SUBSPACE ALGORITHMS FOR THE STOCHASTIC IDENTIFICATION PROBLEM[1]

Peter Van Overschee[2] **Bart De Moor** [3]

ESAT , Department of Electrical Engineering , Katholieke Universiteit Leuven
Kardinaal Mercierlaan 94 , 3001 Leuven (Heverlee) , Belgium
tel: 32/16/220931 fax: 32/16/221855
email: vanovers@esat.kuleuven.ac.be , demoor@esat.kuleuven.ac.be

Abstract

In this paper, we derive a new 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 only QR and **QSVD** decompositions. We also provide an interpretation of the principal directions as states of a non-steady state Kalman filter.

1. Introduction

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

$$x_{k+1} = Ax_k + w_k \quad , \quad y_k = Cx_k + v_k \tag{1}$$

where $x_k \in \Re^n$ is a state vector. Both $w_k \in \Re^n$ and $v_k \in \Re^l$ are assumed to be zero mean, white, gaussian noise sequences with covariance matrix [4]:

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

It is assumed that the stochastic process is stationary, i.e. : $\mathbf{E}[x_k] = 0$ and $\mathbf{E}[x_k x_k^t] = \Sigma$, where the state covariance matrix

Σ is independent of the time k. This implies that A is a stable matrix.
The central problem discussed in this paper is the identification of a state space model from the data y_k and the determination of the noise covariance matrices.

The main contributions of this paper are the following :

- Since the pioneering papers by Akaike [1], canonical correlations (which were introduced by Hotelling [5] 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 and others (e.g. [2],[7]) boils down to applying canonical correlation analysis to two matrices that are double infinite. 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.

[1]The research reported in this paper was partially supported by the Belgian Program on Interuniversity 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 and by a research project supported by Philips. The scientific responsibility is assumed by its authors.

[2]Research Assistant of the Belgian National Fund for Scientific Research

[3]Research Associate of the Belgian National Fund for Scientific Research

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

- In order to find the state space model, we derive a finite-dimensional vector sequence which, in the case of double infinite block Hankel matrices, would be a valid state sequence of the stochastic model. For the semi-infinite matrix problem, the sequence is not an exact state sequence but it corresponds to the output of j ($\rightarrow \infty$) non-steady state Kalman filters that have used a finite number i of output data as input to form an estimate of the state sequence. The state space model is then derived from the sequences by solving a least squares problem.

- It is a common belief that ARMA model identification requires non-linear estimation methods, requiring iterative solution techniques [9]. In that framework, the main problems arise from convergence difficulties and/or the existence of local minima. With the correlation approach, this ARMA modelling is 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 SVD of the triangular factors and is compeletely data driven instead of covariance driven.

This paper is organised as follows : In section 2, we summarize the main properties of linear time-invariant stochastic processes.

In section 3, we'll describe how the projection of past and future block Hankel matrices of output data, leads to the definition of the state sequences. From these sequences, it is easy to derive the stochastic state space model. This will be done in section 4. In section 5 we'll explore the relationship with the non-steady state Kalman filter. In section 6 we'll introduce our main mathematical-geometrical tool : principal angles and directions. We'll also describe a numerical robust computation scheme. In section 7, we'll describe a new, numerically robust algorithm to calculate the state space model of the stochastic system directly from output data. An illustrative examples is given in section 8.

2. Linear time - invariant stochastic processes

In this section, we'll summarize the main properties of linear time invariant stochastic processes.
It is common to assume in (1) that x_k and v_k, w_k are mutually independent and hence

$$\mathbf{E}[x_k v_k^t] = 0 \quad \text{and} \quad \mathbf{E}[x_k w_k^t] = 0$$

Then we find the following Lyapunov equation for the state covariance matrix $\Sigma \stackrel{\text{def}}{=} \mathbf{E}(x_k x_k^t)$

$$\Sigma = A\Sigma A^t + Q \tag{2}$$

Defining the output covariance matrices $\Lambda_i \stackrel{\text{def}}{=} \mathbf{E}(y_{k+i}y_k^t)$ we find for Λ_0 : $\Lambda_0 = C\Sigma C^t + R$. Defining $G \stackrel{\text{def}}{=} A\Sigma C^t + S$, we get :

$$\begin{aligned} \Lambda_i &= CA^{i-1}G & (3) \\ \Lambda_{-i} &= G^t(A^{i-1})^t C^t \qquad i = 1, 2, \ldots & (4) \end{aligned}$$

The model (1) can be converted into a so-called *forward innovation model* and a *backward innovation model* (see e.g. [12]).
The forward innovation model is obtained by applying a Kalman filter to the stochastic system (1), which results in :

$$z_{k+1} = Az_k + w_k^f \quad , \quad y_k = Cz_k + v_k^f$$

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 :

$$P = APA^t + (G - APC^t)(\Lambda_0 - CPC^t)^{-1}(G - APC^t)^t \qquad (5)$$

The backward innovations model can be described as follows :

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

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 :

$$N = A^t NA + (C^t - A^t NG)(\Lambda_0 - G^t NG)^{-1}(C^t - A^t NG)^t \qquad (6)$$

In [8] a robust algorithm to solve these Riccati equations based on the Schur decomposition is presented.

