P. Valcker

GSVD-based stochastic realization*

Johan Suykens, Bart De Moor, Joos Vandewalle

ESAT
(Electronics, Systems, Automation and Technology)
Katholieke Universiteit Leuven
Kardinaal Mercierlaan 94
B-3030 Leuven (Heverlee)
Belgium
tel: 32/16/22 09 31
fax: 32/16/22 18 55
email: demoor@esat.kuleuven.ac.be
or na.demoor@na-net.stanford.edu

ESAT-SISTA Report 1990-~~03~~ 59

July 9, 1990

*This research work was carried out at the ESAT Laboratorium of the Katholieke Universiteit Leuven, in the framework of the Community Research Programma ESPRIT Basic Research Action nr.3280, and was partially sponsored by the European Commission.

Abstract

A solution to the stochastic realization problem is given here in terms of a generalized singular value decomposition (GSVD) of block Hankel and block Toeplitz matrices, related to the future and the past outputs of a MIMO time-discrete, time-invariant, linear, finite-dimensional, second-order stationary stochastic process. The process-order estimation is based on the canonical angles between transformed spaces of the future and the past outputs, which can be expressed directly by means of the GSVD. This estimation is obtained after applying a so-called Orthogonality Theorem. The GSVD also proposes a framework to classify different realization schemes and is interesting from a numerical viewpoint.

Keywords: approximate stochastic realization, canonical angles, RV-coefficient, GSVD (QPQ-SVD, PQ-SVD, PSVD)

Contents:

1 Introduction

In the recent past the problem of finding a finite dimensional Markovian representation (state-space model) from the knowledge of the autocovariance sequence of a second-order stationary stochastic process, received much attention because of applications in system identification, digital filtering, signal processing, time series modelling, aerospace engineering, metrology, geophysical data processing etc. Several realization schemes were introduced [7][5], based on the maximization of performance indices, subject to certain constraints. This paper aims to unify the different approaches within a generalized singular value decomposition (GSVD) framework: for each optimization criterion with its constraints, we have to apply a certain GSVD-configuration in order to get a state vector sequence of the stochastic process and to estimate the process order.

The process order will then be determined as an amount of interaction between the past and the future of the stochastic process, based on that GSVD and a so-called Orthogonality Theorem. The principle is not based on finding an intersection between transformed past and future output spaces, but on linear *dependencies* in terms of that Orthogonality Theorem. A state vector sequence is then obtained as a transformation of the past or future outputs of the process. Another difference with previous methods is that no use is made of a backward state-space representation: all reasoning is based on a *forward* state-space model, which has its future and past input-output matrix equations. The GSVD offers then a unifying framework for different realization schemes and allows to compute them in a numerical reliable way.

Notations:

A^t: *transpose of matrix A*
A^+: *matrix related to the future*
A^-: *matrix related to the past*
A^{-t}: *inverse of* A^t
$(A)^+$: *pseudo-inverse of A*
$A(i:j, k:l)$: *submatrix of A consisting of rows i to j and columns k to l of A*
$span_{col(row)}(A)$: *space generated by the columns (rows) of A*
$A|B$: *projection of* $span_{col}(A)$ *on* $span_{col}(B)$
I_i: *identity matrix* $(i \times i)$
$\underline{x}$: *columnvector*
$\underline{x}_k$: *columnvector at time k*
$x_k(i)$: *i-th component of* $\underline{x}_k$
$\{.\}$: *set of .*
$\hat{.}$: *estimation of .*
$E\{.\}$: *expectation operator of .*
GSVD: generalized singular value decomposition
OSVD: ordinary singular value decomposition
PSVD: product SVD
QSVD: quotient SVD
PQ-SVD: product-quotient SVD
QPQ-SVD: quotient-product-quotient SVD

2 Markovian representation of a stochastic process

Let us suppose that a state space representation (called Markovian representation) of the form:

$$\begin{aligned} \underline{x}_{k+1} &= A.\underline{x}_k + \underline{v}_k \\ \underline{y}_k &= C.\underline{x}_k + \underline{w}_k \end{aligned} \quad (1)$$

can serve as a model for a MIMO stochastic process, which is time-discrete, time-invariant, linear, finite-dimensional and second order stationary. Here $\{\underline{x}_k\}$ is the $(n \times 1)$ state vector process, $\{\underline{v}_k\}$ $(n \times 1)$ and $\{\underline{w}_k\}$ $(l \times 1)$ are zero mean white Gaussian noise processes with variance σ^2, $\{\underline{y}_k\}$ is the $(l \times 1)$ output vector process.

Theorem 1:

The state space form (1) can be formulated more *algebraically* by means of so-called input-output (I/O) equations:

$$Y^+ = \Gamma^+ X^+ + T^+ V^+ + W^+ \quad (2)$$

$$Y^- = \Gamma^- X^- + T^- V^- + W^- \quad (3)$$

Here the I/O-equation (2) is related to the future of the process and (3) to the past of the process. The matrices Y^+ and W^+ are block Hankel, Y^- and W^- are block Toeplitz matrices, Γ^+ and Γ^- are extended observability matrices. The matrices of the I/O-equation are written out here:

$$Y^+(li \times j) = \begin{bmatrix} \underline{y}_k & \underline{y}_{k+1} & \cdots & \underline{y}_{k+j-1} \\ \underline{y}_{k+1} & \underline{y}_{k+2} & \cdots & \underline{y}_{k+j} \\ \vdots & \vdots & \cdots & \vdots \\ \underline{y}_{k+i-1} & \underline{y}_{k+i} & \cdots & \underline{y}_{k+i+j-2} \end{bmatrix}$$

$$Y^-(li \times j) = \begin{bmatrix} \underline{y}_{k-1} & \underline{y}_k & \cdots & \underline{y}_{k+j-2} \\ \underline{y}_{k-2} & \underline{y}_{k-1} & \cdots & \underline{y}_{k+j-3} \\ \vdots & \vdots & \cdots & \vdots \\ \underline{y}_{k-i} & \underline{y}_{k-i+1} & \cdots & \underline{y}_{k-i+j-1} \end{bmatrix}$$

For the definition of $W^+(li \times j)$ and $W^-(li \times j)$ just replace the symbol y by w.

$$X^+(n \times j) = [\underline{x}_k \; \underline{x}_{k+1} \cdots \underline{x}_{k+j-1}]$$
$$X^-(n \times j) = [\underline{x}_{k-i} \; \underline{x}_{k-i+1} \cdots \underline{x}_{k-i+j-1}]$$

For the definition of $V^+(n \times j)$ and $V^-(n \times j)$ just replace the symbol x by v.

$$\Gamma^+(li \times n) = \begin{bmatrix} C \\ C.A \\ C.A^2 \\ \vdots \\ C.A^{i-1} \end{bmatrix}, \quad \Gamma^-(li \times n) = \begin{bmatrix} C.A^{i-1} \\ C.A^{i-2} \\ \vdots \\ C.A \\ C \end{bmatrix}$$

$$T^+(li \times n) = \begin{bmatrix} O \\ C \\ C.A \\ \vdots \\ C.A^{i-2} \end{bmatrix}, \quad T^-(li \times n) = \begin{bmatrix} C.A^{i-2} \\ \vdots \\ C.A \\ C \\ O \end{bmatrix}$$

proof:
The proof is a straightforward repeated substitution of (1).

Stochastic realization aims then to find the matrices A and C of the state space description (1) from the knowledge of the output vector process and to obtain a state vector sequence. Three aspects are involved with this problem of stochastic realization:

1. estimation of the process order n
2. obtaining a state vector sequence X^+ or X^-
3. computation of A and C.

Let us introduce now some assumptions.

Assumption 1:
There exist matrices $F(li \times n)$ and $G(li \times n)$ such that

$$X^+ = F^t Y^+ \quad (4)$$
$$X^- = G^t Y^- \quad (5)$$

A method for obtaining F and G will be described in chapter 6.

