ELSEVIER

Systems & Control Letters 34 (1998) 303–312

SYSTEMS & CONTROL LETTERS

Experimental evidence showing that stochastic subspace identification methods may fail [1]

Anders Dahlén, Anders Lindquist *, Jorge Mari

Division of Optimization and Systems Theory, Royal Institute of Technology, S-100 44 Stockholm, Sweden

Received 28 July 1997; received in revised form 21 October 1997

Abstract

It is known that certain popular stochastic subspace identification methods may fail for theoretical reasons related to positive realness. In fact, these algorithms are implicitly based on the assumption that the positive and algebraic degrees of a certain estimated covariance sequence coincide. In this paper, we describe how to generate data with the property that this condition is not satisfied. Using these data we show through simulations that several subspace identification algorithms exhibit massive failure.

Keywords: Subspace identification; Positive degree; Algebraic degree; Partial realization; Positive realness; Model reduction

1. Introduction

In [14] it was pointed out that some popular stochastic subspace identification algorithms for time series [3, 18] are based on an assumption, which may not be satisfied, about positive realness of a certain rational function and therefore may fail for generic data.

In this paper we test some of these algorithms on statistical data produced by passing white noise through a stable, rational, time-invariant, linear filter with properties to be explained in Section 4. As expected from the theoretical analysis, the algorithms exhibit massive failure for certain choices of dimensions. Since these dimensions cannot be determined easily from generic data, care has to be exercised when using these algorithms.

The basic theoretical problem with these stochastic subspace identification methods can be reduced to the fact that the positive and algebraic degrees of a partial covariance sequence need not coincide. These concepts are related to covariance extension, to be discussed in Section 2. A secondary problem arises in connection with model reduction, as explained in Section 3. In Section 4, we construct a partial covariance sequence which has the property that the positive and algebraic degrees do not agree. From this sequence we obtain a filter through which we pass white noise to generate test data. In Section 5, we present test results using the stochastic subspace identification algorithms of [3, 18] on this data. Finally, in Section 6 we present our conclusions.

2. Theoretical background

System identification in the form studied here amounts to estimating the matrices (A, B, C, D) in

* Corresponding author. Tel.: +46 8 790 7311; fax: +46 8 22 53 20; e-mail: alq@math.kth.se.

[1] This research was supported by the Swedish Research Council for Engineering Sciences (TFR).

0167-6911/98/$19.00
PII: S0167-6911(98)00020-6

some linear stochastic system

$$(\Sigma)\quad \begin{cases} x(t+1) = Ax(t) + Bw(t), \\ y(t) = Cx(t) + Dw(t), \end{cases} \tag{2.1}$$

driven by a (normalized) white noise $\{w(t)\}$, from a data string

$$\{y_0, y_1, y_2, \ldots, y_N\} \tag{2.2}$$

of observations of the output process $\{y(t)\}$. More precisely, system (2.1) should be such that $\{y(t)\}$ is a (second-order) stationary stochastic process. In particular, this means that A is a stable matrix, having all its eigenvalues less than one in modulus. Then the spectral density of $\{y(t)\}$ is given by $\Phi(z) = W(z)W(1/z)'$ on the unit circle, where $W(z) = C(zI - A)^{-1}B + D$ is the stable transfer function of system (2.1).

This estimation problem can be reduced to determining the spectral density

$$\Phi(\mathrm{e}^{\mathrm{i}\theta}) = \sum_{k=-\infty}^{\infty} \Lambda_k \mathrm{e}^{-\mathrm{i}k\theta}, \tag{2.3}$$

from the observed data (2.2), where

$$\Lambda_k = E\{y(t+k)y(t)'\}, \quad k = 0, 1, 2, \ldots \tag{2.4}$$

are the covariances of the process. To understand this matter we develop some auxiliary results which will be used in the sequel. Since $\Lambda_{-k} = \Lambda_k'$, $\Phi(z)$ may also be additively decomposed as $\Phi(z) = \Phi_+(z) + \Phi_+(1/z)'$, where

$$\Phi_+(z) = \tfrac{1}{2}\Lambda_0 + \Lambda_1 z^{-1} + \Lambda_2 z^{-2} + \cdots. \tag{2.5}$$