3. Orthogonal projections

In this section, we'll start to describe an identification scheme that is based on geometrical insights and that obtaines consistent estimates with semi-infinite block Hankel matrices. This last fact is in contrast with previously described similar algorithms [1],[2],[7] that needed double infinite block Hankel matrices.

In what follows, output block Hankel matrices of the form

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

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 throughout, we assume that $j \to \infty$.

Another important concept arrises with the estimation of the covariance matrices. We assume that they can be estimated as

$$\Lambda_i = lim_{j\to\infty}[\frac{1}{j}\sum_{k=0}^{j-1} y_{k+i}y_k^t] \tag{7}$$

for which we will use the notation $\mathbf{E_j}[\sum_{k=0}^{j-1} y_{k+i}y_k^t]$, with an obvious definition of $\mathbf{E_j}[.] = \lim_{j\to\infty}\frac{1}{j}[.]$. The ensamble average can be replaced by a time average due to ergodicity.

We now find easily for the correlation matrix between the future block output Hankel and the past block output Hankel matrix :

$$\mathbf{E_j}[Y_{i/2i-1}Y_{0/i-1}^t] = \begin{pmatrix} \Lambda_i & \Lambda_{i-1} & \dots & \Lambda_2 & \Lambda_1 \\ \Lambda_{i+1} & \Lambda_i & \dots & \Lambda_3 & \Lambda_2 \\ \dots & \dots & \dots & \dots & \dots \\ \Lambda_{2i-1} & \Lambda_{2i-2} & \dots & \Lambda_{i+1} & \Lambda_i \end{pmatrix} \tag{8}$$

$$
\begin{aligned}
&= \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{i-1} \end{pmatrix} \begin{pmatrix} A^{i-1}G & A^{i-2}G & \dots & G \end{pmatrix} \\
&= \mathcal{O}_i \, \mathcal{C}_i \qquad \text{(say)}
\end{aligned}
$$

If we now 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), we'll find that this projection is an n dimensional subspace of the j dimensional subspace, as can be seen from :

$$
\begin{aligned}
&\mathbf{E_j}[Y_{i/2i-1}Y_{0/i-1}^t] \, (\mathbf{E_j}[Y_{0/i-1}Y_{0/i-1}^t])^{-1} \, Y_{0/i-1} \\
&\qquad = \mathcal{O}_i \, \mathcal{C}_i \, L_i^{-1} \, Y_{0/i-1} \qquad (9)
\end{aligned}
$$

where we define

$$
\begin{aligned}
L_i \quad &\stackrel{\text{def}}{=} \quad \mathbf{E_j}[Y_{0/i-1}Y_{0/i-1}^t] = \begin{pmatrix} \Lambda_0 & \Lambda_{-1} & \Lambda_{-2} & \dots & \Lambda_{1-i} \\ \dots & \dots & \dots & \dots & \dots \\ \Lambda_{i-1} & \Lambda_{i-2} & \Lambda_{i-3} & \dots & \Lambda_0 \end{pmatrix} \\
&= \quad \mathbf{E_j}[Y_{i/2i-1}Y_{i/2i-1}^t]
\end{aligned}
$$

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 $n \times j$ matrix Z_i :

$$
Z_i = \mathcal{C}_i L_i^{-1} Y_{0/i-1} \qquad (10)
$$

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

$$
\begin{aligned}
&\mathbf{E_j}[Y_{0/i-1}Y_{i/2i-1}^t] \, (\mathbf{E_j}[Y_{i/2i-1}Y_{i/2i-1}^t])^{-1} \, Y_{i/2i-1} \\
&\qquad = \mathcal{C}_i^t \, \mathcal{O}_i^t \, L_i^{-1} \, Y_{i/2i-1}. \qquad (11)
\end{aligned}
$$

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 $n \times j$ matrix W_i :

$$W_i = \mathcal{O}_i^t L_i^{-1} Y_{i/2i-1} \tag{12}$$

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 [1]. In this framework and also the one presented in [12], those finite dimensional orthogonal projections can be considered as states of the stochastic process. As we'll see in section 5, for finite i, these sequences are not the real states, but only an optimal Kalman prediction of the states based upon the i previous outputs.