Assumption 2:
Let us assume that the property of *ergodicity* holds for the stochastic process which means that an expected value over several measurements may be replaced by an expected value over time. Suppose then that for $j \gg li$:

$$E\{\begin{bmatrix} \underline{v}_k \\ \underline{w}_k \end{bmatrix}.[\underline{v}_s^t \ \underline{w}_s^t]\} = \frac{1}{j}.\begin{bmatrix} V^+(V^+)^t & V^+(W^+(1:l,:))^t \\ W^+(1:l,:)(V^+)^t & W^+(1:l,:)(W^+(1:l,:))^t \end{bmatrix}.\delta_{ks} \quad (6)$$

($\delta_{ks} = 1$ if $k = s$, otherwise $\delta_{ks} = 0$)
with

$$\frac{1}{j}.V^+(V^+)^t = \sigma^2 I_n \quad (7)$$
$$\frac{1}{j}.W^+(1:l,:)(W^+(1:l,:))^t = \sigma^2 I_l$$

Assumption 3:
Assume for large j:

$$\frac{1}{j}.X^+(V^-)^t = O \quad (8)$$
$$\frac{1}{j}.X^+(W^-)^t = O$$
$$\frac{1}{j}.V^+(X^-)^t = O$$
$$\frac{1}{j}.W^+(X^-)^t = O$$

Theorem 2:
The extended observability matrix Γ^+ can be obtained by a QR-factorization of $[(X^+)^t \ (Y^+)^t]$:

$$[(X^+)^t \ (Y^+)^t] = [Q_1 \ Q_2].\begin{bmatrix} R_{11} & R_{12} \\ O & R_{22} \end{bmatrix} \quad (9)$$

where

$$\Gamma^+ = (R_{11}^{-1} R_{12})^t$$

under the conditions that:

$$\begin{aligned} X^+(V^+)^t &= O \\ X^+(W^+)^t &= O \end{aligned} \tag{10}$$

Proof:
The factorization (9) delivers us 2 equations:

$$\begin{aligned} (X^+)^t &= Q_1 R_{11} \\ (Y^+)^t &= Q_1 R_{12} + Q_2 R_{22} \end{aligned}$$

Eliminate Q_1 in these equations:

$$(Y^+)^t = (X^+)^t R_{11}^{-1} R_{12} + Q_2 R_{22} \tag{11}$$

Equation (11) may be identified with (2):

$$(Y^+)^t = (X^+)^t (\Gamma^+)^t + (T^+ V^+ + W^+)^t$$

because assumption (10) obeys the QR-factorization property that $Q_1^t Q_2 = O$.
Thus Γ^+ may be identified as:

$$\Gamma^+ = (R_{11}^{-1} R_{12})^t$$

□

Thanks to the shiftstructure of Γ^+ we can compute A and C as:

$$\begin{aligned} A &= (\underline{\Gamma^+})^+ . \overline{\Gamma^+} \\ C &= \Gamma^+(1:l,:) \end{aligned}$$

with

$$\underline{\Gamma^+} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{i-2} \end{bmatrix}, \quad \overline{\Gamma^+} = \begin{bmatrix} CA \\ CA^2 \\ CA^3 \\ \vdots \\ CA^{i-1} \end{bmatrix}$$

Remark:
Because conditions (10) differ from (8) by the factor j, the conditions (7) must be adapted into:

$$\begin{aligned} V^+(V^+)^t &= \sigma^2 I_n \\ W^+(1:l,:)(W^+(1:l,:))^t &= \sigma^2 I_l \end{aligned}$$

when the matrices A and C of Γ^+ are used for simulation of the stochastic process.

In the next 2 sections we briefly repeat the concepts of the RV-coefficient [4][5] and the generalized singular value decomposition (GSVD) [1][2].

3 The RV-coefficient as a measure of closeness between spaces

The RV-coefficient, introduced by Robert and Escoufier [4] as a tool for solving a large class of problems arising in multivariate statistical analysis, is defined as follows:

Definition 1:
Suppose that 2 data matrices X and Y are given, then the RV-coefficient between those matrices is defined as:

$$\boxed{RV(X,Y) = \frac{trace(XY^t.YX^t)}{[trace(XX^t)^2.trace(YY^t)^2]^{\frac{1}{2}}}} \tag{12}$$

This RV-coefficient serves as a measure of closeness between the spaces $span_{row}(X)$ and $span_{row}(Y)$ and it has the following properties:

1. $RV(X,Y) \in [0,1]$
2. the larger $RV(X,Y)$, the closer $span_{row}(X)$ and $span_{row}(Y)$ ly to each other.
   Special cases are:
   - $RV(X,Y) = 1$ if and only if $X = kY$ (k is a nonzero scalar)
   - $RV(X,Y) = 0$ if and only if $XY^t = O$
3. $RV(X,Y) = RV(Y,X)$
4. $RV(A^tX,Y) = RV(X,Y)$ if A orthogonal
5. $RV(kX,Y) = RV(X,Y)$ for some nonzero scalar k

This concept will be developed later for measuring the closeness of the spaces $span_{row}(X^+)$ and $span_{row}(X^-)$.

4 The generalized singular value decomposition (GSVD)

As a generalization of the ordinary singular value decomposition (OSVD), a factorization of a set of matrices can be obtained as a so-called GSVD. This is recently introduced by De Moor and Golub [1][2].
The theorem is repeated here for real matrices only:

Theorem 3:
Given a set of k real matrices with compatible dimensions: $\{A_1(n_0 \times n_1), A_2(n_1 \times n_2), \ldots, A_{k-1}(n_{k-2} \times n_{k-1}), A_k(n_{k-1} \times n_k)\}$. Then there exist matrices U_1, V_k, $D_j(j = 1,k)$, $X_j(j = 1, k-1)$, $Z_j(j = 1, k-1)$ such that the next factorization can be obtained:

$$\begin{aligned} A_1 &= U_1 D_1 X_1^{-1} \\ A_2 &= Z_1 D_2 X_2^{-1} \\ \vdots &= \vdots \\ A_i &= Z_{i-1} D_i X_i^{-1} \\ \vdots &= \vdots \\ A_k &= Z_{k-1} D_k V_k^t \end{aligned} \tag{13}$$

with:

- $U_1(n_0 \times n_0)$ and $V_k(n_k \times n_k)$ are orthonormal matrices
- the matrices $D_j (j = 1, k-1)$ have the form:

$$D_j(n_{j-1} \times n_j) = \begin{bmatrix} I(r_{1_j}) & O & \dots & O & O \\ O(r_{1_{j-1}} - r_{1_j}) & O & \dots & O & O \\ O & I(r_{2_j}) & \dots & O & O \\ O & O(r_{2_{j-1}} - r_{2_j}) & \dots & O & O \\ \vdots & \vdots & \dots & \vdots & \vdots \\ O & O & \dots & I(r_{j_j}) & O \\ O & O & \dots & O(n_{j-1} - r_{j-1} - r_{j_j}) & O \end{bmatrix}$$

The dimension of the submatrices of D_j are given between brackets and $r_0 = 0$, $r_j = \sum_{i=1}^{j} = rank(A_j)$. The integers r_{i_j} are ranks of certain matrices, given in a constructive proof of this theorem.

- the matrix D_k has the form:

$$D_k(n_{k-1} \times n_k) = \begin{bmatrix} D_{1_k}(r_{1_k}) & O & \dots & O & O \\ O(r_{1_{k-1}} - r_{1_k}) & O & \dots & O & O \\ O & D_{2_k}(r_{2_k}) & \dots & O & O \\ O & O(r_{2_{k-1}} - r_{2_k}) & \dots & O & O \\ \vdots & \vdots & \dots & \vdots & \vdots \\ O & O & \dots & D_{k_k}(r_{k_k}) & O \\ O & O & \dots & O(n_{k-1} - r_{k-1} - r_{k_k}) & O \end{bmatrix}$$

with $r_k = \sum_{i=1}^{k} r_{i_k} = rank(A_k)$ and the $r_{i_k} \times r_{i_k}$ matrices D_{i_k} are diagonal with positive diagonal elements.