Now, inserting into Eq. (2.4) the output process y as defined by Eq. (2.1), a simple calculation yields

$$\Lambda_k = CA^{k-1}\bar{C}' \quad \text{for } k = 1, 2, 3, \ldots, \tag{2.6}$$

where $\bar{C} = CPA' + DB'$, and $P := E\{x(0)x(0)'\}$ is the unique solution of the Lyapunov equation $P = APA' + BB'$. Consequently, we obtain from Eq. (2.5) that

$$\Phi_+(z) = C(zI - A)^{-1}\bar{C}' + \tfrac{1}{2}\Lambda_0. \tag{2.7}$$

In the context of system identification, statistical estimates are used to approximate the covariances. Such estimates could be based on the ergodic limit

$$\Lambda_k = \lim_{T\to\infty} \frac{1}{T+1} \sum_{t=0}^{T} y_{t+k} y_t', \tag{2.8}$$

which holds almost surely [4]. Implicitly or explicitly the stochastic subspace identification procedures of [3, 18] are based on truncated ergodic sums such as

$$\Lambda_k \approx \frac{1}{T+1} \sum_{t=0}^{T} y_{t+k} y_t' \quad \text{for } k \leqslant N - T. \tag{2.9}$$

If $T < N$ is sufficiently large, these estimates will be good provided k is not too large. However, only a finite number of covariances

$$\Lambda_0, \Lambda_1, \Lambda_2, \ldots, \Lambda_v \tag{2.10}$$

can be determined from Eq. (2.2) via Eq. (2.9); in fact we must have $v \ll N$. To focus on the essential questions of this paper, we assume that Eq. (2.10) is a bona fide *partial covariance sequence* in the sense that the (block) Toeplitz matrix of Eq. (2.10) is positive definite.

The key idea of stochastic subspace identification is now to reconstruct $(A, C, \bar{C})$, appearing in Eq. (2.7), by solving the equations

$$\Lambda_k = CA^{k-1}\bar{C}' \quad \text{for } k = 1, 2, \ldots, v \tag{2.11}$$

for $(A, C, \bar{C})$, where the Λ_k are estimated as in Eq. (2.9). This is the *partial realization problem* [13, 11, 9]. Moreover, Eq. (2.7) is a *minimal* partial realization if it has the smallest possible Mc Millan degree, and yet satisfies Eq. (2.11). We call this the *algebraic degree* of Eq. (2.10).

However, for Φ to be a bona fide coercive spectral density we must have

$$\Phi(\mathrm{e}^{\mathrm{i}\theta}) > 0 \quad \text{for all real } \theta. \tag{2.12}$$

This is achieved by requiring that $\Phi_+(z)$ is (strictly) positive real, i.e. $\Phi_+(z)$ is analytic for $|z| \geqslant 1$ and satisfies $\Re[\Phi_+(z)] > 0$ in this region. The partial realization problem with the additional constraint that $\Phi_+(z)$ be positive real is called the *stochastic partial realization problem* [12], and the corresponding minimal degree is called the *positive degree* of Eq. (2.10).

In stochastic subspace identification the triplet $(A, C, \bar{C})$ is determined, either impicitly or explicitly, by a minimal factorization of a (block) Hankel matrix H_{ij} of the estimated covariances (2.10). In fact, by

Eq. (2.11), we have

$$H_{ij} := \begin{bmatrix} \Lambda_1 & \Lambda_2 & \cdots & \Lambda_j \\ \Lambda_2 & \Lambda_3 & \cdots & \Lambda_{j+1} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_i & \Lambda_{i+1} & \cdots & \Lambda_{i+j} \end{bmatrix} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{i-1} \end{bmatrix} \begin{bmatrix} \bar{C} \\ \bar{C}A' \\ \vdots \\ \bar{C}(A')^{j-1} \end{bmatrix}', \tag{2.13}$$

where the indices i, j are chosen so that $i + j = v$ and $|i - j| \leqslant 1$. Then C and $\bar{C}$ can be determined directly as the first block row of the left and right factor of Eq. (2.13), respectively. Likewise, A can be determined by removing the first block row and the last block row respectively from Γ_i, the left factor of Eq. (2.13), to obtain $\overline{\Gamma}_i$ and $\underline{\Gamma}_i$ respectively. Then $\overline{\Gamma}_i = \underline{\Gamma}_i A$, and therefore $A = \underline{\Gamma}_i^{\dagger} \overline{\Gamma}_i$, where $\underline{\Gamma}_i^{\dagger}$ is the pseudo-inverse of $\underline{\Gamma}_i$.

