ux}$ be measured from the system (1), where $L \geq \max(\hat{n}_b, \hat{n}_a)$. If the signal $s(k) = A(q)^{-1}u(k)$ is l.p.e. of order $\max(\hat{n}_b + n_a, n_b + \hat{n}_a) + 1$ $\forall k \in [-M, M]$, then the matrix $\Lambda^M_{xux}(\hat{n}_b, \hat{n}_a)$, defined in (5), is positive definite for $\min(\hat{n}_b - n_b, \hat{n}_a - n_a) \leq 0$, provided that $2M + 1 \geq \max(\hat{n}_b + n_a, n_b + \hat{n}_a) + 1$.
For the case $M \to \infty$ the result is the same, but l.p.e. should be replaced by p.e. The proof for the $M \to \infty$ case is similar to the finite case, shown below.

Proof: The proof is by contradiction. Suppose that $\Lambda^M_{xux}(\hat{n}_b, \hat{n}_a)$ is not positive definte, then there exists a non-zero vector $\gamma_p = \left[\begin{array}{cccccc} d_0 & \cdots & d_{\hat{n}_b} & c_1 & \cdots & c_{\hat{n}_a} \end{array}\right]^T$ such that

$$\gamma_p^T \Lambda^M_{xux}(\hat{n}_b, \hat{n}_a)\gamma_p = 0 \Leftrightarrow \forall k \in [-M, M],\ \omega^{\hat{n}_b, \hat{n}_a}_{ux}(k)^T \gamma_p = 0$$

where the right side can be written as

$$\forall k \in [-M, M], \quad [D(q)A(q) + C(q)B(q)]\frac{u(k)}{A(q)} = 0$$

$$D(q) = \sum_{i=0}^{\hat{n}_b} d_i q^{-i}, \quad C(q) = \sum_{i=1}^{\hat{n}_a} c_i q^{-i}$$

It can be noticed that $[D(q)A(q) + C(q)B(q)] \neq 0$, since the following structures are always different for the case $\min(\hat{n}_b - n_b, \hat{n}_a - n_a) \leq 0$

$$\frac{d_0 q^{-0} + \cdots + d_{\hat{n}_b} q^{-\hat{n}_b}}{c_1 q^{-1} + \cdots + c_{\hat{n}_a} q^{-\hat{n}_a}} \neq -\frac{b_0 + b_1 q^{-1} + \cdots + b_{n_b} q^{-n_b}}{1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a}}$$

where $B(q)$ and $A(q)$ are coprime. Finally the proposition $[D(q)A(q)+C(q)B(q)] \neq 0$ is a contradiction, because $s(k)$ is l.p.e. of order $\max(\hat{n}_b + n_a, n_b + \hat{n}_a) + 1$ $\forall k \in [-M, M]$. ■

Remark 1.1: It is important to notice that given $Z_{ux}^{L,M}$, with $L \geq \max(\hat{n}_b, \hat{n}_a)$, it is possible to compute the signal $s(k)$ $\forall k \in [-M-p+1, M]$, where $p = \max(\hat{n}_b + n_a, n_b + \hat{n}_a) + 1$, which is an intermediate signal between the input $u(k)$ and the output $x(k)$

$$x(k) = B(q)s(k), \quad s(k) = A(q)^{-1}u(k)$$

This signal $s(k)$ $\forall k \in [-M-p+1, M]$ can be obtained through the set of equations

$$\begin{aligned} \forall k \in [-M-\hat{n}_b, M],\ u(k) &= A(q)s(k) \\ \forall k \in [-M-\hat{n}_a, -M-i-1],\ x(k) &= B(q)s(k) \end{aligned}$$

where $i = \min(\hat{n}_b - n_b, \hat{n}_a - n_a)$, which can be written by using the $(2M+p) \times (2M+p)$ Sylvester matrix, defined in theorem 1.5

$$v_{ux} = \mathcal{S}(A, q^{-2M-i-1}B)\, v_s$$

$$v_{ux} = \begin{bmatrix} u(M) \\ \vdots \\ u(-M-\hat{n}_b) \\ x(-M-i-1) \\ \vdots \\ x(-M-\hat{n}_a) \end{bmatrix}, \ v_s = \begin{bmatrix} s(M) \\ \vdots \\ s(-M-p+1) \end{bmatrix}$$

This set of equations has a unique solution, by theorem 1.5, since $A(q)$ and $B(q)$ are coprime.