- nonsingular matrices $X_j(n_j \times n_j)$ and $Z_j(n_{j-1} \times n_{j-1})$ with $j = 1, k-1$, where Z_j is either $Z_j = X_j^{-t}$ or $Z_j = X_j$ (i.e. both choices are always possible).

Proof: see [2].

Because one can choose either $Z_j = X_j^{-t}$ or $Z_j = X_j$ we can define the following GSVD-nomenclature:

Definition 2:

- If for the matrix pair (A_i, A_{i+1}) in theorem 3, we have that $Z_i = X_i$, we say that the factorization of that pair is of the **P-type**.
- On the other hand, if for a matrix pair (A_i, A_{i+1}) in theorem 3, we have that $Z_i = X_i^{-t}$ the factorization of that pair is of the **Q-type**.
- The name of a chain of factorizations of the matrices $A_i (i = 1, k)$ as in theorem 3 is then obtained by simply enumerating the different types.

Example:
A **QPQ**-SVD of a set of matrices $\{A_1(n_0 \times n_1), A_2(n_1 \times n_2), A_3(n_2 \times n_3), A_4(n_3 \times n_4)\}$ means:

$$\begin{aligned} A_1 &= U_1 D_1 X_1^{-1} \\ A_2 &= X_1^{-t} D_2 X_2^{-1} \\ A_3 &= X_2 D_3 X_3^{-1} \\ A_4 &= X_3^{-t} D_4 V_4^t \end{aligned} \quad (14)$$

with $U_1^t U_1 = I$, $V_4^t V_4 = I$ (U_1 and V_4 are square); X_1, X_2, X_3 are nonsingular and square matrices and D_1, D_2, D_3, D_4 have the same dimensions as respectively A_1, A_2, A_3, A_4.

In the following section, this GSVD-framework will be applied in order to classify different stochastic realization schemes. Due to its extremely analysing property, the GSVD will give very much geometrical and algebraical insight into the problem of stochastic realization.

5 Classification of realization schemes by means of the GSVD

Concepts of canonical correlation analysis [13], principal component analysis [7] and RV-coefficient optimization [5] were already applied to the stochastic realization problem in the past. A unifying approach is given here within a GSVD framework.

5.1 Strategy to obtain X^+ and X^-

In order to evaluate dependencies between Y^+ and Y^- we look at the following optimization problem:

$$extr_{\underline{l},\underline{m}} \{\underline{l}^t H \underline{m}\} \quad (15)$$

under constraints:

$$\begin{aligned} L^t R^+ L &= \Delta_L \\ M^t R^- M &= \Delta_M \end{aligned}$$

which is called a symmetric stochastic realization problem because the constraints of L and M have the same structure and where

- $H = Y^+(Y^-)^t$, $R^+ = Y^+(Y^+)^t$, $R^- = Y^-(Y^-)^t$
- $L(li \times li)$ and $M(li \times li)$ contain respectively the li solution vectors $\underline{l}$ and $\underline{m}$ of the optimization problem.
- Δ_L and Δ_M are diagonal matrices

An Orthogonality Theorem (see chapter 6) will tell us then how to evaluate dependencies between $L^t Y^+$ and $M^t Y^-$ and how we can select vectors $\underline{l}_i^t Y^+$ and $\underline{m}_i^t Y^-$, which contribute to respectively X^+ and X^-. Other stochastic realization problems like the unsymmetric one and the symmetric with orthonormality constraints on the solution matrices will also be studied.

5.2 A QPQ-SVD approach to the symmetric stochastic realization problem

We will now derive the solution of the following optimization problem:

$$extr_{\underline{l},\underline{m}} \{\delta_{L_k}^{-1/2} .\underline{l}^t H \underline{m}.\, \delta_{M_k}^{-1/2}\} \tag{16}$$

under constraints:

$$\begin{aligned} L^t R^+ L = \Delta_L &= diag\{\delta_{L_1}, \dots, \delta_{L_{li}}\} \\ M^t R^- M = \Delta_M &= diag\{\delta_{M_1}, \dots, \delta_{M_{li}}\} \end{aligned}$$

This stochastic realization problem is called symmetric because the constraints for L and M have the same structure. The Lagrangian function has the form:

$$\mathcal{L}(\underline{l},\underline{m},\lambda,\gamma) = \delta_{L_k}^{-1/2} .\underline{l}^t H \underline{m}.\, \delta_{M_k}^{-1/2} - \lambda.(\delta_{L_k}^{-1/2}.\, \underline{l}^t R^+ \underline{l}.\, \delta_{L_k}^{-1/2} - 1) - \gamma.(\delta_{M_k}^{-1/2}.\, \underline{m}^t R^- \underline{m}.\, \delta_{M_k}^{-1/2} - 1)$$

One can check that the constraints $\underline{l}_k^t R^+ \underline{l}_s = 0$ and $\underline{m}_k^t R^- \underline{m}_s = 0$ with $s \neq k$ are not relevant to the problem. Let us now define new variables $\underline{l}_\Delta$ and $\underline{m}_\Delta$ as:

$$\begin{aligned} \underline{l}_\Delta &= \underline{l}.\delta_{L_k}^{-1/2} \\ \underline{m}_\Delta &= \underline{m}.\delta_{M_k}^{-1/2} \end{aligned} \tag{17}$$

Thus

$$\mathcal{L}(\underline{l}_\Delta,\underline{m}_\Delta,\lambda,\gamma) = \underline{l}_\Delta^t H \underline{m}_\Delta - \lambda.(\underline{l}_\Delta^t R^+ \underline{l}_\Delta - 1) - \gamma.(\underline{m}_\Delta^t R^- \underline{m}_\Delta - 1)$$

The extremal solutions are given by:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \underline{l}_\Delta} = 0 &: \quad H\underline{m}_\Delta = 2\lambda R^+ \underline{l}_\Delta \\ \frac{\partial \mathcal{L}}{\partial \underline{m}_\Delta} = 0 &: \quad H^t \underline{l}_\Delta = 2\gamma R^- \underline{m}_\Delta \\ \frac{\partial \mathcal{L}}{\partial \lambda} = 0 &: \quad \underline{l}_\Delta^t R^+ \underline{l}_\Delta = 1 \\ \frac{\partial \mathcal{L}}{\partial \gamma} = 0 &: \quad \underline{m}_\Delta^t R^- \underline{m}_\Delta = 1 \end{aligned}$$

Clearly these equations have several solutions for $\underline{l}_\Delta$ and $\underline{m}_\Delta$, λ and γ. After some manipulations we get the following generalized eigenvalue problems:

$$\begin{aligned} H(R^-)^{-1} H^t.L_\Delta &= R^+.L_\Delta.\Lambda \\ H^t(R^+)^{-1} H.M_\Delta &= R^-.M_\Delta.\Gamma \end{aligned} \tag{18}$$

with $L_\Delta(li \times li)$ and $M_\Delta(li \times li)$ containing the solution vectors $\underline{l}_\Delta$ and $\underline{m}_\Delta$; Λ and Γ are diagonal matrices which depend on the values λ and γ. These matrices L_Δ and M_Δ are thus defined as:

$$\begin{aligned} L_\Delta &= L\Delta_L^{-1/2} \\ M_\Delta &= M\Delta_M^{-1/2} \end{aligned}$$

Equations (18) mean:

$$\begin{aligned} Y^+(Y^-)^t.[Y^-(Y^-)^t]^{-1}.Y^-(Y^+)^t.L_\triangle &= Y^+(Y^+)^t.L_\triangle.\Lambda \\ Y^-(Y^+)^t.[Y^+(Y^+)^t]^{-1}.Y^+(Y^-)^t.M_\triangle &= Y^-(Y^-)^t.M_\triangle.\Gamma \end{aligned} \tag{19}$$