Assume that a positive real $\Phi_+(z)$ is obtained from the *identified* matrices A, C and $\bar{C}$, and the *estimated* zero lag covariance Λ_0. Then, B and D can be determined from the identified $(A, C, \bar{C}, \Lambda_0)$ by first finding a $P \geqslant 0$ such that

$$M(P) = \begin{bmatrix} P - APA' & \bar{C}' - APC' \\ \bar{C} - CPA' & \Lambda_0 - CPC' \end{bmatrix} \geqslant 0, \tag{2.14}$$

in general by solving an algebraic Riccati equation, and then factoring $M(P)$ as

$$M(P) = \begin{bmatrix} B \\ D \end{bmatrix} \begin{bmatrix} B \\ D \end{bmatrix}'. \tag{2.15}$$

Modulo a model-reduction step to be discussed in Section 3, all stochastic subspace algorithms are essentially variations of this scheme.

There are, however, several problems with this approach. In fact, if r is the algebraic and p the positive degree of Eq. (2.10), we have

$$\operatorname{rank} H_{ij} \leqslant r \leqslant p, \tag{2.16}$$

where each inequality could be strict. The stochastic subspace identification procedures briefly described above, on which the methods of [3, 18] are based, require that these quantities are equal in order to guarantee the positivity condition mentioned. This is an assumption that is generally implicit.

Only the first equality in Eq. (2.16) is a generic property. To illustrate this, let us consider the *scalar* output case. For the partial covariance sequence $(\Lambda_0, \Lambda_1, \Lambda_2, \ldots, \Lambda_v) := (1, 0, 0, \ldots, \frac{1}{2})$, $\operatorname{rank} H_{ij} = 1$, while $r = p = v$ [9]. This situation is a rare event, but having strict inequality in the second inequality of Eq. (2.16) is not. In fact, it is easy to see that the algebraic degree has the generic value $r = [v/2]$, any other value being rare. The situation for the positive degree is much more complex. In fact, it was shown by Byrnes and Lindquist [7] that, for any integer n such that $[v/2] \leqslant n \leqslant v$, there is a nonempty set of scalar covariance sequences (2.10) which is open in $\mathbb{R}^{v+1}$ so that n is the positive degree of Eq. (2.10).

This shows that the positive degree has no generic value. Also there is no easy way to determine the positive degree; see [7] for details. The statistical test data, which will be used in Section 5 to induce failure of the subspace identification procedures, will have the property that the algebraic and positive degrees of the corresponding partial covariance sequence (2.10) differ for certain v. The result from [7] just stated shows that such failure is a nonrare event.

The problems show up when determining B and D from Eq. (2.14). In fact, in order to perform the factorization (2.14) we must have $M(P) \geqslant 0$. However, by the Kalman–Yakubovich–Popov Lemma (see, e.g., [2]), the existence of a positive definite P such that $M(P) \geqslant 0$ is equivalent to $\Phi_+(z)$ being positive real.

3. Subspace identification algorithms

In the stochastic subspace identification methods studied here the factorization of the estimated Hankel matrix of Eq. (2.13) is performed by singular-value decomposition, either directly on H_{ij}, as in [3], or on some weighted version

$$\hat{H}_{ij} = QH_{ij}R', \tag{3.1}$$

where Q and R are nonsingular matrices, as in [17–19]. Singular value decomposition is a reliable numerical procedure for determining the numerical rank of a matrix. Hence when, factoring Eq. (3.1), the numerical rank of $\hat{H}_{ij}$, n, is chosen so that the decreasingly ordered singular values $\{\sigma_k\}$ are approximately zero for $k > n$.

In the basic stochastic subspace identification method of Van Overschee and De Moor [17–19], Q and R are chosen to be the inverted Cholesky factors of the block Toeplitz matrices of Eq. (2.10)

and $\Lambda_0, \Lambda_1', \Lambda_2', \ldots, \Lambda_\nu'$, respectively. As explained in [1] (also see [14]) this choice of weights, to which we shall refer as the *canonical weights*, is natural, since then the singular values $\{\sigma_k\}$ are the canonical correlation coefficients, i.e. the cosines of the angles between the spaces spanned by the future and past observations; see [14]. All canonical correlation coefficients are less than one in modulus.

In our context the methods of Van Overschee and De Moor [17–19] amount to factoring the square block Hankel matrix H_τ obtained by setting $i = j = \tau$ in H_{ij}. Two integer parameters need to be chosen when implementing the algorithms: τ and the desired dimension n of the system. The numerical rank determination discussed above selects $n = \operatorname{rank} H_\tau$, but the subspace identification procedures allow for choosing $n < \operatorname{rank} H_\tau$; we shall refer to this as *model reduction*. For model order selection see [10, 21]. In [17–19], model reduction is done after performing the singular value decomposition of $\hat{H}_\tau$, by setting some nonzero singular values equal to zero but keeping the matrices of left and right singular vectors fixed. From this rank-reduced matrix the system matrices A, C, and $\bar{C}$ are computed, even though it is no longer a Hankel matrix. Therefore the question whether such model reduction preserves positivity has so far no answer; see [14] for a discussion of this issue. Probably the answer is negative in general. If $\hat{H}_\tau$ is of full rank, a model reduction step is obligatory, at least if the process y is scalar. In fact, since $i = j = \tau$, the procedure of determining A described in Section 2 requires us to take $n < \tau$.