Theorem 1.4: Let the data $Z_{ux}^{L,M}$ be measured from the system (1), where $L \geq \max(\hat{n}_b, \hat{n}_a)$. Then the following $p \times p$ matrix, with $p := \max(\hat{n}_b + n_a, n_b + \hat{n}_a) + 1$

$$L_{ux}^M(\hat{n}_b, \hat{n}_a, i) := \frac{1}{2M+1}\sum_{k=-M}^{M} w_{ux}^{\hat{n}_b,\hat{n}_a,i}(k) w_{ux}^{\hat{n}_b,\hat{n}_a,i}(k)^T$$

$$w_{ux}^{\hat{n}_b,\hat{n}_a,i}(k) := \begin{bmatrix} u(k) \\ \vdots \\ u(k-\hat{n}_b) \\ x(k-i-1) \\ \vdots \\ x(k-\hat{n}_a) \end{bmatrix} \quad (9)$$

for $i = \min(\hat{n}_b - n_b, \hat{n}_a - n_a) \geq 0$, is positive definite if and only if $s(k) = A(q)^{-1}u(k)$ is l.p.e. of order p $\forall k \in [-M, M]$, provided that $2M + 1 \geq p$.

For the case $M \to \infty$ the result is the same, but l.p.e. should be replaced by p.e. The proof for the $M \to \infty$ case is similar to the finite case, shown below.

Proof: Let the vector $w_{ux}^{\hat{n}_b,\hat{n}_a,i}(k)$ be written like

$$w_{ux}^{\hat{n}_b,\hat{n}_a,i}(k) = \frac{1}{A(q)}\begin{bmatrix} A(q)u(k) \\ \vdots \\ A(q)u(k-\hat{n}_b) \\ B(q)u(k-i-1) \\ \vdots \\ B(q)u(k-\hat{n}_a) \end{bmatrix}$$

This vector can be represented in terms of the Sylvester matrix of the theorem 1.5, for the polynomials $A(q)$ and $q^{-i-1}B(q)$

$$w_{ux}^{\hat{n}_b,\hat{n}_a,i}(k) = \mathcal{S}(A, q^{-i-1}B)v_s^p(k)$$

where the matrix $\mathcal{S}(A, q^{-i-1}B)$ is $p \times p$ and non-singular by theorem 1.5, and

$$v_s^p(k) := \begin{bmatrix} s(k) \\ \vdots \\ s(k-p+1) \end{bmatrix}, \quad s(k) := \frac{u(k)}{A(q)}$$

Now, the following product can be obtained

$$L_{ux}^M(\hat{n}_b, \hat{n}_a, i) = \mathcal{S}(A, q^{-i-1}B)\Pi_s^M(p)\mathcal{S}(A, q^{-i-1}B)^T$$

where $\Pi_s^M(p)$ is the same as in definition 1.1. Thus the proof follows from theorem 1.5 and from definition 1.1. ■

Theorem 1.5: **Sylvester's theorem**. Let the polynomials $p(x)$ and $q(x)$ be defined by

$$\begin{aligned} p(x) &= p_0 + p_1 x + \cdots + p_m x^m \\ q(x) &= q_0 + q_1 x + \cdots + q_n x^n \end{aligned}$$

The polynomials $p(x)$ and $q(x)$ are relatively prime (coprime) if and only if $\det \mathcal{S}(p, q) \neq 0$, where the Sylvester matrix, $\mathcal{S}(p, q)$ is defined as

$$\left[\begin{array}{cccccccc} p_0 & p_1 & \cdots & p_{m-1} & p_m & 0 & \cdots & 0 \\ 0 & p_0 & p_1 & \cdots & p_{m-1} & p_m & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & p_0 & p_1 & \cdots & p_{m-1} & p_m \\ \hline q_0 & q_1 & \cdots & q_{n-1} & q_n & 0 & \cdots & 0 \\ 0 & q_0 & q_1 & \cdots & q_{n-1} & q_n & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & q_0 & q_1 & \cdots & q_{n-1} & q_n \end{array}\right]$$

The submatrix related to the polynomial $p(x)$ is $\bar{n} \times (\bar{m} + \bar{n})$ and the submatrix related to the polynomial $q(x)$ is $\bar{m} \times (\bar{m} + \bar{n})$, so the matrix $\mathcal{S}(p, q)$ is $(\bar{m} + \bar{n}) \times (\bar{m} + \bar{n})$, where $\min(\bar{m} - m, \bar{n} - n) = 0$.