Let us now take a **QPQ -SVD of the set** $\{(Y^+)^t, Y^+, (Y^-)^t, Y^-\}$ as in (14):

$$\begin{aligned} (Y^+)^t &= U_1 D_1 X_1^{-1} \\ Y^+ &= X_1^{-t} D_2 X_2^{-1} \\ (Y^-)^t &= X_2 D_3 X_3^{-1} \\ Y^- &= X_3^{-t} D_4 V_4^t \end{aligned} \tag{20}$$

From this expression (20) the numerical advantage of a GSVD can be seen: no products are made of matrices Y^+ and Y^-, but an explicit factorization is obtained and hence there will be no loss of accuracy.
Because of the orthonormality properties of U_1 and V_4 and the nonsingularity of X_1, X_2, X_3 we get:

$$\begin{aligned} (D_2 D_3).(D_4 D_4^t)^{-1}.(D_2 D_3)^t.[X_1^{-1} L_\triangle] &= (D_1^t D_1).[X_1^{-1} L_\triangle].\Lambda \\ (D_2 D_3)^t.(D_1^t D_1)^{-1}.(D_2 D_3).[X_3^{-1} M_\triangle] &= (D_4 D_4^t).[X_3^{-1} M_\triangle].\Gamma \end{aligned} \tag{21}$$

The matrices $D_1^t D_1$, $D_2 D_3$ and $D_4 D_4^t$ are diagonal matrices. If the matrices $D_1^t D_1$ and $D_2 D_3$ are equal to the identity matrix (which is normally the case), then we obtain the following eigenvalue problems:

$$\begin{aligned} (D_4 D_4^t)^{-1}.[X_1^{-1} L_\triangle] &= [X_1^{-1} L_\triangle].\Lambda \\ (D_4 D_4^t)^{-1}.[X_3^{-1} M_\triangle] &= [X_3^{-1} M_\triangle].\Gamma \end{aligned} \tag{22}$$

In order to obtain a solution of (22), which satisfies the constraints $L_\triangle^t R^+ L_\triangle = I_{li}$ and $M_\triangle^t R^- M_\triangle = I_{li}$, we apply a Cholesky factorization on $D_1^t D_1$ and $D_4 D_4^t$:

$$\begin{aligned} D_1^t D_1 &= R_L^t R_L \\ D_4 D_4^t &= R_M^t R_M \end{aligned} \tag{23}$$

Let us now change (22) into:

$$\begin{aligned} R_L (D_4 D_4^t)^{-1} R_L^{-1}.[R_L X_1^{-1} L_\triangle] &= [R_L X_1^{-1} L_\triangle].\Lambda \\ R_M (D_4 D_4^t)^{-1} R_M^{-1}.[R_M X_3^{-1} M_\triangle] &= [R_M X_3^{-1} M_\triangle].\Gamma \end{aligned} \tag{24}$$

Because of the diagonality of R_L and R_M:

$$\begin{aligned} (D_4 D_4^t)^{-1}.[R_L X_1^{-1} L_\triangle] &= [R_L X_1^{-1} l_\triangle].\Lambda \\ (D_4 D_4^t)^{-1}.[R_M X_3^{-1} M_\triangle] &= [R_M X_3^{-1} M_\triangle].\Gamma \end{aligned} \tag{25}$$

The matrix $(D_4 D_4^t)^{-1}$ is diagonal and hence there exist always the next solution:

$$\begin{aligned} \Lambda = \Gamma &= (D_4 D_4^t)^{-1} \\ R_L X_1^{-1} L_\triangle &= I_{li} \\ R_M X_3^{-1} M_\triangle &= I_{li} \end{aligned} \tag{26}$$

One can check that the constraints are satisfied. The solution matrices L_Δ and M_Δ are thus:

$$\begin{aligned} L_\Delta &= X_1 R_L^{-1} = X_1 \\ M_\Delta &= X_3 R_M^{-1} = X_3(D_4 D_4^t)^{-1/2} \end{aligned} \tag{27}$$

We see that the solution of the optimization problem can be completely written in function of the QPQ-factorization of the matrix set $\{(Y^+)^t, Y^+, (Y^-)^t, Y^-\}$ in a very elegant way. Relations between $L^t Y^+$ and $M^t Y^-$ can be detected in:

$$\begin{aligned} L_\Delta^t Y^+.(Y^-)^t M_\Delta &= X_1^t.(X_1)^{-t} D_2 X_2^{-1}.X_2 D_3 X_3^{-1}.X_3(D_4 D_4^t)^{-1/2} \\ &= (D_4 D_4^t)^{-1/2} = \Lambda^{1/2} = \Gamma^{1/2} = S = diag(s_1, \ldots, s_{li}) \end{aligned} \tag{28}$$

Note that $(s_i)_{max}$ gives the global maximum of (16) and $(s_i)_{min}$ the global minimum and all elements of S have values between 0 and 1 because of the orthonormality of $L_\Delta^t Y^+$ and $M_\Delta^t Y^-$ and they can be interpreted as the cosine of the canonical angles between the spaces $span_{row}(L_\Delta^t Y^+)$ and $span_{row}(M_\Delta^t Y^-)$.

Instead of optimizing (16), we can also look at:

$$\begin{aligned} RV(L_\Delta^t Y^+, M_\Delta^t Y^-) &= \frac{trace(L_\Delta^t H (R^-)^{-1} H^t L_\Delta)}{[trace(I_{li})^2]^{1/2}.[trace(I_{li})^2]^{1/2}} \\ &= \frac{trace(M_\Delta^t H^t (R^+)^{-1} H M_\Delta)}{[trace(I_{li})^2]^{1/2}.[trace(I_{li})^2]^{1/2}} \end{aligned} \tag{29}$$

and formulate the symmetric stochastic realization problem as:

$$\begin{aligned} extr_{\underline{l}_\Delta} \{\underline{l}_\Delta^t H (R^-)^{-1} H^t \underline{l}_\Delta\} \quad &, \quad L_\Delta^t R^+ L_\Delta = I_{li} \\ extr_{\underline{m}_\Delta} \{\underline{m}_\Delta^t H^t (R^+)^{-1} H \underline{m}_\Delta\} \quad &, \quad M_\Delta^t R^- M_\Delta = I_{li} \end{aligned} \tag{30}$$

This gives us 2 Lagrangian functions:

$$\begin{aligned} \mathcal{L}_L(\underline{l}_\Delta, \lambda) &= \underline{l}_\Delta^t H (R^-)^{-1} H^t \underline{l}_\Delta - \lambda.(\underline{l}_\Delta^t R^+ \underline{l}_\Delta - 1) \\ \mathcal{L}_M(\underline{m}_\Delta, \gamma) &= \underline{m}_\Delta^t H^t (R^+)^{-1} H \underline{m}_\Delta - \gamma.(\underline{m}_\Delta^t R^- \underline{m}_\Delta - 1) \end{aligned} \tag{31}$$

The solution to the problem is given by:

$$\begin{aligned} \frac{\partial \mathcal{L}_L}{\partial \underline{l}_\Delta} = 0 &: \quad H(R^-)^{-1} H^t.\underline{l}_\Delta = \lambda R^+.\underline{l}_\Delta \\ \frac{\partial \mathcal{L}_M}{\partial \underline{m}_\Delta} = 0 &: \quad H^t(R^+)^{-1} H.\underline{m}_\Delta = \gamma R^-.\underline{m}_\Delta \\ \frac{\partial \mathcal{L}_L}{\partial \lambda} = 0 &: \quad \underline{l}_\Delta^t R^+ \underline{l}_\Delta = 1 \\ \frac{\partial \mathcal{L}_M}{\partial \gamma} = 0 &: \quad \underline{m}_\Delta^t R^- \underline{m}_\Delta = 1 \end{aligned} \tag{32}$$

We can conclude that this is the same eigenvalue problem as (18).

