Automatica, Vol. 29, No. 3, pp. 649–660, 1993
Printed in Great Britain.

0005–1098/93 $6.00 + 0.00

Subspace Algorithms for the Stochastic Identification Problem*†

PETER VAN OVERSCHEE‡§|| and BART DE MOOR‡¶

A new subspace algorithm consistently identifies stochastic state space models directly from given output data, using only semi-infinite block Hankel matrices.

Key Words—System identification; stochastic systems; stochastic approximation; Kalman filters; difference equations; QR and quotient singular value decomposition.

Abstract—In this paper, we derive a new subspace algorithm to consistently identify stochastic state space models from given output data without forming the covariance matrix and using only semi-infinite block Hankel matrices. The algorithm is based on the concept of principal angles and directions. We describe how they can be calculated with QR and Quotient Singular Value Decomposition. We also provide an interpretation of the principal directions as states of a non-steady state Kalman filter bank.

1. INTRODUCTION

LET $y_k \in \Re^l$, $k = 0, 1, \ldots, K$ be a data sequence that is generated by the following system:

$$x_{k+1} = Ax_k + w_k, \tag{1}$$

$$y_k = Cx_k + v_k, \tag{2}$$

where $x_k \in \Re^n$ is a state vector. The vector sequence $w_k \in \Re^n$ is a process noise while the vector sequence $v_k \in \Re^l$ is a measurement noise. They are both assumed to be zero mean, white, Gaussian with covariance matrix**

$$\mathbf{E}\left[\begin{pmatrix} w_k \\ v_k \end{pmatrix}(w_p^t \quad v_p)\right] = \begin{pmatrix} Q & S \\ S^t & R \end{pmatrix}\delta_{kp},$$

where δ_{kp} is the Kronecker delta. It is assumed that the stochastic process is zero mean stationary, i.e.: $\mathbf{E}[x_k] = 0$ and $\mathbf{E}[x_k x_k^t] = \Sigma$ (say). The state covariance matrix Σ is independent of the time k. This implies that A is a stable matrix (all of its poles are strictly inside the unit circle).

The central problem discussed in this paper is the identification of a state space model from the data y_k (including the determination of the system order n) and the determination of the noise covariance matrices. The state space model should be equivalent up to within second order statistics of the output.

The main contributions of this paper are the following:

- Since the pioneering papers by Akaike (1975), canonical correlations (which were first introduced by Jordan (1875) in linear algebra and then by Hotelling (1936) in the statistical community) have been used as a mathematical tool in the stochastic realization problem. In this paper we show how the approach by Akaike (1975) and others (e.g. Arun and Kung, 1990; Larimore, 1990) reduces to applying canonical correlation analysis to two matrices that are double infinite (i.e. have an infinite number of rows and columns). A careful analysis reveals the nature of this double infinity and we manage to reduce the canonical correlation approach to a semi-infinite matrix problem, i.e. only the number of columns needs to be very large while the number of (block) rows remains sufficiently small. This observation is extremely relevant with respect to the use of updating techniques.
- In order to find the state space model, we derive a finite-dimensional vector sequence (principal directions) which, in the case of double infinite block Hankel matrices, would be a valid state sequence of the stochastic

* Received 2 January 1992; received in final form 12 September 1992. The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor B. Wahlberg under the direction of Editor P. C. Parks.

† The research reported in this paper was partially supported by the Belgian Program on Inter University Attraction Poles initiated by the Belgian State Science Policy Programming (Prime Minister's Office) and the European Community Research Program ESPRIT, Basic Research Action nr. 3280. The scientific responsibility is assumed by its authors.

‡ Katholieke Universiteit Leuven, Department of Electrical Engineering—ESAT. Kardinaal Mercierlaan 94, 3001 Heverlee, Belgium. email: vanoverschee @esat.kuleuven.ac.be Fax: 32/16/221855.

§ Research Assistant of the Belgian National Fund for Scientific Research.

|| Author to whom all correspondence should be sent.

¶ Research Associate of the Belgian National Fund for Scientific Research.

** The expected value operator is denoted by $\mathbf{E}[.]$.

model. This sequence would correspond to the outputs of an infinite number of steady state Kalman filters with an infinite number of output measurements as inputs. For the semi-infinite matrix problem, the sequence corresponds to the output of an infinite number of non-steady state Kalman filters that have only used a finite number of output data as input. These state sequences are obtained directly from the output data, without any need for the state space model. The state space model is then derived from these sequences by solving a least squares problem. Figure 1 illustrates the difference between this approach and the classical one, where first the system matrices are identified, where after the states are determined through a Kalman filter.

In our point of view, the notion of a state is largely underestimated in the context of system identification, while it is well accepted in the context of control system design. Most identification approaches are based on an input–output (transfer matrix) framework.

- We give a precise interpretation of the fact that state space models, obtained via principal angles and directions, are approximately balanced in the stochastic sense.
- It is a common belief that ARMA model identification requires non-linear estimation methods, requiring iterative solution techniques (see e.g. Ljung, 1987). In that framework, the main problems arise from convergence difficulties and/or the existence of local minima. With the subspace approach of this paper, this ARMA modelling is basically reduced to the solution of an eigenvalue problem, for which numerically robust and always convergent algorithms exist.
- We derive a numerically robust square root algorithm, that mainly uses QR-decomposition and Quotient Singular Value Decomposition (QSVD) of the triangular factors and is completely data driven instead of covariance driven. The fact that only one of the dimensions of the matrices involved needs to go to infinity is very important since then updating techniques can be used.

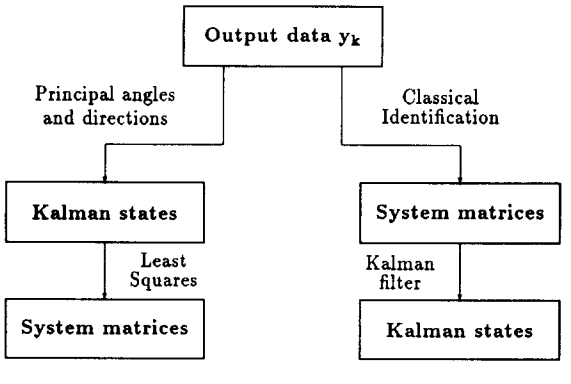

FIG. 1. The left hand side shows the new approach: first the Kalman states, then the system matrices. The right hand side is the classical approach: first the system, and then an estimate of the states.

We should mention the work by Aoki (1987), in which a "covariance" framework for the stochastic realization is derived (in contrast to our square root version). Another aspect is that we emphasize more than is done in Aoki (1987), is the careful analysis of the dynamics of the algorithm as a function of the number of block rows of the Hankel matrices involved. This paper is organized as follows. In Section 2, we summarize the main properties of linear time-invariant stochastic processes, including the non-uniqueness of the state space description. In Section 3, we briefly present the solution via the classical realization approach, which is based on the fact that the output covariance matrices can be considered as Markov parameters of a deterministic linear time-invariant discrete time system. In Section 4, we describe how the projection of past and future block Hankel matrices of output data, leads to the definition of the state sequences. In Section 5 we explore the relationship of these state sequences with the non-steady state Kalman filter. From these sequences, it is easy to derive the stochastic state space model. This will be done in Section 6. The second part of this paper consists of a description of a numerical robust algorithm to calculate the state sequences and the state space matrices. In Section 7 we provide a geometrical interpretation of the main mathematical–geometrical tool: principal angles and directions. We will also show how the state sequences can be calculated as principal directions. In Section 8, we describe a numerical robust computation scheme based on the QR and quotient singular value decomposition. We also put everything together and arrive at a new, numerically robust algorithm to calculate the state space model of the stochastic system directly from output data. Some illustrative examples are given in Section 9.

2. LINEAR TIME-INVARIANT STOCHASTIC PROCESSES

In this section, we summarize the main properties of linear time invariant stochastic processes, including the non-uniqueness of a state space description.