To arrive at a consitent identification scheme, we also need to define the $(i+1)^{th}$ projections as :

- The orthogonal projection of the row space of $Y_{i+1/2i-1}$ onto that of $Y_{0/i}$ which is

  $$Y_{i+1/2i-1} Y_{0/i}^t L_{i+1}^{-1} Y_{0/i} = \mathcal{O}_{i-1} \mathcal{C}_{i+1} L_{i+1}^{-1} Y_{0/i}$$

  A basis for the row space of this orthogonal projection is generated by the rows of

  $$Z_{i+1} = \mathcal{C}_{i+1} L_{i+1}^{-1} Y_{0/i} \tag{13}$$

- The orthogonal projection of the row space of $Y_{0/i-2}$ onto that of $Y_{i-1/2i-1}$, which is

  $$Y_{0/i-2} Y_{i-1/2i-1}^t L_{i+1}^{-1} Y_{i-1/2i-1} = \mathcal{C}_{i-1}^t \mathcal{O}_{i+1}^t L_{i+1}^{-1} Y_{i-1/2i-1}$$

A basis for the row space of this projection is generated by the rows of

$$W_{i+1} = \mathcal{O}_{i+1}^t L_{i+1}^{-1} Y_{i-1/2i-1} \tag{14}$$

Figure 1 illustrates these projections graphically :

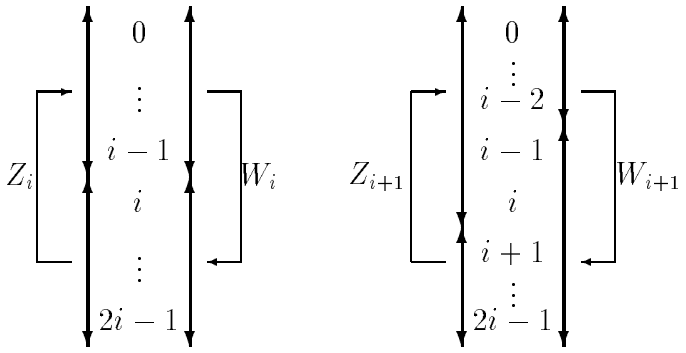


Figure 1 : Definition of the vector sequences Z_i, W_i, Z_{i+1} and W_{i+1}. The thick lines represent the span of the data, the thin lines represent the projections.

4. How to find A, C, G and Λ_0

In this section, we'll show how the system matrices A, C, G and Λ_0 can be identified consistently, by solving a set of linear equations in least squares sense.

Define the matrices P_Z and Z_r such that $Z_{i+1} = P_Z Z_i + Z_r$ where by definition we take $Z_r Z_i^t = 0$. Postmultiplying this equation with Z_i^t we find the least squares solution P_Z : $P_Z = Z_{i+1} Z_i^t (Z_i Z_i^t)^{-1}$
Let us now prove that $P_Z = A$.

$$Z_{i+1} = P_Z Z_i + Z_r$$

$\Rightarrow \quad \mathcal{C}_{i+1} L_{i+1}^{-1} \mathbf{E_j}[Y_{0/i} Y_{0/i-1}^t] L_i^{-1} \mathcal{C}_i^t = P_Z \, \mathcal{C}_i L_i^{-1} \mathcal{C}_i^t$

$$\Rightarrow \quad \mathcal{C}_{i+1} \begin{pmatrix} I & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & I \\ 0 & 0 & \dots & 0 \end{pmatrix} L_i^{-1} \mathcal{C}_i^t = P_Z \, \mathcal{C}_i L_i^{-1} \mathcal{C}_i^t$$

$\Rightarrow \quad A \, \mathcal{C}_i L_i^{-1} \mathcal{C}_i^t = P_Z \, \mathcal{C}_i L_i^{-1} \mathcal{C}_i^t$

$\Rightarrow \quad A = P_Z$

The conclusion is that A is the least squares solution of

$$\min_{P_Z} \| Z_{i+1} - P_Z Z_i \|_F^2 \quad \text{as} \;\; j \to \infty \tag{15}$$

where $\|.\|_F$ is the Frobeniusnorm. Let us now show that the matrix C can also be determined from Z_i by solving the least squares problem.

$$\min_{P_C} \| (\; y_i \;\; y_{i+1} \;\; \dots \;\; y_{i+j-1} \;) - P_C Z_i \|_F^2$$

Define P_C and Z_y by

$$y = (\; y_i \;\; y_{i+1} \;\; \dots \;\; y_{i+j-1} \;) = P_C Z_i + Z_y$$

where by definition $Z_y Z_i^t = 0$. Then from postmultiplying $y = P_C Z_i + Z_y$ with $Y_{0/i-1}^t$ and taking the expected value, we find : $C = P_C$. The preceeding derivation could also be applied to the n-dimensional projection of the past onto the future, which is given by the matrix W_i (12),which will lead to the matrices A and G. It is straightforward to derive that : A^t is the least squares solution of

$$\min_{P_W} \| W_{i+1} - P_W W_i \|_F^2 \tag{16}$$

and G^t is the least squares solution of

$$\min_{P_G} \| (\; y_{i-1} \;\; y_i \;\; \dots \;\; y_{i+j-2} \;) - P_G W_i \|_F^2 \tag{17}$$

Λ_0 can then be found as : $\Lambda_0 = \mathbf{E_j}(\sum_{k=0}^{j-1} y_k y_k^t)$

5. Relation with the Kalman filter

In this section, we'll 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 respectively $i+1$ time steps.