In Aoki [3] the model reduction step consists in setting some of the smallest singular values of H_τ equal to zero, which is equivalent to replacing the triplet $(A, C, \bar{C})$ obtained from the nonreduced H_τ by a triplet of "northwest corners" $(A_{11}, C_1, \bar{C}_1)$ in the sense that

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \qquad C = [C_1 \quad C_2],$$

$$\bar{C} = [\bar{C}_1 \quad \bar{C}_2].$$

This is called *principal subsystem truncation*. Since Aoki is using an unweighted Hankel matrix, i.e., $Q = R = I$ in Eq. (3.1), we do not know whether this reduction procedure preserves positivity either. However, in Theorem 7 of [14] it was shown that principal subsystem truncation preserves positivity if the canonical weights are used, provided $\{y(t)\}$ has coercive spectral density *and* τ is sufficiently large.

Consequently, failure of the stochastic subspace identification procedures may have several causes. The noncoincidence of positive and algebraic degrees prior to model reduction is the primary reason, but, as we have pointed out above, there are secondary factors, such as those connected to model reduction.

In order to illustrate the primary cause for failure, we first test our data on a modified stochastic subspace identification procedure with the following properties. First, we take $i = \tau + 1$ and $j = \tau$, that is we consider a rectangular matrix $H_{\tau+1,\tau}$. Hence, the choice $n = \tau$ is possible. Secondly, we use the canonical weights and do model reduction by principal subsystem truncation so that positivity (if it is present in the unreduced solution) is preserved, at least in the limit as $N \to \infty$ and $\tau \to \infty$ ([14], Theorem 16). We shall refer to this algorithm as the *modified subspace identification algorithm*. Let us stress that this algorithm is not introduced as an alternative to the subspace algorithms mentioned above, but merely as a means to factor out the primary cause of failure and to avoid the obligatory model reduction step mentioned above.

Then we shall test the three stochastic subspace algorithms in the book [19] by Van Overschee and De Moor. Algorithms 1 and 2 differ in the computation of the matrices A and C. Algorithm 2 does it exactly as described in Section 2, while Algorithm 1 does it in a slightly modified form, which theoretically is essentially equivalent. Algorithm 3 is of a different character. An ad hoc modification is made which produces a positive real solution *provided it first produces a stable A*. The problem here is not only that it is unclear what precisely this modification implies for the solution, but that positivity is not guaranteed, because, as for any of the subspace algorithms considered here, there is no guarantee that the identified A will be stable. In fact, it was shown in [6], that such stability is not a generic property: there are nonempty open sets in the parameter space where stability fails.

4. Construction of a test example

Next we construct a scalar covariance sequence, relevant partial sequences of which have the property that the positive degree is greater than the algebraic degree. This construction, which is motivated by the proof of Theorem 3 in [14] and the underlying theory

in [8], is based on the degree-one rational function

$$V(z)=\frac{1}{2}\frac{z+1+\varepsilon}{z+1-\varepsilon}, \tag{4.1}$$

which is stable, having a pole of modulus less than one, but is not positive real for any $\varepsilon>0$.

Expanding $V(z)$ as a Laurent series for $|z|\geqslant 1$ we obtain

$$V(z)=\tfrac{1}{2}c_0+c_1z^{-1}+c_2z^{-2}+\cdots,$$

where $c_0=1$ and $c_k=\varepsilon(\varepsilon-1)^{k-1}$ for $k>0$. Now, it is well-known [1] that $c_0,c_1,\ldots,c_v$ is a bona fide partial covariance sequence of some purely nondeterministic random sequence if and only if

$$|\gamma_k|<1, \quad k=0,1,2,\ldots,v-1, \tag{4.2}$$

where $\gamma_0,\gamma_1,\gamma_2,\ldots$ are its Schur parameters. Obviously, since $V(z)$ is not positive real, this is not the case for all v. It can be shown [8, 14] that these Schur parameters can be generated by the nonlinear dynamical system