First we will develop some (well-known) structural relations for linear time-invariant stochastic processes. Since w_k and v_k are zero mean white noise vector sequences, independent

of x_k, we know that: $\mathbf{E}[x_k v_k^t] = 0$ and $\mathbf{E}[x_k w_k^t] = 0$. Then we find the Lyapunov equation for the state covariance matrix $\Sigma \overset{\text{def}}{=} \mathbf{E}[x_k x_k^t] = A\Sigma A^t + Q$. Defining the output covariance matrices $\Lambda_i \overset{\text{def}}{=} \mathbf{E}[y_{k+i} y_k^t]$ we find for Λ_0:

$$\Lambda_0 = C\Sigma C^t + R. \tag{3}$$

Defining

$$G \overset{\text{def}}{=} A\Sigma C^t + S, \tag{4}$$

we get

$$\Lambda_i = CA^{i-1}G, \tag{5}$$

$$\Lambda_{-i} = G^t(A^{i-1})^t C^t \quad i = 1, 2, \ldots \tag{6}$$

The model (1)–(2) can be converted into a so-called forward innovation model and a backward innovation model (see e.g. Pal, 1982). The forward innovation model is obtained by applying a Kalman filter to the stochastic system (1)–(2): $z_{k+1} = Az_k + w_k^f$, $y_k = Cz_k + v_k^f$, where the forward noise sequences w_k^f and v_k^f are given by $w_k^f = K_f v_k^f$ and $v_k^f = y_k - Cz_k$. Here K_f is the Kalman gain: $K_f = (G - APC^t)(\Lambda_0 - CPC^t)^{-1}$ and P is the forward state covariance matrix, which can be determined from the forward Riccati equation:

$$P = APA^t + (G - APC^t) \times (\Lambda_0 - CPC^t)^{-1}(G - APC^t)^t. \tag{7}$$

Associated with every forward model (1)–(2), there is a backward model that can be obtained as follows. Define the minimum variance estimate of x_k based on x_{k+1} as (Papoulis, 1984): $\Pi(x_k \mid x_{k+1}) = \mathbf{E}[x_k x_{k+1}^t](\mathbf{E}[x_{k+1} x_{k+1}^t])^{-1} x_{k+1}$. We now have:

$$\begin{aligned} x_k &= \Pi(x_k \mid x_{k+1}) + (x_k - \Pi(x_k \mid x_{k+1})) \\ &= \mathbf{E}[x_k x_{k+1}^t](\mathbf{E}[x_{k+1} x_{k+1}^t])^{-1} x_{k+1} \\ &\quad + (x_k - \Pi(x_k \mid x_{k+1})) \\ &= \mathbf{E}[x_k x_{k+1}^t]\Sigma^{-1} x_{k+1} + (x_k - \Pi(x_k \mid x_{k+1})) \\ &= \mathbf{E}[x_k(x_k^t A^t + w_k^t)]\Sigma^{-1} x_{k+1} \\ &\quad + (x_k - \Pi(x_k \mid x_{k+1})) \\ &= \Sigma A^t \Sigma^{-1} x_{k+1} + (x_k - \Pi(x_k \mid x_{k+1})). \end{aligned}$$

Now define the backward state $s_k = \Sigma^{-1} x_k$, then

$$s_k = A^t s_{k+1} + w_k, \tag{8}$$

where $w_k = \Sigma^{-1}(x_k - \Pi(x_k \mid x_{k+1}))$. For the output equation, we obtain in a similar way:

$$y_k = G^t s_k + v_k, \tag{9}$$

with $v_k = (y_k - \Pi(y_k \mid x_{k+1}))$.

Associated with the general backward model (8)–(9) is the backward innovations model:

$$\begin{aligned} r_{k-1} &= A^t r_k + w_k^b, \\ y_k &= G^t r_k + v_k^b, \end{aligned}$$

with $w_k^b = K_b v_k^b$ and $v_k^b = y_k - G^t r_k$. Here K_b is the backward Kalman gain: $K_b = (C^t - A^t N^{-1} G)(\Lambda_0 - G^t N^{-1} G)^{-1}$ and N is the backward state covariance matrix, which can be determined from the backward Riccati equation:

$$N = A^t NA + (C^t - A^t NG) \times (\Lambda_0 - G^t NG)^{-1}(C^t - A^t NG)^t. \tag{10}$$

It is well known that the stochastic model for the data y_k is not unique. Not only can we introduce an arbitrary similarity transformation for the state ($x_k \to Tx_k$) as with all state space models. In addition, there is a whole set of possible state covariance matrices and covariance matrices Q, R and S that give the same second order output covariance matrices. This fact was described in detail by Faurre (1976) and we provide here only a summary of some interesting results.

The set of stochastic models that generate the output covariance matrices Λ_i is characterized as follows.

(1) The matrices A, C and G are unique up to within a similarity transformation.
(2) The set of all state covariance matrices Σ is closed and bounded: $P \leq \Sigma \leq N^{-1}$ where P and N are the solutions of the forward (7), respectively backward (10) Riccati equations. (Inequalities are to be interpreted in the sense of nonnegative definiteness.)
(3) For every state covariance matrix Σ satisfying these bounds, the associated matrices R, S and Q follow from the equations $Q = \Sigma - A\Sigma A^t$, $S = G - A\Sigma C^t$ and $R = \Lambda_0 - C\Sigma C^t$.

Definition 1. A stochastic model will be labeled as balanced if the solutions to the forward and backward Riccati equations are diagonal and equal.

3. THE CLASSICAL REALIZATION APPROACH

In this section, we briefly present the classical realization approach, which is based on the fact that the output covariances can be considered as Markov parameters of a linear time invariant system.

In what follows, output block Hankel matrices of the form

$$Y_{0|i-1} = \begin{pmatrix} y_0 & y_1 & y_2 & \cdots & y_{j-2} & y_{j-1} \\ y_1 & y_2 & y_3 & \cdots & y_{j-1} & y_j \\ y_2 & y_3 & y_4 & \cdots & y_j & y_{j+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ y_{i-1} & y_i & y_{i+1} & \cdots & y_{i+j-3} & y_{i+j-2} \end{pmatrix}.$$

will play a crucial role. The first subscript denotes the time index of the upper left element

while the second subscript is the time index of the bottom left element. For all output block Hankel matrices, the number of columns will be j and for all theoretical derivations, we assume that $j \to \infty$. How to deal with a finite number of output measurements will become clear in Section 8.

An important remark concerns the estimation of the covariance matrices. Hereto we assume that they can be estimated as

$$\Lambda_i = \lim_{j\to\infty}\left[\frac{1}{j}\sum_{k=0}^{j-1} y_{k+i}y_k^t\right],$$

for which we will use the notation $\mathbf{E}_\mathbf{j}\left[\sum_{k=0}^{j-1} y_{k+i}y_k^t\right]$, with an obvious definition of $\mathbf{E}_\mathbf{j}[.] = \lim_{j\to\infty}\frac{1}{j}[.]$. So, due to ergodicity and the infinite number of data at our disposition, we can replace the expectation operator $\mathbf{E}$ (average over an infinite number of experiments) with the different operator $\mathbf{E}_\mathbf{j}$ applied to the sum of variables (average over one, infinitely long, experiment).

The classical realization approach now follows directly from equations (5)–(6). Consider the correlation matrix between the future block output Hankel and the past block output Hankel matrix:

$$\mathbf{E}_\mathbf{j}[Y_{i|2i-1}Y_{0|i-1}^t]$$

$$= \begin{pmatrix} \Lambda_i & \Lambda_{i-1} & \Lambda_{i-2} & \cdots & \Lambda_2 & \Lambda_1 \\ \Lambda_{i+1} & \Lambda_i & \Lambda_{i-1} & \cdots & \Lambda_3 & \Lambda_2 \\ \Lambda_{i+2} & \Lambda_{i+1} & \Lambda_i & \cdots & \Lambda_4 & \Lambda_3 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \Lambda_{2i-1} & \Lambda_{2i-2} & \Lambda_{2i-3} & \cdots & \Lambda_{i+1} & \Lambda_i \end{pmatrix}.$$

$$= \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{i-1} \end{pmatrix} \times (A^{i-1}G \quad A^{i-2}G \quad \cdots \quad AG \quad G),$$

$$= \mathcal{O}_i\mathcal{C}_i \quad \text{(say)}. \tag{11}$$

Hence we obtain, as $j\to\infty$, a rank deficient block Toeplitz matrix. Its rank is equal to the state dimension n and a model for A, G and C can be obtained from the factorization into an extended observability and controllability matrix using for instance the approaches via the SVD as in Zeiger and McEwen (1974) or Kung (1978).

Note that the rank deficiency is guaranteed to hold as long as i is larger than the largest controllability or observability index of the system.

4. ORTHOGONAL PROJECTIONS

In this section, we describe a new approach that avoids the formation of the covariance matrix (11). The identification scheme is based on geometrical insights and in contrast with previously described similar algorithms (Akaike (1975); Arun and Kung (1990); Larimore (1990)) that needed double infinite block Hankel matrices, consistent estimates are obtained with semi-infinite output block Hankel matrices.

For sake of elegance of the proofs, covariance matrices will still be formed in the theoretical motivation that follows. The "square root" algorithm described in Section 8 will avoid this and will thus have far better numerical robustness.

First, we define the orthogonal projection of semi-infinite matrices as follows.