The proof consists of two parts : 1/ The state covariance matrices $\Sigma_i = \mathbf{E_j}(Z_i Z_i^t)$ and $\Sigma_{i+1} = \mathbf{E_j}(Z_{i+1} Z_{i+1}^t)$ are two iterates of the difference Riccati equation associated with the forward model and 2/ Z_{i+1} can be formed from Z_i as : $Z_{i+1} = AZ_i + K_i(y_i - CZ_i)$ with K_i equal to the Kalman gain.

1/ From the definition of the sequences Z_i and Z_{i+1} we find easily :

$$\begin{aligned} \Sigma_i = \mathbf{E_j}(Z_i Z_i^t) &= \mathcal{C}_i L_i^{-1} \mathcal{C}_i^t \\ \Sigma_{i+1} = \mathbf{E_j}(Z_{i+1} Z_{i+1}^t) &= \mathcal{C}_{i+1} L_{i+1}^{-1} \mathcal{C}_{i+1}^t \end{aligned}$$

If we rewrite Σ_{i+1} in function of Σ_i with the help of the matrix inversion lemma, we get (see [4]) :

$$\Sigma_{i+1} = A\Sigma_i A^t + (A\Sigma_i C^t - G)(\Lambda_0 - C\Sigma_i C^t)^{-1}(C\Sigma_i A^t - G^t) \quad (18)$$

which is one step of the forward difference Riccati equation. The Kalman gain associated with this equation is at time step i :

$$K_i = (G - A\Sigma_i C^t)(\Lambda_0 - C\Sigma_i C^t)^{-1}$$

2/ We'll 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) : $Z_{i+1} = N_i Z_i + M_i y_i$, and by postmultiplying this equation with $Y_{0/i-1}^t$ and taking the expected value, we find : $A = N_i + M_i C$, which leads to : $Z_{i+1} = AZ_i + M_i(y_i - CZ_i)$. If we postmultiply this equation with y_i^t and take the expected value we get : $G = A\Sigma_i C^t + M_i(\Lambda_0 - C\Sigma_i C^t)$ 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 :

$$Z_1 = K_1 y_0 = G\Lambda_0^{-1} y_0 = \mathcal{C}_1 L_1^{-1} y_0$$

Through induction, this clearly proves that the sequence Z_0, Z_1, ..., Z_i, Z_{i+1} is a sequence of subsequent optimal Kalman state estimates. This has to be interpreted as follows : the l^{th} 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 : $x_{l-1} = 0$, $\mathbf{E}(x_{l-1} x_{l-1}^t) = 0$. Figure 2 illustrates this graphically. So, in this way every column of Z_i represents an optimal state estimate. And we can interprete the columns of Z_i as the outputs of a bank of j Kalman filters that estimate j optimal states in parallel.

$$\begin{array}{cc} Z_0 & \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \\ Y_{0/i-1} & \begin{bmatrix} y_0 & & y_{l-1} & & y_{j-1} \\ \vdots & & \vdots & & \vdots \\ y_{i-1} & & y_{l+i-2} & & y_{i+j-2} \end{bmatrix} \\ Z_i & \begin{bmatrix} z_i & \cdots & z_{l+i-1} & \cdots & z_j \end{bmatrix} \end{array}$$

Figure 2 : Interpretation of the sequence Z_i as Kalman filter state estimates based upon i measurements.

6. Orthogonal projections and principal angles and directions.

In this section, we briefly review orthogonal projections and the corresponding matrix algebra. We explain the notion of principal angles and directions between subspaces.

6.1 Orthogonal projections

Given two matrices $U \in \Re^{p\times j}$ and $V \in \Re^{q\times j}$, where $j \geq max(p,q)$ and $rank(U) = p$ and $rank(V) = q$. Then, the row space of $UV^t(VV^t)^{-1}V$ is the orthogonal projection of the row space of U onto that of V while the rowspace of $VU^t(UU^t)^{-1}U$ is the orthogonal projection of the row space of V onto that of U. We can combine these two projections in the matrix

$$U^t(UU^t)^{-1}UV^t(VV^t)^{-1}V \tag{19}$$

the row space of which is the orthogonal projection of the row space of U onto that of V. The column space of (19) is the orthogonal projection of the row space of V onto that of U.