$$\begin{aligned}\alpha_{t+1}&=\frac{\alpha_t}{1-\gamma_t^2}, \quad \alpha_0=1,\\ \gamma_{t+1}&=\frac{-\gamma_t\alpha_t}{1-\gamma_t^2}, \quad \gamma_0=\varepsilon,\end{aligned} \tag{4.3}$$

which evolves along the curve $1-\gamma_t^2=(2/\kappa-\alpha_t)\alpha_t$, where $\kappa=2/(2-\varepsilon^2)$. Hence γ_t can be eliminated in the first of the equations (4.3) to yield

$$\alpha_{t+1}=\frac{\kappa}{2-\kappa\alpha_t}, \quad \alpha_0=1, \tag{4.4}$$

Setting $\alpha_t=v_t/u_t$, Eq. (4.4) can be replaced by the linear system

$$\begin{bmatrix}u_{t+1}\\ v_{t+1}\end{bmatrix}=\begin{bmatrix}2/\kappa & -1\\ 1 & 0\end{bmatrix}\begin{bmatrix}u_t\\ v_t\end{bmatrix},$$
$$\begin{bmatrix}u_0\\ v_0\end{bmatrix}=\begin{bmatrix}1\\ 1\end{bmatrix}. \tag{4.5}$$

Since κ is greater than one in modulus, the coefficient matrix of Eq. (4.5) has complex eigenvalues and is thus, modulo a constant scalar factor, similar to

$$\begin{bmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{bmatrix},$$

where $\theta:=\arctan\sqrt{\kappa^2-1}$ is small if ε is small. Hence α_t is the slope of a line through the origin in $\mathbb{R}^2$ which rotates counter-clockwise with the constant angle θ, so that $\arctan\alpha_{t+1}=\arctan\alpha_t+\theta$.

Sooner or later, therefore, the slope α_t will either change sign or become infinite at which time condition (4.2) will fail, as can be seen from the first of equations (4.3). Supposing that ε is chosen so that finite escape does not occur, and let v be the last step prior to the slope changing sign. Then $c_0,c_1,\ldots,c_v$ will be a partial covariance sequence, but $c_0,c_1,\ldots,c_{v+1}$ will not. Clearly, however, v can be made arbitrarily large by just choosing ε small enough.

Let $\varphi_v(z)=z^v+\varphi_{v1}z^{v-1}+\cdots+\varphi_{vv}$ be the Szegö polynomial determined from $\gamma_0,\gamma_1,\ldots,\gamma_{v-1}$ via the Szegö–Levinson recursion

$$\varphi_{t+1}(z)=z\varphi_t(z)-\gamma_tz^t\varphi_t^*(z^{-1}), \quad \varphi_0(z)=1,$$

and set $r_v:=\prod_{k=0}^{v-1}(1-\gamma_k^2)$. Then the maximum-entropy filter

$$W(z)=\frac{\sqrt{r_v}}{\varphi_v(z)} \tag{4.6}$$

is stable and minimum phase [5, 20]. Let $\Phi_+(z)$ be the positive real part of the spectral density $\Phi(z):=W(z)W(z^{-1})$, and define the infinite covariance sequence $\Lambda_0,\Lambda_1,\Lambda_2,\ldots$ via the Laurent expansion (2.5). Clearly this is a covariance extension of $c_0,c_1,\ldots,c_v$. In fact, $\Lambda_k=c_k$ for $k=0,1,2,\ldots,v$.

Now let us consider the algebraic degree r and the positive degree p of the partial sequence

$$\Lambda_0,\Lambda_1,\Lambda_2,\ldots,\Lambda_k \tag{4.7}$$

as k varies. In this context, recall that the subspace algorithms studied here implicitly require that $p=r$. In our example, by construction, $p>r$ for $2\leqslant k\leqslant 2(v-1)$. For $k\geqslant 2v-1$, we always have $r=p$. Note that we can choose v arbitrarily large and thus construct an arbitrarily long interval for which the condition $p=r$ will fail.

In Fig. 1 we show a typical situation for $v=5$, depicting p (marked by $(\times)$) and r (marked by $(\bullet)$) as a function of k.