Definition 2. Orthogonal projections of semi-infinite matrices. Given two matrices $P_1 \in \mathfrak{R}^{p_1\times j}$ and $P_2 \in \mathfrak{R}^{p_2\times j}$, with $j\to\infty$. The orthogonal projection of the row space of P_1 onto the row space of P_2 is defined as: $\mathbf{E}_\mathbf{j}[P_1P_2^t](\mathbf{E}_\mathbf{j}[P_2P_2^t])^{-1}P_2 \in \mathfrak{R}^{p_1\times j}$.

Now, we compute the orthogonal projection of the row space of $Y_{i|2i-1}$ (the future) onto the row space of $Y_{0|i-1}$ (the past). It follows from the rank deficiency of the covariance matrix $\mathbf{E}_\mathbf{j}[Y_{i|2i-1}Y_{0|i-1}^t]$ (11) that the row space of this projection will be an n-dimensional subspace of the j-dimensional ambient space. This can also be seen from

$$\begin{aligned}\mathbf{E}_\mathbf{j}[Y_{i|2i-1}Y_{0|i-1}^t](\mathbf{E}_\mathbf{j}[Y_{0|i-1}Y_{0|i-1}^t])^{-1}Y_{0|i-1} &= \mathcal{O}_i(A^{i-1}G\cdots AG\ G)L_i^{-1}Y_{0|i-1} \\ &= \mathcal{O}_i\mathcal{C}_iL_i^{-1}Y_{0|i-1},\end{aligned}$$

where we define

$$\begin{aligned}L_i &\overset{\text{def}}{=} \mathbf{E}_\mathbf{j}[Y_{0|i-1}Y_{0|i-1}^t] \\ &= \begin{pmatrix} \Lambda_0 & \Lambda_{-1} & \Lambda_{-2} & \cdots & \Lambda_{1-i} \\ \Lambda_1 & \Lambda_0 & \Lambda_{-1} & \cdots & \Lambda_{2-i} \\ \Lambda_2 & \Lambda_1 & \Lambda_0 & \cdots & \Lambda_{3-i} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \Lambda_{i-1} & \Lambda_{i-2} & \Lambda_{i-3} & \cdots & \Lambda_0 \end{pmatrix} \\ &= \mathbf{E}_\mathbf{j}[Y_{i|2i-1}Y_{i|2i-1}^t].\end{aligned}$$

The last equality is due to stationarity. Hence, a basis for the n-dimensional projection of the row space of $Y_{i|2i-1}$ onto that of $Y_{0|i-1}$ is formed by the rows of the matrix Z_i:

$$Z_i = \mathcal{C}_iL_i^{-1}Y_{0|i-1}. \tag{12}$$

In a similar way, we find for the orthogonal projection of the row space of $Y_{0|i-1}$ onto the row space of $Y_{i|2i-1}$:

$$\begin{aligned}&\boldsymbol{E}_j[Y_{0|i-1}Y_{i|2i-1}^t](\mathbf{E}_\mathbf{j}[Y_{i|2i-1}Y_{i|2i-1}^t])^{-1}Y_{i|2i-1} \\ &\quad = \mathcal{C}_i^t\mathcal{O}_i^tL_i^{-1}Y_{i|2i-1}.\end{aligned}$$

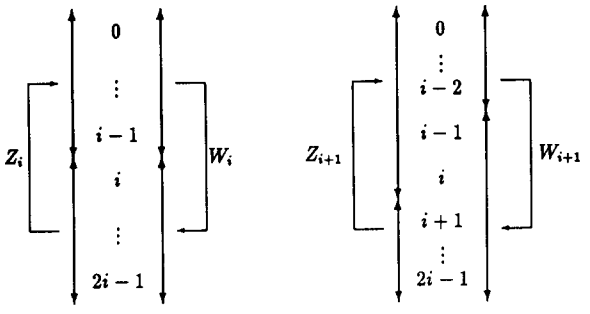

FIG. 2. Definition of the vector sequences Z_i, W_i, Z_{i+1} and W_{i+1}. The thick lines represent the span of the data, the thin lines represent the projections: "from–onto".

Hence, a basis for the n-dimensional projection of the row space of $Y_{0|i-1}$ onto that of $Y_{i|2i-1}$ is formed by the rows of the matrix W_i:

$$W_i = \mathcal{O}_i^t L_i^{-1} Y_{i|2i-1}. \tag{13}$$

The conclusion is that both the orthogonal projections of the row space of $Y_{i|2i-1}$ onto that of $Y_{0|i-1}$ and the other way around, are finite dimensional subspaces of dimension n, with a lot of structure. Of course, in an infinite dimensional setting (i.e. $i \to \infty$ and $j \to \infty$) this result has been known since the pioneering work of Akaike (1975). In this framework and also the one presented in Pal (1982), those finite dimensional orthogonal projections can be considered as states of the stochastic process. Hence, it is tempting to consider the sequence Z_i (12) as the "forward" state and the sequence W_i (13) as the "backward" state sequences. However, we will see in Section 5 that for finite i, these sequences are just an optimal Kalman prediction of the states based upon the i previous outputs.

To arrive at a consistent identification scheme, we also need to define:

- $\mathcal{O}_{i-1}\mathcal{C}_{i+1}L_{i+1}^{-1}Y_{0|i}$ which is the orthogonal projection of the row space of $Y_{i+1|2i-1}$ onto that of $Y_{0|i}$. A basis for the row space of this orthogonal projection is generated by the rows of

$$Z_{i+1} = \mathcal{C}_{i+1}L_{i+1}^{-1}Y_{0|i}. \tag{14}$$

- $\mathcal{C}_{i-1}^t\mathcal{O}_{i+1}^t L_{i+1}^{-1}Y_{i-1|2i-1}$ which is the orthogonal projection of the row space of $Y_{0|i-2}$ onto that of $Y_{i-1|2i-1}$. A basis for the row space of this projection is generated by the rows of

$$W_{i+1} = \mathcal{O}_{i+1}^t L_{i+1}^{-1}Y_{i-1|2i-1}. \tag{15}$$

Figure 2 illustrates these projections graphically.

5. RELATION WITH THE KALMAN FILTER

In this section, we show that the sequences Z_i and Z_{i+1} can be considered as the outputs of a bank of non-steady state Kalman filters after i, respective $i+1$ time steps. We also show that the covariance matrices of these sequences are two iterates of the forward Riccati difference equation.

Theorem 1. Given the sequences $Z_i = \mathcal{C}_i L_i^{-1} Y_{0|i-1}$ and $Z_{i+1} = \mathcal{C}_{i+1}L_{i+1}^{-1}Y_{0|i}$, then:

- $\Sigma_i = \mathbf{E_j}(Z_i Z_i^t)$ and $\Sigma_{i+1} = \mathbf{E_j}(Z_{i+1}Z_{i+1}^t)$ satisfy the forward Riccati difference equation: $\Sigma_{i+1} = A\Sigma_i A^t + (A\Sigma_i C^t - G)(\Lambda_0 - C\Sigma_i C^t)^{-1}(C\Sigma_i A^t - G^t)$.
- $Z_{i+1} = AZ_i + K_i(Y_{i|i} - CZ_i)$ with $K_i = (G - A\Sigma_i C^t)(\Lambda_0 - C\Sigma_i C^t)^{-1}$ equal to the Kalman gain.

Proof. From the definition of the sequences $Z_i = \mathcal{C}_i L_i^{-1} Y_{0|i-1}$ and $Z_{i+1} = \mathcal{C}_{i+1}L_{i+1}^{-1}Y_{0|i}$ we find that $\Sigma_i = \mathbf{E_j}(Z_i Z_i^t) = \mathcal{C}_i L_i^{-1}\mathcal{C}_i^t$ and $\Sigma_{i+1} = \mathbf{E_j}(Z_{i+1}Z_{i+1}^t) = \mathcal{C}_{i+1}L_{i+1}^{-1}\mathcal{C}_{i+1}^t$. If we rewrite Σ_{i+1} with the help of the matrix inversion lemma, we get:

$$\begin{aligned}
\Sigma_{i+1} &= \mathcal{C}_{i+1}L_{i+1}^{-1}\mathcal{C}_{i+1}^t \\
&= (A\mathcal{C}_i \quad G)\begin{pmatrix} L_i^t & \mathcal{C}_i^t C^t \\ C\mathcal{C}_i & \Lambda_0 \end{pmatrix}^{-1}\begin{pmatrix} \mathcal{C}_i^t A^t \\ G^t \end{pmatrix} \\
&= (A\mathcal{C}_i \quad G)\begin{pmatrix} L_i^{-1} + L_i^{-1}\mathcal{C}_i^t C^t \Delta^{-1} C\mathcal{C}_i L_i^{-1} & -L_i^{-1}\mathcal{C}_i^t C^t \Delta^{-1} \\ -\Delta^{-1}C\mathcal{C}_i L_i^{-1} & \Delta^{-1} \end{pmatrix}\begin{pmatrix} \mathcal{C}_i^t A^t \\ G^t \end{pmatrix} \\
&\quad \text{with } \Delta = \Lambda_0 - C\mathcal{C}_i L_i^{-1}\mathcal{C}_i^t C^t \\
&= (A\mathcal{C}_i \quad G)\begin{pmatrix} L_i^{-1}\mathcal{C}_i^t A^t + L_i^{-1}\mathcal{C}_i^t C^t \Delta^{-1} C\mathcal{C}_i L_i^{-1}\mathcal{C}_i^t A^t - L_i^{-1}\mathcal{C}_i^t C^t \Delta^{-1} G^t \\ -\Delta^{-1}C\mathcal{C}_i L_i^{-1}\mathcal{C}_i^t A^t + \Delta^{-1}G^t \end{pmatrix} \\
&= A\mathcal{C}_i L_i^{-1}\mathcal{C}_i^t A^t + A\mathcal{C}_i L_i^{-1}\mathcal{C}_i^t C^t \Delta^{-1} C\mathcal{C}_i L_i^{-1}\mathcal{C}_i^t A^t - A\mathcal{C}_i L_i^{-1}\mathcal{C}_i^t C^t \Delta^{-1} G^t \\
&\quad - G\Delta^{-1}C\mathcal{C}_i L_i^{-1}\mathcal{C}_i^t A^t + G\Delta^{-1}G^t \\
&= A\Sigma_i A^t + A\Sigma_i C^t \Delta^{-1} C\Sigma_i A^t - A\Sigma_i C^t \Delta^{-1} G^t - G\Delta^{-1}C\Sigma_i A^t + G\Delta^{-1}G^t \\
&= A\Sigma_i A^t + (A\Sigma_i C^t - G)\Delta^{-1}(C\Sigma_i A^t - G^t),
\end{aligned}$$

and with $\Delta = \Lambda_0 - C\Sigma_i C^t$, we find:

$$\Sigma_{i+1} = A\Sigma_i A^t + (A\Sigma_i C^t - G)(\Lambda_0 - C\Sigma_i C^t)^{-1} \times (C\Sigma_i A^t - G^t),$$

which is one step of the forward difference Riccati equation. The Kalman gain associated with this equation at time step i is: $K_i = (G - A\Sigma_i C^t)(\Lambda_0 - C\Sigma_i C^t)^{-1}$. We prove now that Z_{i+1} is formed out of Z_i with the help of this Kalman gain. First, we can write (with N_i and M_i unknown)

$$\begin{aligned}
Z_{i+1} &= N_i Z_i + M_i Y_{i|i} \\
\mathscr{C}_{i+1} L_{i+1}^{-1} Y_{0|i} &= N_i \mathscr{C}_i L_i^{-1} Y_{0|i-1} + M_i Y_{i|i}, \\
\mathscr{C}_{i+1} L_{i+1}^{-1} \mathbf{E_j}[Y_{0|i} Y_{0|i-1}^t] L_i^{-1} \mathscr{C}_i^t &= N_i \mathscr{C}_i L_i^{-1} \mathscr{C}_i^t + M_i \mathbf{E_j}[Y_{i|i} Y_{0|i-1}^t] L_i^{-1} \mathscr{C}_i^t, \\
A \mathscr{C}_i L_i^{-1} \mathscr{C}_i^t &= N_i \mathscr{C}_i L_i^{-1} \mathscr{C}_i^t + M_i (\Lambda_i \quad \Lambda_{i-1} \quad \cdots \quad \Lambda_1) L_i^{-1} \mathscr{C}_i^t, \\
A \mathscr{C}_i L_i^{-1} \mathscr{C}_i^t &= N_i \mathscr{C}_i L_i^{-1} \mathscr{C}_i^t + M_i C \mathscr{C}_i L_i^{-1} \mathscr{C}_i^t, \\
A\Sigma_i &= N_i \Sigma_i + M_i C \Sigma_i, \\
A &= N_i + M_i C.
\end{aligned} \tag{16}$$

So, $N_i = A - M_i C$, and we can rewrite (16) as:

$$\begin{aligned}
Z_{i+1} &= AZ_i + M_i(Y_{i|i} - CZ_i), \\
\mathscr{C}_{i+1} L_{i+1}^{-1} \mathbf{E_j}[Y_{0|i} Y_{i|i}^t] &= A \mathscr{C}_i L_i^{-1} \mathbf{E_j}[Y_{0|i-1} Y_{i|i}^t] + M_i(\mathbf{E_j}[Y_{i|i} Y_{i|i}^t] - C \mathscr{C}_i L_i^{-1} \mathbf{E_j}[Y_{0|i-1} Y_{i|i}^t]) \\
\mathscr{C}_{i+1} L_{i+1}^{-1} \begin{pmatrix} \Lambda_i^t \\ \Lambda_{i-1}^t \\ \cdots \\ \Lambda_0 \end{pmatrix} &= A \mathscr{C}_i L_i^{-1} \begin{pmatrix} \Lambda_i^t \\ \Lambda_{i-1}^t \\ \cdots \\ \Lambda_1^t \end{pmatrix} + M_i \left(\Lambda_0 - C \mathscr{C}_i L_i^{-1} \begin{pmatrix} \Lambda_i^t \\ \Lambda_{i-1}^t \\ \cdots \\ \Lambda_1^t \end{pmatrix} \right) \\
\mathscr{C}_{i+1} (0 \quad 0 \quad \cdots \quad I)^t &= A\Sigma_i C^t + M_i(\Lambda_0 - C\Sigma_i C^t), \\
G &= A\Sigma_i C^t + M_i(\Lambda_0 - C\Sigma_i C^t).
\end{aligned}$$

So $M_i = (G - A\Sigma_i C^t)(\Lambda_0 - C\Sigma_i C^t)^{-1} = K_i$ which is exactly the Kalman gain. □

So, if we start from $Z_0 = 0$ and $\Sigma_0 = 0$, we find $Z_1 = K_1 Y_{0|0} = G\Lambda_0^{-1} Y_{0|0} = \mathscr{C}_1 L_1^{-1} Y_{0|0}$. Through induction, this clearly proves that the sequence $Z_0, Z_1, \ldots, Z_i, Z_{i+1}$ is a sequence of subsequent optimal Kalman state estimates. This has to be interpreted as follows: the lth column of Z_i is the optimal prediction of the state x_{l+i-1} based on the measurements $y_{l-1}, y_l, \ldots, y_{l+i-2}$ and with initial conditions for the predicted state: $\hat{x}_{l-1} = 0$, $\mathbf{E}(\hat{x}_{l-1} \hat{x}_{l-1}^t) = 0$. Figure 3 illustrates this graphically. So, in this way every column of Z_i represents an optimal state estimate. And we can interpret the columns of Z_i as the outputs of a bank of j Kalman filters that estimate j optimal states in parallel.

It should be noted that a larger past (larger i) will lead to a better suppression of the influence of the error on the initial state Z_0, and thus to a smaller variance on the final estimates of the system matrices. This is because the initial state at i lags in the past is set to zero, which is correct on the average, but the true initial state value produces "noise" in the forward and backward estimates of the states (see also Section 9 for an illustration of this effect).

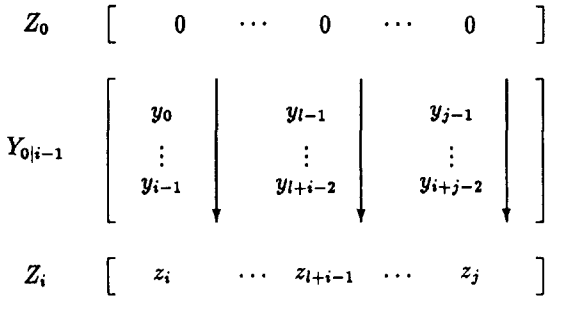

FIG. 3. Interpretation of the sequence Z_i as Kalman filter state estimates based upon i measurements.

6. HOW TO FIND A, C, G AND Λ_0

In this section, we show how the system matrices A, C, G and Λ_0 can be identified consistently. All this provided that the matrices Z_i (12), W_i (13), Z_{i+1} (14) and W_{i+1} (15) are available.