6.2 Geometrical motivation

The notion of an angle between two vectors can be generalized to angles between subspaces by using a recursive procedure. This concept goes back to Jordan [6]. Before giving a formal definition, let us first present a verbal recipe. Suppose we are given two matrices $U \in \Re^{p\times j}$ and $V \in \Re^{q\times j}$. The first principal angle θ_1 (the smallest one) is obtained as follows : Choose unit vectors $u_1 \in \mathbf{R}(U^t)$ and $v_1 \in \mathbf{R}(V^t)$ and minimize the angle between them (The notation $\mathbf{R}(.)$ refers to the column space (range) of the matrix between brackets.) Next choose a unit vector $u_2 \in \mathbf{R}(U^t)$ orthogonal to u_1 and $v_2 \in \mathbf{R}(V^t)$ 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 is now formalized :

Definition 1 Principal angles and directions
The principal angles $\theta_1 \leq \theta_2 \leq \ldots \leq \pi/2$ between the row spaces

$\mathbf{R}(U^t)$ *and* $\mathbf{R}(V^t)$ *of two matrices* $U \in \Re^{p \times j}$ *and* $V \in \Re^{q \times j}$ *and the corresponding principal directions* $u_i \in \mathbf{R}(U^t)$ *and* $v_i \in \mathbf{R}(V^t)$ *are defined recursively as*

$$cos\theta_k = max_{\ u \in \mathbf{R}(U^t)\ ,\ v \in \mathbf{R}(V^t)\ } \ u^t v = u_k^t v_k$$

subject to $\|u\| = \|v\| = 1$ *and for* $k > 1$ *:* $u^t u_i = 0$ *and* $v^t v_i = 0$ *for* $i = 1, \ldots, k-1$

From this definition, it is not too difficult to show that

Lemma 1
The principal directions and cosines of the angles between the row spaces of two matrices U *and* V *are the singular vectors and values of the matrix* $U^\dagger U V^\dagger V$*.*

If both U and V are of full row rank then of course $U^\dagger U V^\dagger V = U^t (U U^t)^{-1} U V^t (V V^t)^{-1} V$.

6.3 The QR-decomposition and the **QSVD**

Lemma 2
Any pair of matrices $A \in \Re^{m \times n}$ *and* $B \in \Re^{p \times n}$ *can be decomposed as*

$$A = U_a S_a X^t \quad and \quad B = U_b S_b X^t$$

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

$$S_a = \begin{array}{c} \\ r_{ab} - r_b \\ r_a + r_b - r_{ab} \\ m - r_a \end{array} \begin{array}{c} \begin{array}{cccc} r_{ab} - r_b & r_a + r_b - r_{ab} & r_{ab} - r_a & n - r_{ab} \end{array} \\ \begin{pmatrix} I & 0 & 0 & 0 \\ 0 & D_a & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

$$S_b = \begin{array}{c} \\ p - r_b \\ r_a + r_b - r_{ab} \\ r_{ab} - r_a \end{array} \begin{array}{c} \begin{array}{cccc} r_{ab} - r_b & r_a + r_b - r_{ab} & r_{ab} - r_a & n - r_{ab} \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & D_b & 0 & 0 \\ 0 & 0 & I & 0 \end{pmatrix} \end{array}$$

where

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

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

The original formulation of the **QSVD** (which used to be called *the generalized SVD*, but see also [3]) is due to Paige [11] and Van Loan [13].

We are now ready to derive the following Theorem (which to our knowledge is new) :

Theorem 1
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 $R_{21}^t = USX^t$ and $R_{22}^t = VTX^t$, 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*

$$(\; Q_1 U \;) \; S \; (\; Q_1 US + Q_2 VT \;)^t$$

where the nonzero elements of S are the cosines of the principal angles, the left factor contains the principal directions in the row space of A and the right factor contains the principal directions in the row space of B.

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

$$\begin{aligned} A^\dagger A &= Q_1 Q_1^t \\ B^\dagger B &= (Q_1 US + Q_2 VT)(S^t U^t Q_1^t + T^t V^t Q_2^t) \end{aligned}$$

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

$$A^\dagger AB^\dagger B = (Q_1U)\ S\ (Q_1US + Q_2VT)^t$$

which is clearly an **SVD**. It follows from Lemma 1 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.

7. Principal angles and directions and output block Hankel matrices.

We have derived in section 4 that the matrices A and C could be estimated from the matrices Z_i and Z_{i+1} while A and G 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.
First, consider

$$Y^t_{i/2i-1}L_i^{-1}\mathcal{O}_i\mathcal{C}_iL_i^{-1}Y_{0/i-1} = W_i^tZ_i \tag{20}$$

We know from Lemma 1 that the **SVD** of (20) delivers the principal angles and directions between past and future.
Recall that the rows of Z_i form a basis for the projection of the row space of $Y_{i/2i-1}$ onto the row space of $Y_{0/i-1}$. Consider the RQ-decomposition of the output block Hankel matrix :

$$\frac{Y_{0/2i-1}}{\sqrt{j}} = \begin{pmatrix} Y_{0/i-1} \\ Y_{i/2i-1} \end{pmatrix} = \begin{pmatrix} R_{11} & 0 \\ R_{21} & R_{22} \end{pmatrix} \begin{pmatrix} Q_1^t \\ Q_2^t \end{pmatrix} \tag{21}$$

where all blocks in R are square : $li \times li$. The scalar $\sqrt{j}$ is used to be conform with the definition of the operator $\mathbf{E_j}$.

