CONCLUDING REMARKS

We have shown that the impulse response of a linear time-invariant filter h(t) can be recovered from the covariance of the (nonstationary) process obtained when this filter is excited with a unilateral white noise of arbitrary distribution. Moreover, this is true even if the filter is not minimum phase or even unstable. We have also indicated how to estimate the required covariance function from (multichannel) output data alone.

There are still many practical difficulties that need to be resolved. First, it would desirable to extend our results to the case of non-zero initial conditions. Second, our procedure should simplify for filters with a rational transfer function, attempting to produce directly the two polynomials rather than the (infinite) impulse response. Third, effects of measurement and computational errors need to be explicitly addressed.

REFERENCES

1. Astrom, K.J., P. Hazader and J. Sternby (1984). Zeros of sampled system. Automatica , Vol. 20, No. 1.
2. Benveniste , A., M. Goursat and G. Ruget (1980). Robust identification of a nonminimum phase: blind adjust of linear equalizer in data communication. IEEE Trans. on Automatic Control. Vol. AC-25, No. 3.
3. Chu J. J. and Y. Wang (1985). Parameter estimation and identification of nonminimum phase linear stochastic systems. IFAC Identification and System Parameter Estimation York, UK, 3-7 July 1985 991-996.
4. Kailath, T. (1973). Some new algorithms for recursive estimation in constant linear systems. IEEE Trans. on Information Theory. Vol. IT-19, 750-760.
5. Kailath, T., S. Y. Kung and M. Morf(1979). Displacement Ranks of Matrix and Linear Equations. J. Math. Anal. and Appl. Vol. 68, No. 2, 395-407.
6. Kailath, T.,(1987). Signal Processing Applications of Some Moment Problems. in Moments in Mathematics, ed. H. Landau Amer. Math. Society, Providence, RI, 1987.
7. Kormylo, J.J. and J. Mendel (1983). Maximum-likelihood seismic deconvolution. IEEE Trans. on Geosci. Electron., Vol. GE-21, No.1, 72-81.
8. Lev-Ari, H. (1983). Nonstationary lattice-filtering modeling. Ph.D. dissertation, Stanford University, Stanford, CA, 1983
9. Lev-Ari, H. and T. Kailath, (1984). Lattice Filter Parametrization and Modeling of Nonstationary Processes. IEEE Trans. on Information Theory. Vol. IT-30, No. 1, 2-16.
10. Lii, K.J. and M. Rosenblatt (1982). Deconvolution and estimation of transfer function phase and coefficient for non-Gaussian linear process. The Annals of Statistics. Vol. 10, No. 4
11. Matsuoka,T. and T. J. Ulrych (1984). Phase estimation using the bispectrum. Proceedings of the IEEE. October, 1984.
12. Rosenblatt, M. (1980). Linear processes and bispectra. J. Appl. Prob. Vol. 17,, No. 2 , 265-270.
13. Scargle,J.D. (1981). Phase-sensitive deconvolution to model random process. Applied Time Series Analysis Vol. 2, edited by D. F. Findley, Academic Press, 1981, 549-564.
14. Wang, Y., P. F. Yan, J. J. Chu and T. Chang,(1987). Identification of nonminimum phase system with multichannel nonstationary output data. Reprinted Proceeding of IFAC 10th Congress Vol.10, Munich, FRG. July 1987. pp346-351.

A GEOMETRICAL STRATEGY FOR THE IDENTIFICATION OF STATE SPACE MODELS OF LINEAR MULTIVARIABLE SYSTEMS WITH SINGULAR VALUE DECOMPOSITION

B. De Moor, J. Vandewalle, M. Moonen*, L. Vandenberghe* and P. Van Mieghem

ESAT Katholieke Universiteit Leuven, K. Mecierlaan 94, 3030 Heverleeu, Belgium
**Supported by the N.F.W.O. (Belgian National Fund for Scientific Research)*

Abstract

In this paper, some geometrically inspired concepts are studied for the identification of models for multivariable linear time invariant systems from noisy input - output observations. Starting from a fundamental highly structured input-output matrix equation, it is shown how the singular value decomposition allows to estimate the order of the observable part of the system and its state space model matrices. Moreover, conditions for persistance of excitation of the inputs and the behavior of the algorithm when the data are perturbed by noise, can easily be studied from a geometrical point of view . The singular values allow to quantify these concepts. An example of an industrial plant identification is presented.
Keywords: Linear and Total linear least squares, identification, singular value decomposition, observable modes.