Theorem 2. Given the sequences Z_i, Z_{i+1}, W_i and W_{i+1}. The matrices A, C, G and Λ_0 satisfy:

$$A = \mathbf{E}_j[Z_{i+i}Z_i^t](\mathbf{E}_j[Z_iZ_i^t])^{-1}, \tag{17}$$

$$A^t = \mathbf{E}_j[W_{i+1}W_i^t](\mathbf{E}_j[W_iW_i^t])^{-1}, \tag{18}$$

$$C = \mathbf{E}_j[Y_{i|i}Z_i^t](\mathbf{E}_j[Z_iZ_i^t])^{-1}, \tag{19}$$

$$G^t = \mathbf{E}_j[Y_{i-1|i-1}W_i^t](\mathbf{E}_j[W_iW_i^t])^{-1}, \tag{20}$$

$$\Lambda_0 = \mathbf{E}_j[Y_{i|i}Y_{i|i}^t]. \tag{21}$$

Proof of equation (17).

$$\begin{aligned}
&\mathbf{E}_j[Z_{i+1}Z_i^t](\mathbf{E}_j[Z_iZ_i^t])^{-1} \\
&\quad = \mathscr{C}_{i+1}L_{i+1}^{-1}\mathbf{E}_j[Y_{0|i}Y_{0|i-1}^t]L_i^{-1}\mathscr{C}_i^t \\
&\qquad \times (\mathscr{C}_iL_i^{-1}\mathbf{E}_j[Y_{0|i-1}Y_{0|i-1}^t]L_i^{-1}\mathscr{C}_i^t)^{-1} \\
&\quad = \mathscr{C}_{i+1}\begin{pmatrix} I & 0 & \cdots & 0 \\ 0 & I & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & I \\ 0 & 0 & \cdots & 0 \end{pmatrix} \\
&\qquad \times L_i^{-1}\mathscr{C}_i^t(\mathscr{C}_iL_i^{-1}\mathscr{C}_i^t)^{-1} \\
&\quad = A\mathscr{C}_iL_i^{-1}\mathscr{C}_i^t(\mathscr{C}_iL_i^{-1}\mathscr{C}_i^t)^{-1} \\
&\quad = A.
\end{aligned}$$

The proof of equation (18) is similar.

Proof of equation (19).

$$\begin{aligned}
&\mathbf{E}_j[Y_{i|i}Z_i^t](\mathbf{E}_j[Z_iZ_i^t])^{-1} \\
&\quad = \mathbf{E}_j[Y_{i|i}Y_{0|i-1}^t]L_i^{-1}\mathscr{C}_i^t(\mathscr{C}_iL_i^{-1}\mathscr{C}_i^t)^{-1} \\
&\quad = (\Lambda_i \quad \Lambda_{i-1} \quad \cdots \quad \Lambda_1)L_i^{-1}\mathscr{C}_i^t(\mathscr{C}_iL_i^{-1}\mathscr{C}_i^t)^{-1} \\
&\quad = C\mathscr{C}_iL_i^{-1}\mathscr{C}_i^t(\mathscr{C}_iL_i^{-1}\mathscr{C}_i^t)^{-1} \\
&\quad = C.
\end{aligned}$$

The proof of (20) is similar. Equation (21) directly follows from the definition of Λ_0 and of $\mathbf{E}_j$. Hence we have a general outline for finding A, G, C and Λ_0 from output data, if Z_i, Z_{i+1}, W_i and W_{i+1} were available.

Let us as a last remark point out that the matrix A also satisfies the following Sylvester equation.

$$\begin{aligned}
A\mathbf{E}_j[Z_iZ_i^t] + \mathbf{E}_j[W_iW_i^t]A \\
= \mathbf{E}_j[Z_{i+1}Z_i^t] + \mathbf{E}_j[W_iW_{i+1}^t].
\end{aligned}$$

7. THE STATES AS PRINCIPAL DIRECTIONS

In this section we show that the matrices Z_i and W_i are generated from the principal directions between the past and future output data (up to a diagonal scaling). The matrices Z_{i+1} and W_{i+1} are also shown to be equal to principal directions, up to within a similarity transformation.

7.1. *Principal angles and directions*

Definition 3. Principal angles and directions. Given two matrices $U \in \mathfrak{R}^{p\times j}$ and $V \in \mathfrak{R}^{q\times j}$, $(j\to\infty)$ of full row rank and the Singular Value Decomposition (SVD):†

$$U^t(\mathbf{E}_j[UU^t])^{-1}\mathbf{E}_j[UV^t](\mathbf{E}_j[VV^t])^{-1}V = P^tSQ. \tag{22}$$

Then the principal directions between the row spaces of U and V are defined as the left (P) and right (Q) singular vectors (the principal directions in the row space of U, respectively V). The cosines of the principal angles between the row spaces of U and V are defined as the singular values (the diagonal of S). The principal directions and angles between the row spaces of U and V will be denoted as:

$$\mathscr{P}[U, V] = [P, S, Q].$$

Geometrical motivation. The principal angles between two subspaces are a generalization of an angle between two vectors (the concept goes back to Jordan (1875)). Suppose we are given two matrices $U \in \mathfrak{R}^{p\times j}$ and $V \in \mathfrak{R}^{q\times j}$. The first principal angle θ_1 (the smallest one) is obtained as follows: Choose unit vectors $u_1 \in \mathbf{R}_{\text{row}}(U)$ and $v_1 \in \mathbf{R}_{\text{row}}(V)$ and minimize the angle between them.‡ Next choose a unit vector $u_2 \in \mathbf{R}_{\text{row}}(U)$ orthogonal to u_1 and $v_2 \in \mathbf{R}_{\text{row}}(V)$ orthogonal to v_1 and minimize the angle θ_2 between them. This is the second principal angle. Continue in this way until $\min(p, q)$ angles have been found. This informal description can also be formalized.

Definition 4. Principal angles and directions. The principal angles $\theta_1 \le \theta_2 \le \cdots \le \pi/2$ between the row spaces $\mathbf{R}_{\text{row}}(U)$ and $\mathbf{R}_{\text{row}}(V)$ of two matrices $U \in \mathfrak{R}^{p\times j}$ and $V \in \mathfrak{R}^{q\times j}$ and the corresponding principal directions $u_i \in \mathbf{R}_{\text{row}}(U)$ and $v_i \in \mathbf{R}_{\text{row}}(V)$ are defined recursively as

$$\cos\theta_k = \max_{u\in\mathbf{R}_{\text{row}}(U),\ v\in\mathbf{R}_{\text{row}}(V)} u^tv = u_k^tv_k,$$

subject to $\|u\| = \|v\| = 1$ and for $k > 1$, $u^tu_i = 0$ for $i = 1, \ldots, k-1$ and $v^tv_i = 0$ for $i = 1, \ldots, k-1$.

We will show in Section 8 how the principal angles and directions can be calculated efficiently.

† The Singular Value Decomposition is written here as $A = P^tSQ$ instead of $A = PSQ^t$ for convenience of notation.

‡ The notation $\mathbf{R}_{\text{row}}(.)$ refers to the row space of the matrix between brackets.

7.2. *Determination of* Z_i, W_i, Z_{i+1} *and* W_{i+1}

Theorem 3.

Given (with $j \to \infty$):

$$\mathcal{P}[Y_{0|i-2}, Y_{i-1|2i-1}] = [P_{i+1}, S_{i+1}, Q_{i+1}],$$
$$\mathcal{P}[Y_{0|i-1}, Y_{i|2i-1}] = [P_i, S_i, Q_i],$$
$$\mathcal{P}[Y_{0|i}, Y_{i+1|2i-1}] = [P_{i-1}, S_{i-1}, Q_{i-1}].$$

Then:

- The number of elements in S_i different from zero is equal to the state dimension n. The same holds for S_{i-1} and S_{i+1}. This means that there are only n principal angles different from $\pi/2$.
  Denote with P_{i+1}^1, P_i^1, P_{i-1}^1, Q_{i+1}^1, Q_i^1, Q_{i-1}^1 the matrices with the first n principal directions, and with S_{i+1}^1, S_i^1, S_{i-1}^1 the $n \times n$ matrices with elements from S_{i+1}, S_i, S_{i-1} (not zero) on the diagonals.
- Then, a possible choice for the state matrices is given by:§

$$Z_i = (S_i^1)^{1/2} P_i^1, \tag{23}$$

$$W_i = (S_i^1)^{1/2} Q_i^1, \tag{24}$$

$$Z_{i+1} = [(\underline{\mathcal{O}_i})^\dagger \bar{\mathcal{O}}_{i-1}] \times (S_{i-1}^1)^{1/2} P_{i-1}^1, \tag{25}$$

$$W_{i+1} = [(\overline{\mathscr{C}_i^t})^\dagger (\bar{\mathscr{C}}_{i-1})^t] \times (S_{i+1}^1)^{1/2} Q_{i+1}^1. \tag{26}$$

With

$$\mathcal{O}_i \overset{\text{def}}{=} \mathbf{E_j}[Y_{i|2i-1}(P_i^1)^t (S_i^1)^{-1/2}], \tag{27}$$

$$\mathscr{C}_i^t \overset{\text{def}}{=} \mathbf{E_j}[Y_{0|i-1}(Q_i^1)^t (S_i^1)^{-1/2}], \tag{28}$$

$$\bar{\mathcal{O}}_{i-1} \overset{\text{def}}{=} \mathbf{E_j}[Y_{i+1|2i-1}(P_{i-1}^1)^t \times (S_{i-1}^1)^{-1/2}], \tag{29}$$

$$\bar{\mathscr{C}}_{i-1}^t \overset{\text{def}}{=} \mathbf{E_j}[Y_{0|i-2}(Q_{i+1}^1)^t \times (S_{i+1}^1)^{-1/2}]. \tag{30}$$

Proof. Since the rank of $\mathbf{E_j}[Y_{i|2i-1} Y_{0|i-1}^t]$ is equal to n (see (11)), we know that the left hand side of (22) will be of rank n. This implies that there are n singular values different from zero, which proves the fact that there are only n principal angles different from $\pi/2$. From (12) we know that Z_i can be chosen to be any basis for the projection of the row space of $Y_{i|2i-1}$ onto the row space of $Y_{0|i-1}$. From Definition 3, we know that $(S_i^1)^{1/2} P_i^1$ is such as basis. This proves equation (23).