Lemma 3
Let $j \to \infty$ and consider the QR-decomposition as in (21). Then the following facts hold true :
- The column space of R_{21} in the QR-decomposition (21) coincides with the column space of $\mathcal{O}_i$.
- Only n principal angles between the row spaces of $Y_{0/i-1}$ and $Y_{i/2i-1}$ are different from $\pi/2$.

Proof : It is straightforward to derive that

$$\mathbf{E_j}[Y_{i/2i-1}Y_{0/i-1}^t](\mathbf{E_j}[Y_{0/i-1}Y_{0/i-1}^t])^{-1}Y_{0/i-1} = R_{21}Q_1^t$$

It then follows from (9) that the column space of R_{21} coincides with the column space of the extended observability matrix $\mathcal{O}_i$ as $j \to \infty$. The fact that only n principal angles are different from 90^o follows from the rank deficiency in (20). This enables us to estimate the order n of the stochastic system : when $j \to \infty$, n is equal to the number of principal angles different from $\pi/2$. □

Next consider the QR-decomposition of the output block Hankel matrix $Y_{0/2i-1}$ with the following appropriate partitioning (which is also visualized in Figure 1):

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

Then, it follows from Lemma 3 that the column spaces of $(\; R_{41} \;\; .. \;\; R_{43} \;)$, $\begin{pmatrix} R_{31} & R_{32} \\ R_{41} & R_{42} \end{pmatrix}$ and $\begin{pmatrix} R_{21} \\ R_{31} \\ R_{41} \end{pmatrix}$ coincide with the

respective column spaces of $\mathcal{O}_{i-1}$, $\mathcal{O}_i$ and $\mathcal{O}_{i+1}$. Next, we will apply Theorem 1 to these three blocks. Hereto we firsts compute three **QSVD**s :

$$\begin{aligned}(R_{41} \;\; R_{42} \;\; R_{43})^t &= U_{i-1} S_{i-1} X_{i-1}^t \\ R_{44}^t &= V_{i-1} T_{i-1} X_{i-1}^t\end{aligned}$$

$$\begin{aligned}\begin{pmatrix} R_{31} & R_{32} \\ R_{41} & R_{42} \end{pmatrix}^t &= U_i S_i X_i^t \\ \begin{pmatrix} R_{33} & 0 \\ R_{43} & R_{44} \end{pmatrix}^t &= V_i T_i X_i^t\end{aligned}$$

$$\begin{aligned}\begin{pmatrix} R_{21} \\ R_{31} \\ R_{41} \end{pmatrix}^t &= U_{i+1} S_{i+1} X_{i+1}^t \\ \begin{pmatrix} R_{22} & 0 & 0 \\ R_{32} & R_{33} & 0 \\ R_{42} & R_{43} & R_{44} \end{pmatrix}^t &= V_{i+1} T_{i+1} X_{i+1}^t\end{aligned}$$

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, V_{i+1}, S_{i-1}, S_i, S_{i+1}, T_{i-1}, T_i and T_{i+1} with a supperscript 1 denote the first n directions and first n diagonal elements of these matrices. It then follows from Theorem 1 that the column spaces of $\mathcal{O}_{i-1}, \mathcal{O}_i$ and $\mathcal{O}_{i+1}$ are generated by the columns of $X_{i-1}^1 (S_{i-1}^1)^{1/2}$, $X_i^1 (S_i^1)^{1/2}$ and $X_{i+1}^1 (S_{i+1}^1)^{1/2}$. This means that there exist three similarity transformations, M_{i-1}, M_i and M_{i+1} so that

$$\begin{aligned}X_{i-1}^1 (S_{i-1}^1)^{1/2} M_{i-1} &= \mathcal{O}_{i-1} \\ X_i^1 (S_i^1)^{1/2} M_i &= \mathcal{O}_i \\ X_{i+1}^1 (S_{i+1}^1)^{1/2} M_{i+1} &= \mathcal{O}_{i+1}\end{aligned}$$

Without loss of generality, we can set one of these three equal to the identity matrix (which is equivalent with fixing a certain basis in the state space), say $M_i = I_n$. Then we get :

$$M_{i-1} = (S^1_{i-1})^{-1/2}(X^1_{i-1})^\dagger \underline{X^1_i}(S^1_i)^{1/2} \quad (23)$$

$$M_{i+1} = (S^1_{i+1})^{-1/2}(\underline{X^1_{i+1}})^\dagger X^1_i(S^1_i)^{1/2} \quad (24)$$

The notation $\underline{X^1_i}$ refers to the $l(i-1) \times n$ matrix obtained from X^1_i by omission of its last l rows.