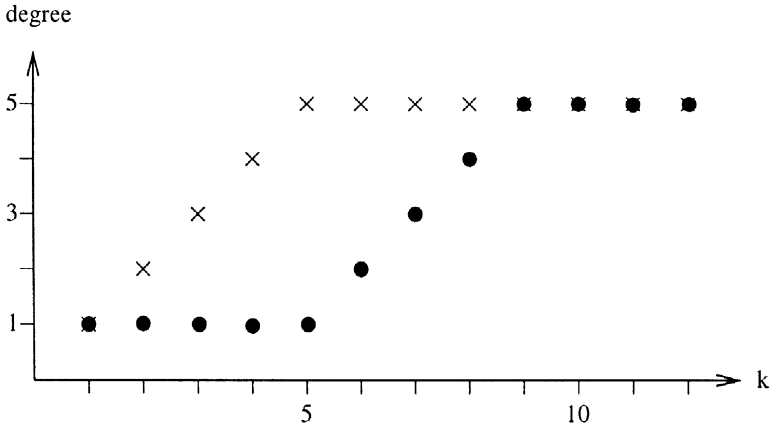

Fig. 1. The positive (×) and algebraic degrees (•) for the case $v = 5$.

5. Simulations

To obtain data we pass white noise through a filter

white noise $\longrightarrow$ $\boxed{W(z)}$ $\longrightarrow$ output process,

with the transfer function (4.6) constructed in Section 4.

The experiments will be carried out in the following manner. In each of the algorithms to be tested there are two integer parameters to be chosen: the size τ of the approximate Hankel matrix, as defined previously for each algorithm, and the desired dimension n of the resulting system. To choose n is to choose which singular values will be set equal to zero. For each choice of τ and n, we make 100 simulations and apply the subspace algorithms to each of these sets of data. The number of failures are recorded in tables. By failure we mean that the estimated triplet $(A, C, \bar{C})$ is such that Eq. (2.7) fails to be positive real.

We begin by testing the modified subspace identification algorithm introduced in Section 3 as a means to illustrate more precisely the failures anticipated from the analysis in Section 4. Thus, in particular, the most important secondary reason for failure – that the model reduction phase may not preserve positivity – has been removed. The index k in Eq. (4.7) needed for the appropriate Hankel matrix of size τ is precisely $k = 2\tau$.

First, to illustrate the behavior expected from Fig. 1, we consider a very simple example for which $v = 5$, which corresponds to the choice $\varepsilon = 0.25$ in Eq. (4.1). Table 1 shows the percentage of failure for different choices of τ and n.

Table 1
Percentage of failures for modified algorithm when $v = 5$

	n	1	2	3	4	5	6	7
τ:	1	100						
	2	99	100					
	3	78	98	100				
	4	58	100	100	100			
	5	0	0	0	0	0		
	6	0	0	0	0	0	41	
	7	0	0	0	0	0	28	69

The dimension of the system generating this data is five. From the theoretical considerations reported above, we expect the algorithm to succeed for $\tau \geqslant 5$ as long as $n \leqslant 5$. This agrees with the experiment as can be seen from the zeros in the table. We also see that the algorithm has an almost massive failure rate in the area where positive and algebraic degrees do not match, as explained in Section 4. Finally, we have large failure rate when $n > 5$; in these cases, the transfer functions of the systems may have approximate common factors in the numerators and the denominators, likely to be unstable in approximately half of the runs.

This is a very simple example, but it exhibits all the characteristics of a larger example. In fact, the corresponding result for $v = 15$, obtained by setting $\varepsilon = 0.097$, is depicted in Table 2. It shows massive failure for $\tau \leqslant 14$, which agrees with theory. Similarly, the algorithm has complete success for $\tau \geqslant 15$ as long as $n \leqslant 15$.

Next, for the case $v = 15$, we test the three stochastic subspace identification methods of Van Overschee and De Moor [19] with canonical weights. These

Table 2
Percentage of failures for modified algorithm when $v = 15$

	n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
τ:	1	85																
	2	86	95															
	3	97	96	100														
	4	93	99	98	99													
	5	99	100	99	100	100												
	6	97	99	100	99	100	100											
	7	96	98	100	100	100	100	100										
	8	60	99	100	100	100	100	100	100									
	9	47	97	100	100	100	100	100	100	100								
	10	45	97	98	98	98	98	100	100	100	100							
	11	48	88	96	95	96	96	96	96	100	100	100						
	12	43	91	97	97	97	97	97	97	97	97	100	100					
	13	42	90	95	95	97	97	98	98	98	98	98	98	100				
	14	53	93	100	100	100	100	100	100	100	100	100	100	100	100			
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	56	
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	37	67