In a similar way as we proved equation (19), it is easy to derive that:

$$\mathcal{O}_i = \mathbf{E_j}[Y_{i|2i-1} Z_i^t](\mathbf{E_j}[Z_i Z_i^t])^{-1}. \tag{31}$$

Together with $\mathbf{E_j}[P_i^1 (P_i^1)^t] = I$, equation (27) follows easily. Following the same reasoning, equations (24) and (28) can be derived. Equation (25) can be derived as follows. We know from (14) that Z_{i+1} is equal to any basis of the n-dimensional projection of the row space of $Y_{i+1|2i-1}$ onto the row space of $Y_{0|i}$. Once again, we know from Definition 3 that $M_{i-1}(S_{i-1}^1)^{1/2} P_{i-1}^1$ is a valid basis, with M_{i-1} an arbitrary similarity transform. In the same way as above, we derive that the $\mathcal{O}_{i-1}$ corresponding to the choice of this basis is equal to:

$$\begin{aligned}\mathcal{O}_{i-1} &= \mathbf{E_j}[Y_{i+1|2i-1}(P_{i-1}^1)^t \times (S_{i-1}^1)^{1/2} M_{i-1}^t] \times (M_{i-1} S_{i-1}^1 M_{i-1}^t)^{-1} \\ &= \mathbf{E_j}[Y_{i+1|2i-1}(P_{i-1}^1)^t \times (S_{i-1}^1)^{-1/2}] M_{i-1}^{-1} \\ &= \bar{\mathcal{O}}_{i-1} M_{i-1}^{-1}.\end{aligned}$$

If we strip the last l rows from $\mathcal{O}_i$ in (27), we must find the same matrix $\mathcal{O}_{i-1}$. Thus:

$$\bar{\mathcal{O}}_{i-1} M_{i-1}^{-1} = \underline{\mathcal{O}_i}.$$

This leads to $M_{i-1} = (\underline{\mathcal{O}_i})^\dagger \bar{\mathcal{O}}_{i-1}$. Which proves equation (25). Equation (26) can be proven in a similar way.

8. A NUMERICALLY EFFICIENT ALGORITHM

We have derived in Section 6 that the matrices A and C could be estimated from the matrices Z_i and Z_{i+1} while A^t and G^t could be estimated from W_i and W_{i+1}. In this section, we show how these matrices can be computed starting from the output data and without explicit calculation of the output covariance matrices.

Up to now, we have been working with an infinite amount of data ($j \to \infty$). However, for all practical applications the number of data is finite. This is why from now on, we will replace the operator $\mathbf{E_j} = \lim_{j \to \infty} \frac{1}{j}[.]$ by $\frac{1}{j}[.]$. The orthogonal projection as defined in Definition 2 thus becomes a normal orthogonal projection of finite size matrices:

$$\frac{1}{j}[UV^t]\left(\frac{1}{j}[VV^t]\right)^{-1} V = UV^t(VV^t)^{-1}V.$$

8.1. *Principal directions and the quotient singular value decomposition*

In this subsection, we derive an elegant algorithm to compute principal angles and directions, which is based on a combination of

§ $A^\dagger$ denotes the Moore–Penrose pseudo-inverse. The matrices $\underline{A}$, $\bar{A}$ denote the matrix A with, respectively the last l and the first l rows omitted.

the QR-decomposition and the QSVD. The QSVD is a generalization for two matrices of the SVD for one matrix.

Lemma 1. Any pair of matrices $A \in \mathfrak{R}^{m\times n}$ and $B \in \mathfrak{R}^{p\times n}$ can be decomposed as

$$A = U_a S_a X^t,$$
$$B = U_b S_b X^t,$$

where $U_a \in \mathfrak{R}^{m\times m}$, $U_b \in \mathfrak{R}^{p\times p}$ are orthonormal, $X \in \mathfrak{R}^{n\times n}$ is square non-singular and $S_a \in \mathfrak{R}^{m\times n}$ and $S_b \in \mathfrak{R}^{p\times n}$ have the following quasi-diagonal structure:

$$S_a = \begin{matrix} \\ r_{ab}-r_b \\ r_a+r_b-r_{ab} \\ m-r_a \end{matrix} \begin{pmatrix} \overset{r_{ab}-r_b}{I} & \overset{r_a+r_b-r_{ab}}{0} & \overset{r_{ab}-r_a}{0} & \overset{n-r_{ab}}{0} \\ 0 & D_a & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$S_b = \begin{matrix} \\ p-r_b \\ r_a+r_b-r_{ab} \\ r_{ab}-r_a \end{matrix} \begin{pmatrix} \overset{r_{ab}-r_b}{0} & \overset{r_a+r_b-r_{ab}}{0} & \overset{r_{ab}-r_a}{0} & \overset{n-r_{ab}}{0} \\ 0 & D_b & 0 & 0 \\ 0 & 0 & I & 0 \end{pmatrix},$$

where

$$r_{ab} = \operatorname{rank}\begin{pmatrix} A \\ B \end{pmatrix},$$

and

$$D_a^2 + D_b^2 = I_{r_a+r_b-r_{ab}}.$$

The original formulation of the QSVD (which used to be called the generalized SVD, but see also De Moor and Golub (1989)) is due to Paige and Saunders (1981) and Van Loan (1976). We are now ready to derive the following theorem (which to our knowledge is new).

Theorem 4. Consider the QR-decomposition of the concatenated matrix

$$\begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} R_{11} & 0 \\ R_{21} & R_{22} \end{pmatrix}\begin{pmatrix} Q_1^t \\ Q_2^t \end{pmatrix},$$

where both A and B are of full row rank and consider a QSVD of the matrix pair

$$R_{21}^t = USX^t,$$
$$R_{22}^t = VTX^t,$$

where U, V are orthonormal, S and T quasi-diagonal and X square nonsingular. Then, the principal angles and directions between the row spaces of A and B are given by:

$$\mathcal{P}[A, B] = [(Q_1U)^t, S, (Q_1US + Q_2VT)^t],$$

Proof. Because A and B are of full row rank, we have

$$A^t(AA^t)^{-1}A = Q_1Q_1^t,$$
$$B^t(BB^t)^{-1}B = (Q_1US + Q_2VT) \times (S^tU^tQ_1^t + T^tV^tQ_2^t).$$

It follows from the properties of the QSVD that $S^tS + T^tT = I$ and hence $Q_1US + Q_2VT$ is an orthonormal matrix. We find that

$$A^t(AA^t)^{-1}AB^t(BB^t)^{-1}B = (Q_1U)S(Q_1US + Q_2VT)^t,$$

which is clearly an SVD. It follows from Definition 3 that it contains the principal angles and directions.

The major advantage of this result lies in the fact that principal angles and directions can basically be computed from blocks of the R-factor of a QR-decomposition. This R-factor consists of matrices of which the dimensions are small compared to j.

It should be pointed out that a different way for computing principal angles and directions, without forming covariance matrices, was developed by Björck and Golub (1973).