Let us now clarify how the **QSVD** allows us to calculate the matrices Z_i, W_i, Z_{i+1} and W_{i+1} and at the same time provide the required similarity transformations.
Recall that $Y^t_{i/2i-1}L_i^{-1}[\mathbf{E_j}[Y_{i/2i-1}Y^t_{0/i}]]^{-1}L_i^{-1}Y_{0/i-1} = W_i^t Z_i$ Using the QR-decomposition and the **QSVD** we take

$$Z_i = (S^1_i)^{1/2}(U^1_i)^t \begin{pmatrix} Q^t_1 \\ Q^t_2 \end{pmatrix} \quad (25)$$

$$W_i = (S^1_i)^{1/2}\begin{pmatrix} (S^1_i)^t U^t_i & (T^1_i)^t V^t_i \end{pmatrix} \begin{pmatrix} Q^t_1 \\ Q^t_2 \\ Q^t_3 \\ Q^t_4 \end{pmatrix} \quad (26)$$

At the same time we find

$$Y_{i/2i-1}Y^t_{0/i-1}L_i^{-1}Y_{0/i-1} = \mathcal{O}_i\mathcal{C}_i L_i^{-1}Y_{0/i-1} \quad (27)$$

$$= (X^1_i)(S^1_i)^t(U^1_i)^t \begin{pmatrix} Q^t_1 \\ Q^t_2 \end{pmatrix} \quad (28)$$

The conclusion is that the combined QR-QSVD approach delivers at once the sequences Z_i (25) and W_i (26) and the matrix $\mathcal{O}_i$ (from (27)) as $X^1_i(S^1_i)^{1/2}$ (28).

We can obtain the matrices Z_{i+1} and W_{i+1} by using formulae (13)-(14), but we have also to take into account the similarity transformations M_{i-1} and M_{i+1} as in (23)-(24) :

$$Z_{i+1} = M_{i-1}^{-1}(S_{i-1}^1)^{1/2}(U_{i-1}^1)^t \begin{pmatrix} Q_1^t \\ Q_2^t \\ Q_3^t \end{pmatrix} \tag{29}$$

$$W_{i+1} = M_{i+1}^t(S_{i+1}^1)^{1/2}[\ (S_{i+1}^1)(U_{i+1}^1)^t \ \ (T_{i+1}^1)(V_{i+1}^1)^t\ ] \begin{pmatrix} Q_1^t \\ Q_2^t \\ Q_3^t \\ Q_4^t \end{pmatrix} \tag{30}$$

If we now apply equation (15) to these matrices, we find that A is given by :

$$A_f = (S_i^1)^{-1/2}(\underline{X_i^1})^\dagger X_{i-1}^1 S_{i-1}^1 (\underline{U_{i-1}^1})^t U_i^1 (S_i^1)^{-1/2} \tag{31}$$

Or with equation (16) we find for A_b^t :

$$(S_i^1)^{1/2} X_i^1 (\underline{X_{i+1}^1}^{\ t})^\dagger (\ S_{i+1}^1(U_{i+1}^1)^t \ \ T_{i+1}^1(V_{i+1}^1)^t\ ) \begin{pmatrix} U_i^1 S_i^1 \\ V_i^1 T_i^1 \end{pmatrix} (S_i^1)^{-1/2} \tag{32}$$

The matrix C follows from the least squares solution of () and is given by

$$C = (\ R_{31} \ \ R_{32}\ ) U_i^1 (S_i^1)^{-1/2} = (\ I_l \ \ 0\ ) X_i (S_i^1)^{1/2} \tag{33}$$

The second equality shows that C is equal to the first l rows of the estimated $\mathcal{O}_i$. Finally, the matrix G follows from the least squares solution (17)

$$G = (\ R_{21} \ \ R_{22}\ ) U_i^1 S_i^1 \tag{34}$$

This last equation could also be written in function of the **QSVD**'s, but that would make things unnecessary complicated. Observe

that in these final expressions for A, C and G the orthogonal matrices Q_1, Q_2, Q_3 and Q_4 have cancelled out, so we don't need to calculate them. This observation represents a major reduction in the computational complexity.