A third approach to the symmetric stochastic realization problem is based on obtaining X^+

and X^- respectively as a projection of $(Y^+)^t$ on $(Y^-)^t$ and $(Y^-)^t$ on $(Y^+)^t$. These projections are equal to:

$$\begin{aligned} P^+ &= (Y^+)^t|(Y^-)^t = (Y^-)^t(R^-)^{-1}H^t \\ P^- &= (Y^-)^t|(Y^+)^t = (Y^+)^t(R^+)^{-1}H \end{aligned} \tag{33}$$

Here P^+ is the projection of $span_{col}((Y^+)^t)$ on $span_{col}((Y^-)^t)$ and P^- of $span_{col}((Y^-)^t)$ on $span_{col}((Y^+)^t)$. These projections are related to a linear least squares solution of:

$$\begin{aligned} (Y^-)^t P^+ &= (Y^+)^t \\ (Y^+)^t P^- &= (Y^-)^t \end{aligned} \tag{34}$$

We can approach the symmetric stochastic realization problem then as:

$$extr_{\underline{l}_\Delta, \underline{m}_\Delta} \{\underline{m}_\Delta^t (P^-)^t P^+ \underline{l}_\Delta\} \tag{35}$$

under constraints

$$\begin{aligned} L_\Delta^t (P^-)^t P^- L_\Delta &= I_{l_i} \\ M_\Delta^t (P^+)^t P^+ M_\Delta &= I_{l_i} \end{aligned}$$

It is left to the reader to check that we obtain the same eigenvalue problem as (18) and the same solution as (27).

5.3 A PQ-SVD approach to the unsymmetric stochastic realization problem

In analogy with the RV-coefficient approach to the symmetric stochastic realization problem, this unsymmetric problem can be stated as:

$$\begin{aligned} extr_{\underline{l}_\Delta} \{\underline{l}_\Delta^t H (R^-)^{-1} H^t \underline{l}_\Delta\} \quad &, \quad L_\Delta^t L_\Delta = I_{l_i} \\ extr_{\underline{m}_\Delta} \{\underline{m}_\Delta^t H^t H \underline{m}_\Delta\} \quad &, \quad M_\Delta^t R^- M_\Delta = I_{l_i} \end{aligned} \tag{36}$$

This stochastic realization problem is called unsymmetric because the constraints for L and M have not the same structure. Note that the role of L_Δ and M_Δ may be exchanged in (36). The Lagrangian functions of the problem are:

$$\begin{aligned} \mathcal{L}_L(\underline{l}_\Delta, \lambda) &= \underline{l}_\Delta^t H (R^-)^{-1} H^t \underline{l}_\Delta - \lambda.(\underline{l}_\Delta^t \underline{l}_\Delta - 1) \\ \mathcal{L}_M(\underline{m}_\Delta, \gamma) &= \underline{m}_\Delta^t H^t H \underline{m}_\Delta - \gamma.(\underline{m}_\Delta^t R^- \underline{m}_\Delta - 1) \end{aligned} \tag{37}$$

The solution has to satisfy:

$$\begin{aligned} \frac{\partial \mathcal{L}_L}{\partial \underline{l}_\Delta} = 0 \quad &: \quad H(R^-)^{-1} H^t.\underline{l}_\Delta = \lambda.\underline{l}_\Delta \\ \frac{\partial \mathcal{L}_M}{\partial \underline{m}_\Delta} = 0 \quad &: \quad H^t H \underline{m}_\Delta = \gamma.R^- \underline{m}_\Delta \\ \frac{\partial \mathcal{L}_L}{\partial \lambda} = 0 \quad &: \quad \underline{l}_\Delta^t \underline{l}_\Delta = 1 \\ \frac{\partial \mathcal{L}_M}{\partial \gamma} = 0 \quad &: \quad \underline{m}_\Delta^t R^- \underline{m}_\Delta = 1 \end{aligned} \tag{38}$$

With the same notations as for the symmetric problem we obtain the following eigenvalue problems:

$$\begin{aligned} H(R^-)^{-1}H^t.L_\Delta &= L_\Delta.\Lambda \\ H^tH.M_\Delta &= R^-M_\Delta.\Gamma \end{aligned} \tag{39}$$

Let us now take a **PQ -SVD of the matrix set** $\{Y^+,(Y^-)^t,Y^-\}$.

$$\begin{aligned} Y^+ &= U_1D_1X_1^{-1} \\ (Y^-)^t &= X_1D_2X_2^{-1} \\ Y^- &= X_2^{-t}D_3V_3^t \end{aligned} \tag{40}$$

Thanks to the orthonormality of U_1 and V_3 and the nonsingularity of the matrices X_1 and X_2, (39) can be written as:

$$\begin{aligned} (D_1D_2).(D_3D_3^t)^{-1}(D_1D_2)^t.[U_1^tL_\Delta] &= [U_1^tL_\Delta].\Lambda \\ (D_1D_2)^t.(D_1D_2).[X_2^{-1}M_\Delta] &= (D_3D_3^t).[X_2^{-1}M_\Delta].\Gamma \end{aligned} \tag{41}$$

Normally D_1D_2 is the identity matrix. The matrix $D_2D_3^t$ is diagonal and after applying a Cholesky factorization $D_3D_3^t$

$$D_3D_3^t = R_M^tR_M \tag{42}$$

we obtain

$$\begin{aligned} (D_3D_3^t)^{-1}.[U_1^tL_\Delta] &= [U_1^tL_\Delta].\Lambda \\ (D_3D_3^t)^{-1}.[R_MX_2^{-1}M_\Delta] &= [R_MX_2^{-1}M_\Delta].\Gamma \end{aligned} \tag{43}$$

Because of the diagonality of $(D_3D_3^t)^{-1}$ there always exist a solution of the form:

$$\begin{aligned} \Lambda = \Gamma &= (D_3D_3^t)^{-1} \\ U_1^tL_\Delta &= I_{li} \\ R_MX_2^{-1}M_\Delta &= I_{li} \end{aligned} \tag{44}$$

The solution to (36) is then:

$$\begin{aligned} L_\Delta &= U_1 \\ M_\Delta &= X_2(D_3D_3^t)^{-1/2} \end{aligned} \tag{45}$$

The solution of the optimization problem can thus be written as a function of the PQ-factorization of the matrix set $\{Y^+,(Y^-)^t,Y^-\}$. We can remark that for the unsymmetric stochastic realization problem:

$$L_\Delta^tY^+.(Y^-)^tM_\Delta = (D_3D_3^t)^{-1/2} = \Lambda^{1/2} = \Gamma^{1/2} = S \tag{46}$$

The diagonal elements of S are not necessarily lying between 0 and 1 in this case, because the orthonormality of $L_\Delta^tY^+$ was not required.

5.4 A PSVD approach to the symmetric stochastic realization problem with orthonormality constraints to the solution matrices

The symmetric stochastic realization problem with orthonormality constraints to the solution matrices, formulated with the RV-coefficient approach, takes the form:

$$\begin{aligned} extr_{\underline{l}_\triangle} \{\underline{l}_\triangle^t H H^t \underline{l}_\triangle\} \quad &, \quad L_\triangle^t L_\triangle = I_{l_i} \\ extr_{\underline{m}_\triangle} \{\underline{m}_\triangle^t H^t H \underline{m}_\triangle\} \quad &, \quad M_\triangle^t M_\triangle = I_{l_i} \end{aligned} \tag{47}$$

This problem is in accordance with the principal components approach (PC-H) of Arun and Kung [7]. The derivatives of the Lagrangian functions propose the following solution:

$$\begin{aligned} H H^t . L_\triangle &= L_\triangle . \Lambda \\ H^t H . M_\triangle &= M_\triangle . \Gamma \end{aligned} \tag{48}$$

After applying a **PSVD to the matrix pair** $\{Y^+, (Y^-)^t\}$:

$$\begin{aligned} Y^+ &= U_1 D_1 X_1^{-1} \\ (Y^-)^t &= X_1 D_2 V_2^t \end{aligned} \tag{49}$$

we get:

$$\begin{aligned} (D_1 D_2).(D_1 D_2)^t.[U_1^t L_\triangle] &= [U_1^t L_\triangle].\Lambda \\ (D_1 D_2)^t.(D_1 D_2).[V_2^t M_\triangle] &= [V_2^t M_\triangle].\Gamma \end{aligned} \tag{50}$$

Because of the diagonality of $D_1 D_2$, there exist always a solution of the form:

$$\begin{aligned} \Lambda = \Gamma &= (D_1 D_2)^2 \\ L_\triangle &= U_1 \\ M_\triangle &= V_2 \end{aligned} \tag{51}$$

Let us remark that:

$$L_\triangle^t Y^+ .(Y^-)^t M_\triangle = D_1 D_2 = \Lambda^{1/2} = \Gamma^{1/2} = S$$

As in the PQ-case, the diagonal elements of S are not necessarily lying between 0 and 1, because the orthonormality of $L_\triangle^t Y^+$ and $M_\triangle^t Y^-$ was not required.