8.2. *Algorithm*

Consider the QR-decomposition of the output block Hankel matrix $Y_{0|2i-1}$ with the following appropriate partitioning (which is also visualized in Fig. 2):

$$\frac{Y_{0|2i-1}}{\sqrt{j}} = \begin{matrix} \\ l(i-1) \\ l \\ l \\ l(i-1) \end{matrix} \begin{pmatrix} \overset{l(i-1)}{R_{11}} & \overset{l}{0} & \overset{l}{0} & \overset{l(i-1)}{0} \\ R_{21} & R_{22} & 0 & 0 \\ R_{31} & R_{32} & R_{33} & 0 \\ R_{41} & R_{42} & R_{43} & R_{44} \end{pmatrix} \times \begin{pmatrix} Q_1^t \\ Q_2^t \\ Q_3^t \\ Q_4^t \end{pmatrix}. \tag{32}$$

It should be noted that Verhaegen (1991) uses a similar QR-decomposition in his identification algorithms. Now, compute three QSVDs:

(1) $$(R_{41} \quad R_{42} \quad R_{43})^t = U_{i-1}S_{i-1}X_{i-1}^t,$$
$$R_{44}^t = V_{i-1}T_{i-1}X_{i-1}^t,$$

(2) $$\begin{pmatrix} R_{31} & R_{32} \\ R_{41} & R_{42} \end{pmatrix}^t = U_iS_iX_i^t,$$
$$\begin{pmatrix} R_{33} & 0 \\ R_{43} & R_{44} \end{pmatrix}^t = V_iT_iX_i^t,$$

$$(3)\qquad \begin{pmatrix} R_{21} \\ R_{31} \\ R_{41} \end{pmatrix}^t = U_{i+1} S_{i+1} X_{i+1}^t,$$

$$\begin{pmatrix} R_{22} & 0 & 0 \\ R_{32} & R_{33} & 0 \\ R_{42} & R_{43} & R_{44} \end{pmatrix}^t = V_{i+1} T_{i+1} X_{i+1}^t.$$

With n, the state space dimension from the number of principal angles different from $\pi/2$, let X_{i-1}, X_i, X_{i+1}, U_{i-1}, U_i, U_{i+1}, V_{i-1}, V_i and V_{i+1} with superscript 1 denote the first n columns of the respective matrices. Also, let S_{i-1}, S_i, S_{i+1}, T_{i-1}, T_i and T_{i+1} with superscript 1 denote the $n \times n$ upper left sub-matrices of the respective matrices. Now, equations (27)–(30) can be evaluated, and after some straightforward calculations, we find:‡

$$\begin{aligned} \mathcal{O}_i &\overset{j}{=} X_i^1 (S_i^1)^{1/2}, \\ \mathcal{C}_i^t &\overset{j}{=} R_{1:2,1:2} U_i^1 (S_i^1)^{1/2}, \\ \tilde{\mathcal{O}}_{i-1} &\overset{j}{=} X_{i-1}^1 (S_{i-1}^1)^{1/2}, \\ \tilde{\mathcal{C}}_{i-1}^t &\overset{j}{=} R_{1:1,1:1} U_{i+1}^1 (S_{i+1}^1)^{1/2}. \end{aligned}$$

The similarity transformation of equation (25) can be calculated as follows:

$$[(\underline{\mathcal{O}_i})^\dagger \tilde{\mathcal{O}}_{i-1}] \overset{j}{=} (S_i^1)^{-1/2} (\underline{X_i^1})^\dagger X_{i-1}^1 (S_{i-1}^1)^{1/2}.$$

The state sequences can be found from equations (23)–(26) as:

$$Z_i \overset{j}{=} (S_i^1)^{1/2} (U_i^1)^t Q_{1:2}^t, \tag{33}$$

$$\begin{aligned} W_i &\overset{j}{=} (S_i^1)^{1/2} [(U_i^1 S_i^1)^t Q_{1:2}^t \\ &\quad + (V_i^1 T_i^1)^t Q_{3:4}^t], \end{aligned} \tag{34}$$

$$\begin{aligned} Z_{i+1} &\overset{j}{=} (S_i^1)^{-1/2} (\underline{X_i^1})^\dagger X_{i-1}^1 \\ &\quad \times S_{i-1}^1 (U_{i-1}^1)^t Q_{1:3}^t, \\ W_{i+1} &\overset{j}{=} [(\overline{\mathcal{C}_i})^\dagger \tilde{\mathcal{C}}_{i-1}^t] (S_{i+1}^1)^{1/2} \\ &\quad \times [(U_{i+1}^1 S_{i+1}^1)^t Q_{1:1}^t \\ &\quad + (V_{i+1}^1 T_{i+1}^1)^t Q_{2:4}^t]. \end{aligned}$$

Finally, from equations (17)–(21), we find the system matrices as:

$$\begin{aligned} A &\overset{j}{=} (S_i^1)^{-1/2} (\underline{X_i^1})^\dagger X_{i-1}^1 S_{i-1}^1 (\underline{U_{i-1}^1})^t U_i^1 (S_i^1)^{-1/2}, \\ C &\overset{j}{=} \text{the first } l \text{ rows of } \mathcal{O}_i \\ &\overset{j}{=} [X_i^1 (S_i^1)^{1/2}]_{\text{first } l \text{ rows}}, \\ G^t &\overset{j}{=} \text{the last } l \text{ rows of } \mathcal{C}_i^t \\ &\overset{j}{=} [R_{1:2,1:2} U_i^1 (S_i^1)^{1/2}]_{\text{last } l \text{ rows}}, \\ \Lambda_0 &\overset{j}{=} R_{3:3,1:3} R_{3:3,1:3}^t. \end{aligned}$$

‡ $R_{k:l,r:s}$ denotes the sub-matrix taken from block row k to block row l, and from block column r to block column s. The blocks are determined from the partitioning defined in (32). $Q_{k:l}^t$ denotes the submatrix of Q^t (in the same RQ decomposition), from block row k to block row l. The symbol $\overset{j}{=}$ means that the equality holds when $j \to \infty$.

The expressions for C and G^t are a logical consequence of (19)–(20) and the way $\mathcal{O}_i$ ($\mathcal{C}_i^t$) is calculated (31). Equation (18) is not repeated here since both ways ((17) and (18)) of calculating A are equivalent (for j large enough).

Note, that in these final expressions, the orthogonal matrices Q_1, Q_2, Q_3 and Q_4 have canceled out, so they do not have to be calculated. Only the R factor has to be calculated. Another observation concerns the "covariance" matrices $Z_i Z_i^t$ and $W_i W_i^t$. From (33)–(34) we find $Z_i Z_i^t = S_i^1 = W_i W_i^t$. Hence, these products are diagonal and equal to each other. So, the algorithm returns stochastic models that are balanced (for finite i). If $i \to \infty$, we have stochastically balanced models as defined in Pal (1982).

9. AN EXAMPLE

9.1. *Taking the same linear relations*

Estimates of the state sequence Z_i are obtained by taking linear combinations (F_i) of the past output matrix $Y_{0|i-1}$ as $Z_i = F_i Y_{0|i-1}$. Now, in earlier work (Akaike (1975); Arun and Kung (1990); Larimore (1990)), the shifted state sequence Z_{i+1} is computed by applying the same linear combinations (F_i) to the shifted Hankel matrix $Y_{1|i}$ as $Z_{i+1} = F_i Y_{1|i}$. While this trick works for purely deterministic systems (see e.g. Moonen *et al.* (1989)), it does not work for stochastic systems for finite i, as our formulas in the previous section clearly indicate. In general, one can state that for small i, the bias for the classical approach (same linear combinations) becomes worse as some of the poles of A move towards the unit circle. Only for $i \to \infty$, this classical approach will be exact. The following easy example illustrates this clearly. Applying formulae (12) and (14) to a block Hankel matrix with, respectively, one and two block rows (SISO system), gives:

$$\begin{aligned} Z_1 &= \mathcal{C}_1 L_1^{-1} Y_{0|0} = G \Lambda_0^{-1} Y_{0|0}, \\ Z_2 &= (AG \quad G) \begin{pmatrix} \Lambda_0 & \Lambda_1 \\ \Lambda_1 & \Lambda_2 \end{pmatrix}^{-1} Y_{0|1}. \end{aligned}$$

We then find that

$$\begin{aligned} Z_2 Z_1^t (Z_1 Z_1^t)^{-1} &= (AG \quad G) \begin{pmatrix} \Lambda_0 & \Lambda_1 \\ \Lambda_1 & \Lambda_2 \end{pmatrix}^{-1} \\ &\quad \times \mathbf{E}_j [Y_{0|1} Y_{0|0}^t] \Lambda_0^{-1} G^t \\ &\quad \times (G \Lambda_0^{-1} G^t)^{-1} \\ &= A. \end{aligned}$$

If on the other hand we take, as is usually done, the same linear combinations for Z_2 as for Z_1 (namely $G\Lambda_0^{-1}$), we find:

$$Z_2 = G \Lambda_0^{-1} Y_{1|1},$$

and

$$Z_2Z_1^t(Z_1Z_1^t)^{-1} = \Lambda_1/\Lambda_0 = CG/(C\Sigma C^t + R),$$

which has nothing to do with A. For a numerical example, consider the very simple SISO system with only one state ($Q = BB^t$, $S = BD^t$, $R = DD^t$):

$$A = 0.9490, \quad B = 0.6574,$$
$$C = 0.1933, \quad D = 0.8981.$$

Figure 4 (top) shows the eigenvalues of the identified A matrix (a scalar), in function of the parameter i ($j = 1000$) for three algorithms:
(a) taking the same linear combinations as described in this section;
(b) the algorithm as described in Section 8;
(c) the algorithm described in Aoki (1987).
For every i, 100 Monte Carlo experiments were performed. The mean value of these 100 experiments is plotted for every i on Fig. 4 (the exact value is 0.949). Clearly the algorithm described in this section (which is the one that is normally used in the literature), calculates heavily biased solutions for small i. The bias is still significant for i, as large as 8 (eight times larger than the dimension of the state vector).