8. An example

Estimates of the state sequence Z_i are obtained by taking linear combinations of the past output matrix $Y_{0/i-1}$ as $Z_i = Q_i Y_{0/i-1}$. Now, in earlier work [2],[7],[12], the shifted state sequence Z_{i+1} is computed by applying the same linear combinations Q_i to the shifted Hankel matrix $Y_{1/i}$ as $Z_{i+1} = Q_i Y_{1/i}$. While this trick works for purely deterministic systems (see [10]), it doesn't 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 wurse as some of the poles of A move towards the unit circle. Only for $i \rightarrow \infty$, this classical approach will be exact (as e.g. in [12]). The following easy example illustrates this clearly. Applying formulae (10) and (13) to a block Hankel matrix with respectively one and two block rows (SISO system), gives :

$$\begin{aligned} Z_1 &= \mathcal{C}_1 L_1^{-1} Y_{0/0} = G\Lambda_0^{-1} Y_{0/0} \\ Z_2 &= \begin{pmatrix} AG & G \end{pmatrix} \begin{pmatrix} \Lambda_0 & \Lambda_1 \\ \Lambda_1 & \Lambda_2 \end{pmatrix}^{-1} Y_{0/1} \end{aligned}$$

We then find that

$$P_Z = Z_2 Z_1^t (Z_1 Z_1^t)^{-1} = A$$

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

$$P_Z = Z_2 Z_1^t (Z_1 Z_1^t)^{-1} = \Lambda_1/\Lambda_0 = CG/(C\Sigma C^t + R)$$

which has nothing to do with A !
Consider the very simple SISO system with only one state : $A = 0.9490, B = -0.4424, C = 0.8725, D = -0.9666$. Figure 3 shows the eigenvalues of the identified A matrix (a scalar), in function of the parameter i ($j = 1000$) for the 2 algorithms : Stars (*) : taking the same linear combinations as described in this section, O's : (o) the algorithm as described in section 7. For every i, 100 Monte Carlo experiments were performed. The mean value of these 100 experiments is plotted for every i on Figure 3 (the exact value is ofcourse 0.949). Clearly the algorithm described in this section (which is the one that is normally used in the liturature), calculates heavily biased solutions for small i. On the other hand, the new algorithm (of section 7) doesn't suffer at all from this problem.

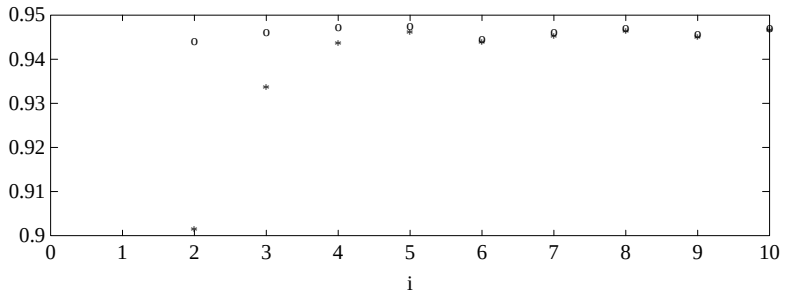


Figure 3 : Stars (*) : Mean of the estimated eigenvalues in function of i for the algorithm of section 8. O's (o) : The same for the algorithm of section 7.

9. Conclusions

In this paper, we have derived a new algoritm 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 the 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** decompositions. We have also interpreted the principal directions as non-steady state Kalman filter states. With a couple of examples, we have showed that the new algorithm is superior to the classical canonical correlation algorithms, especially for small i.

References

[1] Akaike H. *Markovian Representation of Stochastic processes by Canonical Variables.* SIAM Journal on Control, 13, 1, pp. 162-173, January 1975.

[2] Arun K.S., Kung S.Y. *Balanced Aproximation of Stochastic Systems.* SIAM Matrix Analysis and Applications, Vol. 11, No 1., pp. 42-68, January 1990.

[3] De Moor B., Golub G.H. *Generalized singular value decompositions : A proposal for a standardized nomenclature.* Numerical Analysis Project Report 89-04, Department of Computer Science, Stanford University, April 1989.

[4] Faure P. *Stochastic realization algorithms.* System identification : Advances and case studies. Eds. Mehra R. and Lainiotis D., Academic Press, 1976.

[5] Hotelling H. *Relations between two sets of variates.* Biometrika, Vol.28, pp. 321-377, 1936.

[6] Jordan C. *Essai sur la géometrie à n dimensions.* Bull. Soc. Math. France, 3, 1875, pp. 103-174.

[7] Larimore W.E. *Canonical Variate Analysis in Identification, Filtering, and Adaptive Control.* Proceedings of the 29th Conference on Decision and Control, Hawaii, December 1990.

[8] Laub A. *A Schur method for solving algebraic Riccati equations.* IEEE Transactions on Automatic Control, AC-24, pp. 913-921, 1979.

[9] Ljung L. *System Identification : theory for the user.* Prentice-Hall Information and System Sciences Series, 1987.

[10] Moonen M., De Moor B., Vandenberghe L., Vandewalle J. *On- and off-line identification of linear state space models.* International Journal of Control, Vol. 49, No. 1, pp. 219-232, January 1989.

[11] Paige C.C., Saunders M.A. *Towards a generalized singular value decomposition.* SIAM J. Numer. Anal., 18, pp. 398-405, (1981).

[12] Pal D. *Balanced Stochastic Realization and Model Reduction.* Master's thesis, Whasington State University, Electrical Engineering, 1982.

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