Finally, the results of this chapter are summarized here in a table.

	Constraints	GSVD-configuration
Symm.	$L^t R^+ L = \Delta_L$ $M^t R^- M = \Delta_M$	QPQ-SVD of $\{(Y^+)^t, Y^+, (Y^-)^t, Y^-\}$
Unsymm.	$L^t L = \Delta_L$ $M^t R^- M = \Delta_M$	PQ-SVD of $\{Y^+, (Y^-)^t, Y^-\}$
Symm.	$L^t L = \Delta_L$ $M^t M = \Delta_M$	PSVD of $\{Y^+, (Y^-)^t\}$

6 Estimation of the process order and obtaining a state vector sequence

In the previous section, 3 approaches to the stochastic realization problem were proposed:

1. a **QPQ-SVD** approach to the symmetric stochastic realization problem
2. a **PQ-SVD** approach to the unsymmetric stochastic realization problem
3. a **PSVD** approach to the symmetric stochastic realization problem with orthonormality constraints on the solution matrices

From the viewpoint of estimating the stochastic process order and obtaining a state vector sequence, the first approach is most important. Indeed, let us interpret the result (27). Because of the orthonormality of $L_\Delta^t Y^+$ and $M_\Delta^t Y^-$, an OSVD of $L_\Delta^t H M_\Delta$:

$$L_\Delta^t H M_\Delta = RSV^t$$

will have a S-matrix with the same diagonal elements as $(D_A D_A^t)^{-1/2}$ (eventually the sequence can change) and there exists always a possible OSVD with $R = T = I_{li}$. The matrix S contains thus the *canonical angles* between $L_\Delta^t Y^+$ and $M_\Delta^t Y^-$ and because of the diagonality of $L_\Delta^t H M_\Delta$ we can interpret directly the angles between $\underline{l}_{\Delta_i}^t Y^+$ and $\underline{m}_{\Delta_i}^t Y^-$. The state vector sequences X^+ and X^- are then determined by selecting those $\underline{l}_{\Delta_i}^t Y^+$ and $\underline{m}_{\Delta_i}^t Y^-$, which are not orthogonal to each other (approximately). This part of the matrices L_Δ and M_Δ will be respectively stored in F and G. The matrices X^+ and X^- are then equal to $X^+ = F^t Y^+$ and $X^- = G^t Y^-$ with $F^t = L^t(1:n,:)$ and $G^t = M^t(1:n,:)$ (suppose that the elements of S are ordered as in an OSVD).
Instead of looking at an intersection between $span_{row}(L_\Delta^t Y^+)$ and $span_{row}(M_\Delta^t Y^-)$ in order to determine X^+ and X^-, we evaluate dependencies between $L^t Y^+$ and $M^t Y^-$ and apply the following philosophy with respect to our linear stochastic model:

Orthogonality is absence of linear dependencies

and

The linear model explains all linear dependencies between X^+ and X^-

thus

Orthogonal vectors $\underline{l}_{\Delta_i}^t Y^+$ and $\underline{m}_{\Delta_i}^t Y^-$ do not contribute to the model

The following theorem (Orthogonality Theorem) will tell us that this philosophy is only valid if $j \gg li$, thus if we take many measurement data into account.

Theorem 4: Orthogonality Theorem
Consider a fixed subspace $\mathcal{S}^r$ in the j dimensional vectorspace of j-tuples $\mathcal{R}^j$ ($j \gg r$). Assume that a random vector $\underline{v}$ is generated with equal probability for all directions in $\mathcal{R}^j$. Then the directional probability density function $p(\alpha, r, j)$ of the angle α between $\underline{v}$ and the fixed subspace $\mathcal{S}^r$ is given by:

$$p(\alpha, r, j) = K(r, j).(sin\alpha)^{j-r-1}(cos\alpha)^{r-1}$$

where $K(r,j) = 2\frac{C_r C_{j-r}}{C_j}$ and the C_k can be obtained recursively as:

- $C_1 = 1$
- $C_2 = \pi$
- $C_k = \frac{2\pi}{k-2} C_{k-2}$, $k > 2$

Remark: the angle α is measured between $\underline{v}$ and its orthogonal projection on $\mathcal{S}^r$.
Proof: see [3]. □

In figure 1 an example of this Orthogonality Theorem is given for $r = 2$. Let us now apply this theorem to our stochastic realization problem. Let $\mathcal{S}^r (r = li)$ be equal to $span_{row}(M^t_\Delta Y^-)$; the vectors $\underline{l}^t_{\Delta_i} Y^+$ play then the role of $\underline{v}$ in the case that they do not contribute to the model. However the angle α does not correspond to the angle between $\underline{l}^t_{\Delta_i} Y^+$ and $\underline{m}^t_{\Delta_i} Y^-$, we may conclude that for large values of j orthogonal vectors $\underline{l}^t_{\Delta_i} Y^+$ and $\underline{m}^t_{\Delta_i} Y^-$ do not contribute to the model. It is also possible to allocate $span_{row}(L^t_\Delta Y^+)$ to $\mathcal{S}^r$. An $n \times n$ diagonal matrix is then obtained as:

$$S_x = X^+(X^-)^t$$

Definition 3:
The matrix S in :

$$L^t_\Delta H M_\Delta = S = (D_A D^t_A)^{-1/2}$$

is called the *interaction* matrix.

This definition arises from the fact that S measures the interaction between the past and the future of the process [10]. So much canonical angles between $span_{row}(L^t_\Delta Y^+)$ and $span_{row}(M^t_\Delta Y^-)$ near to 90^o refer to a low process order and a *low* interaction level between past and future of the process and less near to 90^o refers to a *high* process order.
Now we will give an example of estimating the process order from the interaction matrix.

Example 1:
In figure 2 the canonical angles between $L^t_\Delta Y^+$ and $M^t_\Delta Y^-$ are calculated for different values of j for artificially constructed stochastic processes of order 2 and 3 with 2 outputs; li was chosen equal to 8, $\Delta_L = \Delta_M = I_{li}$. We may conclude that only for $j \gg$, a senseful order estimation is possible, which is an illustration of the Orthogonality Theorem.

Example 2:
In figure 3 a simulation is given of a first order stochastic process

$$\begin{aligned} x_{k+1} &= 0.7x_k + v_k \\ y_k &= 0.5x_k \end{aligned}$$

with $i = 6, k = 7, l = 1, j = 200$. All data before the vertical line were used for the estimation of the model. The process order was estimated equal to 1.