On the other hand, the new algorithm (of Section 8) does not suffer at all from this problem. Neither does the method of Aoki, which is a covariance based method, and is mentioned just for comparison.

Figure 4 (bottom) also shows the variance of the estimated A matrix. Clearly the variance grows smaller when i grows larger (as mentioned in Section 5). It should also be noted that, for small i, the biased algorithm of this section trades off bias for a smaller variance. Even though it is not clear from this example, we found from other experiments that the variance for the algorithm of Aoki tends to be larger than the variance for the algorithm described in this paper.

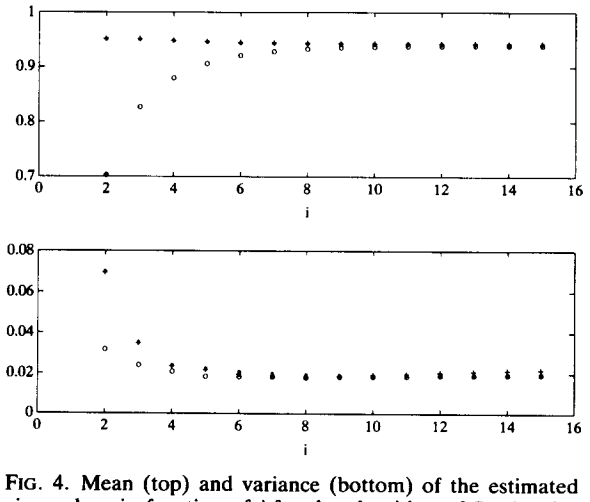

FIG. 4. Mean (top) and variance (bottom) of the estimated eigenvalues in function of i for the algorithm of Section 9.1 (circle), for the algorithm of this paper (star) and for the algorithm of Aoki (cross).

9.2. *An example with undamped poles*

For finite i, the difference between the classical approach and the approach described in this paper becomes more pronounced as A becomes more and more marginally stable. Clearly a limiting case occurs when A has poles exactly on the unit circle. Consider the system:

$$A = \begin{pmatrix} 0.6 & 0.8 & 0 \\ -0.8 & 0.6 & 0 \\ 0 & 0 & 0.8 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$
$$C = \begin{pmatrix} 1 \\ 0.3 \\ 0.5 \end{pmatrix}^t, \quad D = (1),$$

which is a third order system with one output, and complex poles on the unit circle (this is a limiting case of a stable stochastic system). So, there is a pure autonomous sinusoidal mode in the output (initial state $x_0 = (1\ 1\ 0)^t$). The eigenvalues of A are: $0.6 \pm 0.8i$ and 0.8. Once again, we performed 100 Monte Carlo simulations ($i = 4$, $j = 1000$) for the two algorithms. Figure 5 shows the absolute values of the 100 estimates of the poles. The top figure used the algorithm of the previous subsection. The full lines are the absolute values of the exact poles (1 and 0.8). Clearly there is a bias on the estimates of about $0.2 \simeq 20\%$. The second figure shows the estimates with the new algorithm of this paper. The bias is eliminated. The sample mean for these 100 experiments was:

$$m_a = \begin{pmatrix} 0.7993 \\ 0.6733 \end{pmatrix}, \quad m_b = \begin{pmatrix} 0.9983 \\ 0.8080 \end{pmatrix}.$$

The real values are: 1 and 0.8.

10. CONCLUSIONS

In this paper, we have derived a new subspace algorithm for the consistent identification of stochastic state space descriptions, using only semi-infinite block Hankel matrices of output measurements. The explicit formation of the covariance matrix and double infinite block Hankel matrices are avoided. The algorithm is based on principal angles and directions. We have described a numerical robust and fast way to calculate these on the basis of QR and QSVD decomposition. We have also interpreted the principal directions as non-steady state Kalman filter states. With an example, we have illustrated that the new algorithm is superior to the classical canonical correlation algorithms, especially for small i.

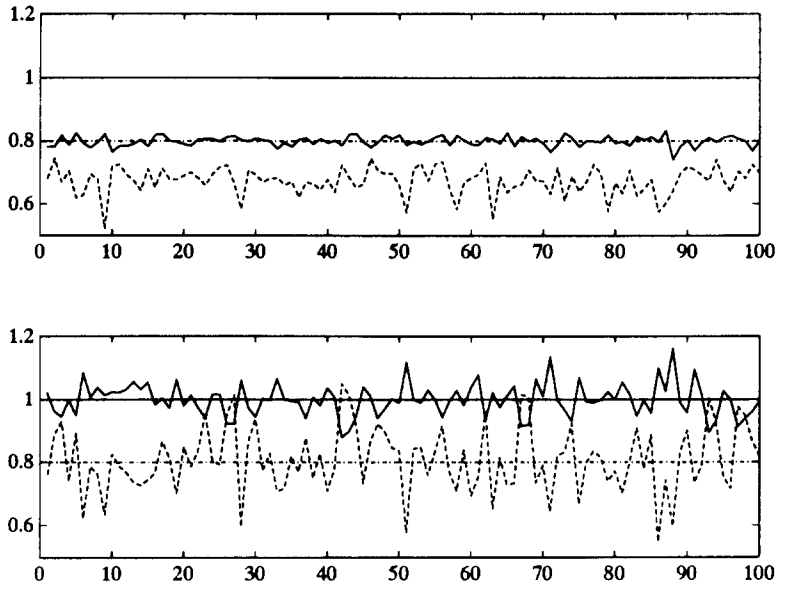

FIG. 5. One hundred estimates of the absolute value of the eigenvalues with the algorithm of Section 9.1 (top). The horizontal lines show the exact results. The bottom figure shows the same for the algorithm of this paper.

REFERENCES

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

Aoki, M. (1987). *State Space Modeling of Time Series.* Springer Verlag, Berlin.

Arun, K. S. and S. Y. Kung (1990). Balanced approximation of stochastic systems. *SIAM J. Matrix Analysis and Applications,* **11,** 42–68.

Björck, A. and G. H. Golub (1973). Numerical methods for computing angles between subspaces. *Numer. Math.*, **27,** 579–594.

De Moor, B. and G. H. Golub (1989). Generalized singular value decompositions: a proposal for a standardized nomenclature. Numerical Analysis Project Report 89-04, Department of Computer Science, Stanford University.

Faurre, P. (1976). Stochastic realization algorithms. In R. Mehra and D. Lainiotis (Eds), *System Identification: Advances and Case Studies.* Academic Press, NY.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika,* **28,** 321–377.

Jordan, C. (1875). Essai sur la géometrie à n dimensions. *Bull. Soc. Math. France,* **3,** 103–174.

Kung, S. Y. (1978). A new identification and model reduction algorithm via singular value decomposition. *Proc. of the 12th Asilomar Conf. on Circuits, Systems and Computers,* Pacific Grove, pp. 705–714.

Larimore, W. E. (1990). Canonical variate analysis in identification, filtering, and adaptive control. *Proc. of the IEEE 29th Conf. on Decision and Control,* HI, pp. 596–604.

Ljung, L. (1987). *System Identification: Theory for the User.* Prentice-Hall information and system sciences series, Englewood Cliffs, NJ.

Moonen, M., B. De Moor, L. Vandenberghe and J. Vandewalle (1989). On- and off-line identification of linear state space models. *Int. J. of Control,* **49,** 219–232.

Paige, C. C. and M. A. Saunders (1981). Towards a generalized singular value decomposition. *SIAM J. Numer. Anal.*, **18,** 398–405.

Pal, D. (1982). Balanced stochastic realization and model reduction. Master's thesis, Washington State University, Electrical Engineering.

Papoulis, A. (1984). *Probability, Random Variables, and Stochastic Processes.* McGraw–Hill, NY.

Van Loan, C. F. (1976). Generalizing the singular value decomposition. *SIAM J. Numer. Anal.*, **13,** 76–83.

Verhaegen, M. (1991). A novel non-iterative MIMO state space model identification technique. *Preprints 9th IFAC/IFORS Symposium on Identification and System Parameter Estimation,* Budapest, Hungary, pp. 1453–1458.

Zeiger, H. P. and A. J. McEwen (1974). Approximate linear realization of given dimension via Ho's algorithm. *IEEE Trans. Aut. Control,* **AC-19,** 153.