Table 3
Percentage of failures for Algorithm 1 when $v = 15$

	n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
τ:	2	91																
	3	100	96															
	4	100	100	100														
	5	100	100	100	100													
	6	100	100	100	100	100												
	7	100	100	100	100	100	100											
	8	100	100	100	100	100	100	100										
	9	0	100	99	100	100	100	100	100									
	10	0	0	96	97	98	100	100	100	100								
	11	0	0	37	87	87	93	100	100	100	100							
	12	0	0	0	58	50	84	78	98	100	100	100						
	13	0	0	0	0	7	67	47	84	76	97	99	100					
	14	0	0	0	0	0	1	11	72	75	84	85	100	100				
	15	0	1	100	0	0	0	0	0	21	55	86	98	100	100			
	16	100	100	100	100	100	100	100	100	100	100	100	100	100	93	0		
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	46	
	18	0	0	0	0	0	0	10	1	32	4	42	24	33	17	0	38	56

algorithms operate in a setting which in our notation corresponds to a square Hankel matrix $H_\tau := H_{\tau,\tau}$, so the index k in Eq. (4.7) is $k = 2\tau - 1$. Algorithm 2 in [19] uses the procedure described in Section 2 to determine A, and the other algorithms use an equivalent shift strategy. Consequently, as pointed out in Section 3, we must take $n < \tau$, and hence we cannot allow for the case that H_τ is full rank and there is no model reduction.

The simulations for Algorithm 1 and Algorithm 2 are reported in Tables 3 and 4, respectively, which show the percentage of failure for different choices of τ and n. As before the dimension of the system generating the data is 15. Since we have at least one model reduction step, we expect the algorithm to succeed for $\tau \geqslant 16$ as long as $n \leqslant 15$. The adherence to this pattern is not as good as for the modified algorithm, and this is probably due to the secondary effects discussed in Section 3. We get massive failure not only in the area where positive and algebraic degrees do not match, but also in certain cases when they do agree. We stress that data producing failure for $\tau < v$ can

Table 4
Percentage of failures for Algorithm 2 when $v = 15$

	n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
τ:	2	94																
	3	100	99															
	4	100	100	99														
	5	100	100	99	100													
	6	100	99	100	100	100												
	7	100	100	100	100	100	100											
	8	100	100	100	100	100	100	100										
	9	100	100	98	99	99	100	100	100									
	10	100	45	99	95	99	100	100	100	100								
	11	100	12	99	71	100	92	99	100	100	100							
	12	100	50	73	24	98	78	92	99	99	100	100						
	13	100	89	19	4	28	37	92	84	89	95	99	100					
	14	94	100	100	0	18	1	8	38	91	98	100	100	100				
	15	62	100	100	4	100	0	2	0	1	3	10	21	54	97			
	16	100	100	100	100	100	100	100	100	100	100	100	100	100	90	0		
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	43	
	18	21	77	13	56	48	6	66	11	78	23	57	31	33	18	0	40	60

Table 5
Percentage of failures for Algorithm 3 when $v = 15$

	n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
τ:	2	20																
	3	11	55															
	4	8	31	63														
	5	6	24	46	79													
	6	2	15	42	65	89												
	7	0	26	49	65	77	97											
	8	0	22	29	48	67	84	92										
	9	0	15	91	100	100	100	100	100									
	10	0	0	96	91	95	98	99	100	100								
	11	0	0	36	58	80	90	99	100	100	100							
	12	0	0	0	7	48	49	67	92	99	100	100						
	13	0	0	0	0	6	11	47	62	68	94	99	100					
	14	0	0	0	0	0	0	11	23	70	50	77	98	100				
	15	0	0	100	0	0	0	0	0	19	18	68	82	100	100			
	16	0	7	100	1	94	0	62	0	1	0	0	0	0	0	0		
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	46	
	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	38	56

be constructed for arbitrarily large v, as explained in Section 4.

The corresponding simulations for Algorithm 3, reported in Table 5, show failure which is due to unstability of A. This is in complete agreement with the theory since stability is not automatic unless $p = r$; see [6]. The failure in the critical region is not as massive, but on the other hand the meaning of the solution is not entirely clear, since the modification needed is quite ad hoc.

Finally, we test Aoki's method [3]. As seen in Table 6, a similar pattern as in the modified subspace algorithm occurs, except that sporadic failures may also occur in the 'good region', as expected from the manner in which model reduction is performed. On the other hand, the algorithm is likely to succeed if n is chosen much smaller than τ.