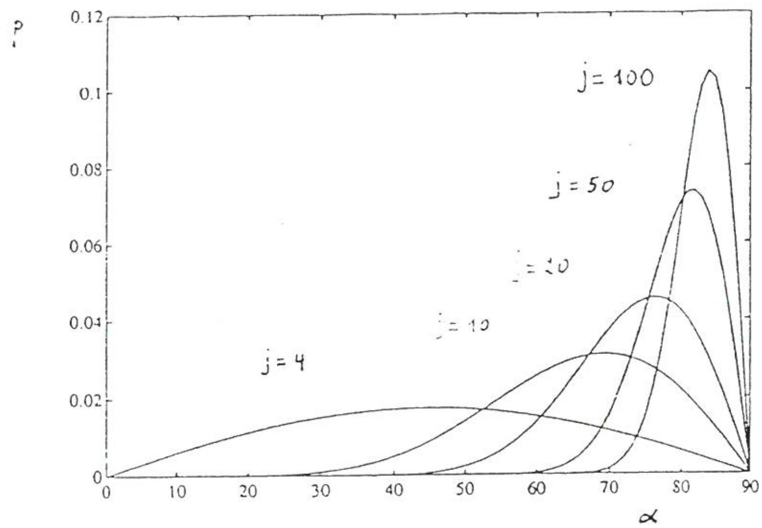
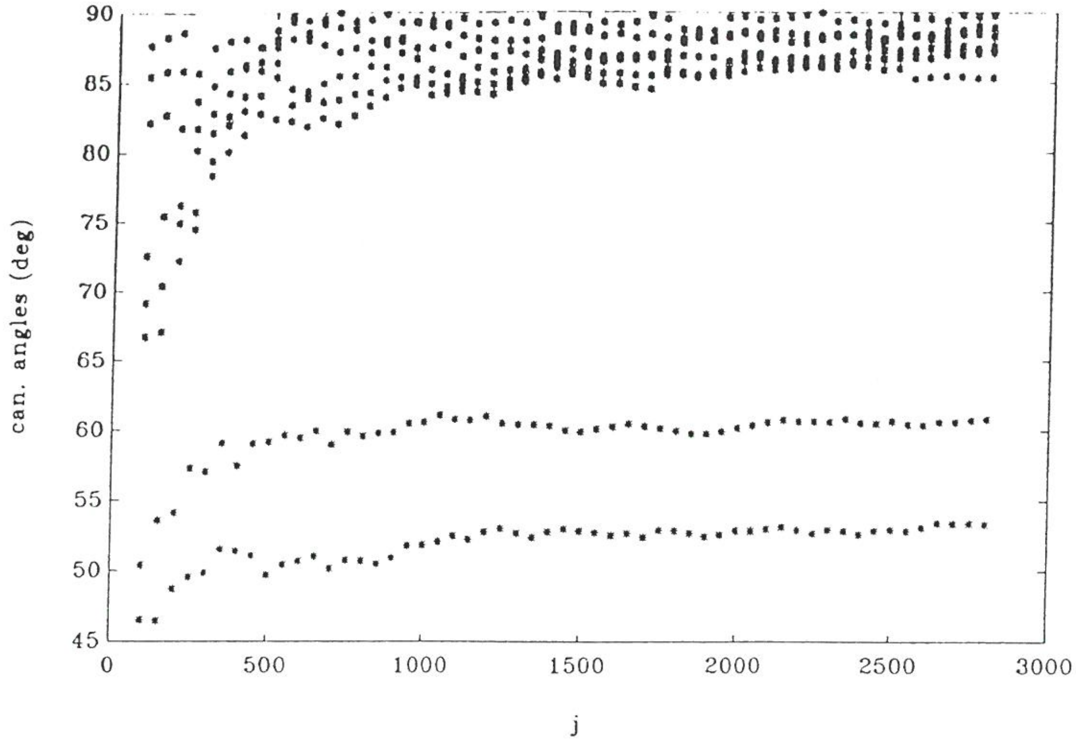


Figure 1: *Probability density function of angle between random vector and fixed subspace S^2 for different values of j*

a.



b.

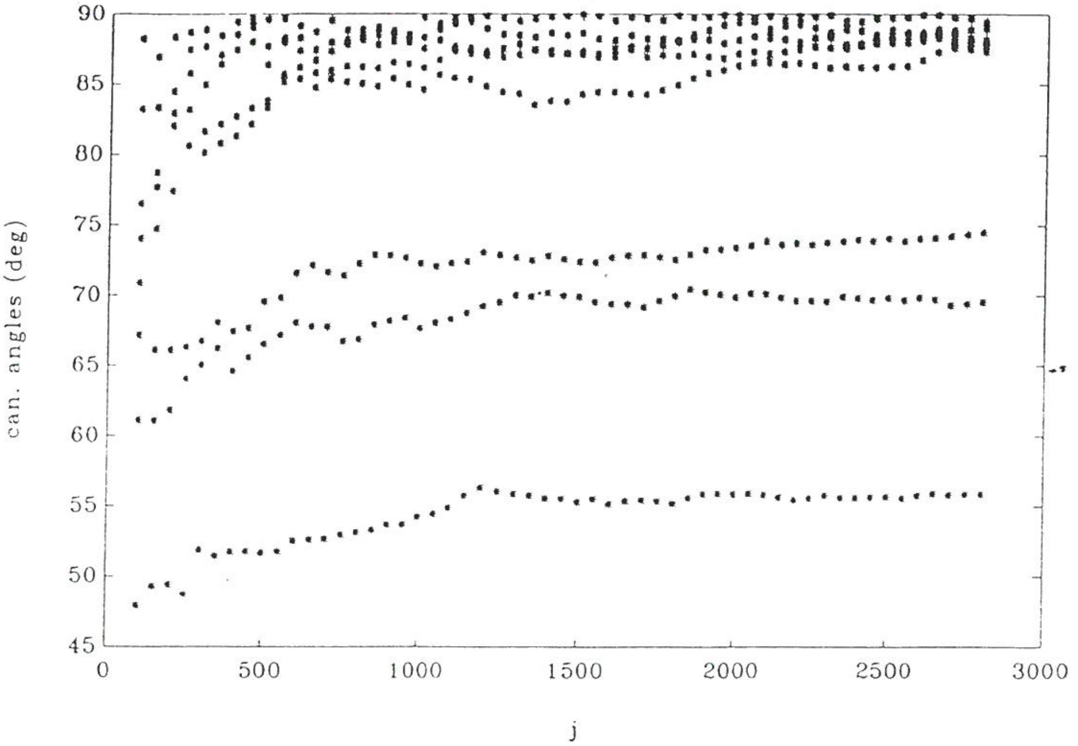


Figure 2: *Stochastic process order estimation from the interaction matrix, fig.a. process order is 2, fig.b. process order is 3 ($li = 8, l = 2$)*