For a proof for this theorem see [2, lemma A3.1].

Theorem 1.6: Let $A, B \in \mathbb{C}^{m\times n}$, with $n \leq m$ and $\mathrm{rank}\,(B) = n$, then

$$\mathrm{rank}\,(B^H A) = \mathrm{rank}\,(A)$$

Proof: Let the singular value decompositions

$$\begin{aligned} A &= U\Sigma V^H = \begin{bmatrix} U_n & U_{m-n} \end{bmatrix} \begin{bmatrix} \Sigma_n \\ O \end{bmatrix} V^H \\ B &= W\Xi Y^H = \begin{bmatrix} W_n & W_{m-n} \end{bmatrix} \begin{bmatrix} \Xi_n \\ O \end{bmatrix} Y^H \end{aligned}$$

where $U, W \in \mathbb{C}^{m\times m}$, $V, Y \in \mathbb{C}^{n\times n}$ are unitary matrices and $U_n, W_n \in \mathbb{C}^{m\times n}$, $U_{m-n}, W_{m-n} \in \mathbb{C}^{m\times(m-n)}$, $\Sigma, \Xi \in \mathbb{R}^{m\times n}$, $\Sigma_n, \Xi_n \in \mathbb{R}^{n\times n}$ and $O \in \mathbb{R}^{(m-n)\times n}$.
By multipliying

$$B^H A = (Y\Xi_n^H W_n^H U_n)\Sigma_n(V^H)$$

it is possible to notice that as V^H is invertible, then it only suffices to proof that $Y\Xi_n^H W_n^H U_n \in \mathbb{C}^{n\times n}$ is invertible.
Ξ_n is invertible, because $\mathrm{rank}(B) = n$, Y is also invertible by definition and finally $W_n^H U_n \in \mathbb{C}^{n\times n}$ is an unitary matrix, then the product $Y\Xi_n^H W_n^H U_n$ is invertible, which completes the proof. ■

REFERENCES

[1] P. Stoica, M. Cedervall and A. Eriksson. Combined instrumental variable and subspace fitting approach to parameter estimation of noisy input-output systems. *IEEE Transactions on Signal Processing*, Vol. 43, Issue 10, pp. 2386-2397, October 1995.

[2] T. Söderström and P. Stoica. *Instrumental variable methods for system identification*. Springer-Verlag. Berlin, Germany, 1983.

[3] L. Ljung. *System identification: theory for the user*, Second edition. Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1999.

[4] G. Liang, D. Wilkes and J. Cadzow. ARMA model order estimation based on the eigenvalues of the covariance matrix. *IEEE Transactions on Signal Processing*, Vol. 41, Issue 10, pp. 3003-3009, October 1993.

[5] H. Duong and I. Landau. On statistical properties of a test for model structure selection using the extended instrumental variable approach. *IEEE Transactions on Automatic Control*, Vol. 39, Issue 1, pp. 211-215, January 1994.

[6] A. Al-Smadi and D. Wilkes. On estimating ARMA model orderss. *IEEE Int. Symp. on Circuits and Systems*, Vol. 2, pp. 505-508, May 1996.

[7] A. Al-Smadi and D. Wilkes. Robust and accurate ARX and ARMA model order estimation of non-gaussian processes. *IEEE Transactions on Signal Processing*, Vol. 50, Issue 3, pp. 759-763, March 2002.

[8] P. Wellstead and R. Rojas. Instrumental product moment model-order testing: extensions and applications. *International Journal of Control*, vol. 35, no. 6, pp. 1013-1027, 1982.

[9] R. Rojas. Instrumental determinant ratio test for multivariable structure identification. *2nd IASTED international congress on telecommunication and control, Telecon85*, Rio de Janeiro, Brasil, vol. 1, pp. 230-233, 1985.

[10] J. Cadzow and O. Solomon. Algebraic approach to system identification. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. 34, Issue 3, pp. 462- 469, June 1986.

[11] C. Davila and H. Chiang. An algorithm for pole-zero system model order estimation. *IEEE Transactions on Signal Processing*, vol. 43, pp. 1013-1016, April 1995.

[12] P. Wellstead. An instrumental product moment test for model order estimation. *Automatica* vol. 14, pp. 89-91, January 1978.

[13] T. Söderström and P. Stoica. *System identification*. Prentice-Hall International, Hemel Hempstead, UK, 1989.