6. Conclusions

In [14] it was pointed out that there is no guarantee that some popular stochastic subspace identification algorithms will actually work for generic data,

Table 6
Percentage of failures for Aoki's method when $v = 15$

	n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
τ:	1	80																
	2	97	96															
	3	100	100	98														
	4	100	99	98	100													
	5	100	100	100	100	100												
	6	100	100	100	100	100	100											
	7	100	100	100	100	100	100	100										
	8	99	87	98	100	100	100	100	100									
	9	0	0	45	90	96	100	100	100	100								
	10	0	0	0	0	70	91	100	100	100	100							
	11	0	0	0	0	0	0	57	95	99	100	100						
	12	0	0	0	0	0	0	0	6	61	97	99	100					
	13	0	0	0	0	0	0	0	0	1	98	62	95	100				
	14	0	0	0	0	0	0	0	0	0	0	0	100	66	100			
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0		
	16	0	4	0	0	0	0	0	0	1	3	4	4	9	15	21	48	
	17	0	0	0	0	4	3	11	7	6	4	8	11	20	29	34	56	68

and theoretical evidence for this fact was provided.

In this paper we constructed statistical data for which said identification algorithms exhibit massive failure. These data were produced by passing white noise through a filter which was constructed in such a way that the corresponding partial covariance sequences of the output process do not fulfill the requirement of stochastic subspace identification that the positive and algebraic degrees coincide.

Failure can also occur if the dimension of the identified system is chosen too high, so that almost cancellations of unstable factors occur between the numerator and the denominator of the transfer functions.

Consequently some care has to be exercised when using these stochastic subspace identification methods. In [15] possible remedies are discussed, and in [16] the authors present an alternative identification procedure which overcomes these difficulties.

References

[1] H. Akaike, Markovian representation of stochastic processes by canonical variables, SIAM J. Control 13 (1975) 162–173.

[2] B.D.O. Anderson, S. Vongpanitlerd, Network Analysis and Synthesis, Prentice-Hall, Englewood Cliffs, NJ, 1973.

[3] M. Aoki, State Space Modeling of Time Series, Springer, Berlin, 1987.

[4] P.J. Brockwell, R.A. Davis, Time Series: Theory and Methods, Springer, Berlin, 1993.

[5] J.P. Burg, Maximum entropy spectral analysis, Ph.D. Thesis, Stanford University, CA, 1975.

[6] C.I. Byrnes, A. Lindquist, The stability and instability of partial realizations, Systems Control Lett. 2 (1982) 99–105.

[7] C.I. Byrnes, A. Lindquist, On the partial stochastic realization problem, IEEE Trans. Automat. Control 42 (1997) 1049–1070.

[8] C.I. Byrnes, A. Lindquist, T. McGregor, Predictability and unpredictability in Kalman filtering, IEEE Trans. Automat. Control 36 (1991) 563–579.

[9] W.B. Gragg, A. Lindquist, On the partial realization problem, Linear Algebra Appl. 50 (1983) 277–319.

[10] E.J. Hannan, M. Deistler, The Statistical Theory of Linear Systems, Series in Probability and Mathematical Statistics, Wiley, New York, 1988.

[11] R.E. Kalman, On partial realizations, transfer functions and canonical forms, Acta Polytech. Scand., Math. Comput. Sci. Ser. MA31 (1979) 9–39.

[12] R.E. Kalman, Realization of covariance sequences, in: Proc. Toeplitz Memorial Conf., Tel Aviv, Israel, 1981.

[13] R.E. Kalman, P.L. Falb, M.A. Arbib, Topics in Mathematical Systems Theory, McGraw-Hill, New York, 1969.

[14] A. Lindquist, G. Picci, Canonical correlation analysis, approximate covariance extension, and identification of stationary time series, Automatica 32 (1996) 709–733.

[15] J. Mari, Rational modeling of time series and applications of geometric control. Ph.D. Thesis, Royal Institute of Technology, 1998.

[16] J. Mari, A. Dahlén, A. Lindquist, A covariance extension approach to identification of time series, submitted.

[17] P. Van Overschee, Subspace identification, theory – implementation – application, Ph.D. Thesis, Katholieke Universiteit Leuven, 1995.

[18] P. Van Overschee, B. De Moor, Subspace algorithms for stochastic identification problem, Automatica J. 3 (1993) 649–660.

[19] P. Van Overschee, B. De Moor, Subspace Identification for Linear Systems — Theory Implementation Applications, Kluwer Academic, Dordrecht, 1996.
[20] B. Porat, Digital Processing of Random Signals, Theory and Methods, Prentice-Hall, Englewood Cliffs, NJ, 1994.
[21] J. Sorelius, T. Söderström, P. Stoica, M. Cedervall, On partial realizations, transfer functions and canonical forms, in: SYSID'97 11th IFAC Symp. on System Identification, vol. 3, 1997, pp. 1131–1136.