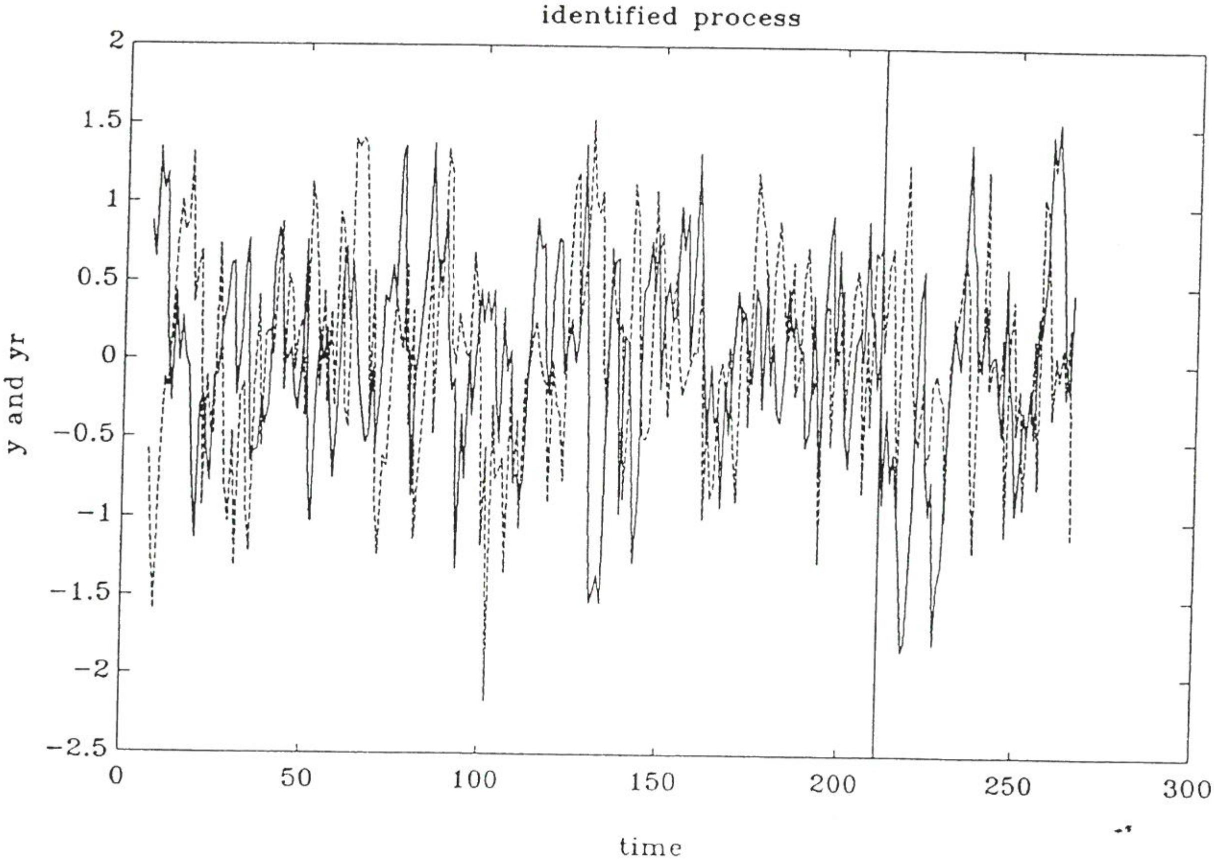


Figure 3: *Original and simulated output of stochastic process:* $x_{k+1} = 0.7x_k + v_k$, $y_k = 0.5x_k$ *with* $i = 6, k = 7, l = 1, j = 200$ *(full line = original).*

7 Stochastic realization algorithm

Finally, we propose as a summary an algorithm to approach the symmetric stochastic realization problem:

1. Given the output process $\{\underline{y}_k\}$, construct the past and future output matrices Y^- and Y^+.

2. Compute a QPQ-SVD of the matrix set $\{(Y^+)^t, Y^+, (Y^-)^t, Y^-\}$:

$$\begin{aligned} (Y^+)^t &= U_1 D_1 X_1^{-1} \\ Y^+ &= X_1^{-t} D_2 X_2^{-1} \\ (Y^-)^t &= X_2 D_3 X_3^{-1} \\ Y^- &= X_3^{-t} D_4 V_4^t \end{aligned}$$

3. Obtain the interaction matrix $S = (D_4 D_4^t)^{-1/2}$ in order to estimate the process order. Select the vectors $\underline{l}^t Y^+$ and $\underline{m}^t Y^-$ belonging to the largest diagonal elements of S (the number of selected vectors is determined by the Orthogonality Theorem) and construct with those vectors X^+ and X^-:

$$\begin{aligned} X^+ &= F^t Y^+ \\ X^- &= G^t Y^- \end{aligned}$$

where $L = X_1$, $M = X_3.S$, $F^t = L^t(1:n,:)$, $G^t = M^t(1:n,:)$ (suppose that the elements of S are ordered as in an OSVD)

4. Take the following QR-factorization:

$$[(X^+)^t\ (Y^+)^t] = [Q_1\ Q_2].\begin{bmatrix} R_{11} & R_{12} \\ O & R_{22} \end{bmatrix}$$

where

$$\Gamma^+ = (R_{11}^{-1} R_{12})^t$$

The matrices A and C of the state space model of the stochastic process can then be obtained easily from Γ^+, because of its shift structure.

$$\begin{aligned} A &= (\underline{\Gamma^+})^+ . \overline{\Gamma^+} \\ C &= \Gamma^+(1:l,:) \end{aligned}$$

with

$$\underline{\Gamma^+} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{i-2} \end{bmatrix}, \quad \overline{\Gamma^+} = \begin{bmatrix} CA \\ CA^2 \\ \vdots \\ CA^{i-1} \end{bmatrix}$$

5. When you use the model for simulation take then $\underline{v}_k$ and $\underline{w}_k$:

$$\underline{v}_k = \frac{1}{\sqrt{N}}.rand(n, N), \quad \underline{w}_k = \frac{1}{\sqrt{N}}.rand(l, N)$$

where $rand$ is a random matrix (see comments of (1)) and N is the total number of samples to be displayed. The division by $\sqrt{N}$ is necessary , because otherwise theorem 2 does not hold.

Remark:
Instead of taking a QPQ-SVD in step 2 of this algorithm, we can get the same result with 3 OSVD's:

$$\begin{aligned}(Y^+)^t &= U_1\Sigma_1(V_1)^t \\ (Y^-)^t &= U_2\Sigma_2(V_2)^t \\ U_1^tU_2 &= USV^t\end{aligned}$$

where S has the same meaning as in (28). The matrices X^+ and X^- are then:

$$\begin{aligned}X^+ &= U^{-1}(1:n,:)U_1^t \\ X^- &= V^{-1}(1:n,:)U_2^t\end{aligned}$$

8 Conclusions

The Markovian representation of a stochastic process was written as a matrix I/O-equation. Under certain conditions a state vector sequence can be obtained as a transformation of the past or future output matrices. This transformation matrix follows then from an optimization problem.
It is shown that the GSVD is very useful as an extremely analysing tool in the problems of symmetric and unsymmetric stochastic realization. It proposes a classification for different possible realization schemes together with a tool for computing them in a numerical reliable way. An interaction matrix, which is important for the process order estimation can also be written in an elegant way thanks to the GSVD.

References

[1] De Moor B., Golub G., *Generalized singular value decomposition: a proposal for a standardized nomenclature*, ESAT-SISTA report 1989-10, Department of Electrical Engineering, Kath. Univ. Leuven (also Numerical Analysis Project Manuscript NA-89-04, Department of Computer Science, Stanford University), April 1989.

[2] De Moor B., Zha H., *A tree of generalizations of the ordinary singular value decomposition*, ESAT-SISTA Report 1989-21, Department of Electrical Engineering, Kath. Univ. Leuven, Oct. 1989.

[3] De Moor B., *Mathematical concepts and techniques for modelling of static and dynamic systems*, Ph.D. Thesis, Kath. Univ. Leuven, June 1988.

[4] Robert P., Escoufier Y., *A unifying tool for linear multivariate statistical methods: the RV-coefficient*, Appl. Statist., vol.25, No.3, pp.257-265, 1976.

[5] Ramos J., Verriest E., *A unifying tool for comparing stochastic realization algorithms and model reduction techniques*, Proc. Amer. Contr. Conf., San Diego, CA, pp.150-155, 1984.

[6] Rao C., *The use and interpretation of principal component analysis in applied research*, Sankhya A, Vol.26, pp.329-358, 1966.

[7] Arun K., Kung S., *Generalized principal component analysis and its application in approximate stochastic realization*, in Modelling and application of stochastic processes, Edited by Desai U., 1986.

[8] Akaike H., *Stochastic theory of minimal realization*, IEEE Trans. Automat. Contr., Vol. AC-19, pp. 667-674, June 1974.

[9] Desai U., Pal D., Kirkpatrick R., *A realization approach to stochastic model reduction*, Int. J. Control, Vol.42, No.4, 821-838, 1985.

[10] Avniel Y., *The scattering matrix associated with a stationary stochastic process: system theoretic properties and role in realization*, in Modelling and application of stochastic processes, Edited by Desai U., 1986.

[11] Moore B., *Singular value analysis of linear systems, pt. I*, Department of electrical engineering, Systems Control report 7081, University of Toronto, Ontario, Canada, 1978.

[12] Bjork A., Golub G., *Numerical methods for computing angles between linear subspaces*, Math. Comp., vol.27, pp.579-594, 1973.

[13] Hotelling H., *Relations between two sets of variates*, Biometrika 28, pp.321-372, 1936.

[14] Golub G., Van Loan C., *Matrix Computations*, John Hopkins University Press, Baltimore M.D., 1983.