1 Introduction

The selection and the identification of appropriate mathematical representations are of central importance in the analysis, design and control of multivariable systems. With acces only to the external input-outputbehavior of a multivariable dynamical process, the internal structure (other than a priori assumed time invariance and linearity being unknown), the problem of constructing a model is a highly non-trivial task. Because of this complexity, reliable and robust general purpose identification schemes have not yet become a standard tool. In most cases, (experimental) observations on the input-output behavior of the system under normal operating conditions are readily available. The most obvious choice for a mathematical model is in a lot of cases a state space representation since the major part of modern system and control theory, such as the design of observers , filters and optimal controllers regards this very efficient and compact representation. In this paper, a new geometrically inspired identification scheme will be presented . It makes use of the numerically reliable key technique of the singular value decomposition and allows to estimate the order of the system under study and to identify its state space model matrices, from possibly noise corrupted multiple input-output measurements. No a priori paramatrization , which may be ill-conditioned with respect to identification, is required, in contrast to the identification techniques reported in [5] [12].The inspiration for the identification approach derived in this paper, can be found in the work reported in [1] [4][7][9][10][11][15]. This paper is organised as follows: In section 2, an important input - output matrix equation is derived, relating input measurements with output measurements in a structured way. The geometrical interpretation in terms of row spaces of the involved matrices, is emphasized. In section 3, the main properties of the singular value decomposition are briefly summarized while in section 4, two different versions of the identification technique are derived. Some more details about the results to be expected when the input-output observations are noisy, are reported in section 5, together with some robustness results of the new approach. In section 6, we present an example of an industrial plant identification.

2 An important input-output matrix equation and its geometrical interpretation.

In this section, a crucial input-output relation will be derived. It is essential for the identification approach to be presented furtheron. We consider linear, discrete time , multivariable time invariant systems with m inputs and l outputs, with state space representation:

$$x[k+1]_{n\times 1} = A_{n\times n}.x[k]_{n\times 1} + B_{n\times m}.u[k]_{m\times 1} \quad (1)$$
$$y[k]_{l\times 1} = C_{l\times n}.x[k]_{n\times 1} + D_{l\times m}.u[k]_{l\times 1} \quad (2)$$

(where necessary, matrix and vector dimensions will be indicated) The matrices A, B, C, D are real, the index k denotes the discrete time and the vectors $u[k], y[k]$ and $x[k]$ are the input, the output and the state at time k. Furthermore, we will also frequently use the set of Markovparamaters H_i of the linear system, defined by $H_0 = D$, $H_i = C.A^{i-1}.B$ $(i > 0)$.All matrices in this paper are assumed to be real.
From manipulation of the state space system description, one can easily obtain the following important input-output matrix equation :

$$Y_h(k,i,j) = \Gamma(i).X(k,j) + H_{tl}(i).U_h(k,i,j) \quad (3)$$

(The subscript h denotes that the matrix has block Hankelstructure , tl is block Toeplitz lower triangular, tu block Toeplitz upper triangular.) The matrices have the following structure:

$$Y_h(k,i,j) = \begin{bmatrix} y[k] & y[k+1] & \dots & y[k+j-1] \\ y[k+1] & y[k+2] & \dots & y[k+j] \\ y[k+2] & y[k+3] & \dots & y[k+j+1] \\ \dots & \dots & \dots & \dots \\ y[k+i-1] & y[k+i] & \dots & y[k+j+i-2] \end{bmatrix}$$

This is a $li \times j$ block Hankelmatrix constructed of $i+j-1$ consecutive output vectors. The matrix $U_h(k,i,j)$ has a similar structure with input vectors. The matrix $H_{tl}(i)$ is a $li \times mi$ lower triangular block Toeplitzmatrix that contains the i first Markovparameters :

$$H_{tl}(i) = \begin{bmatrix} D & 0 & 0 & \dots & 0 \\ CB & D & 0 & \dots & 0 \\ CAB & CB & D & \dots & 0 \\ CA^2B & CAB & CB & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ CA^{i-2}B & CA^{i-3}B & CA^{i-4}B & \dots & D \end{bmatrix}$$

$\Gamma(i)$ is the $il \times n$ 'extended' observability matrix.

$$\Gamma(i) = \begin{bmatrix} C \\ CA \\ CA^2 \\ \dots \\ CA^{i-1} \end{bmatrix}$$

and $X(k,j) = [x[k]x[k+1]\dots x[k+j-1]]$ is the $n \times j$ matrix containing j consecutive states. The geometrical observations to be presented will be the key tool in the analysis of the algorithm. They will provide insight in the mechanism of estimating the dimension of the observable part of the state space, in the conditions for persistant excitation of the input sequence, which has to ensure a reliable identification and in the noise sensitivity of the results. Denote by $\text{span}_{\text{row}}(U_h), \text{span}_{\text{row}}(Y_h)$ the rowspace of the row vectors of U_h, Y_h. It is assumed that the dimensions of the matrices U_h and Y_h are chosen such that $j > \max(li, mi)$ and $(j - mi) > n$. The matrix input - output relation can be interpreted as follows: $\text{span}_{\text{row}}(Y_h)$ is the sum of two rowspaces:
— the row space of the row vectors of $\Gamma(i).X(k,j)$: $\Gamma(i)$ makes linear combinations of the rows of $X(k,j)$. It filters out the non-observable parts of the state . Hence,
$\text{span}_{\text{row}}(\Gamma(i).X(k,j)) = \text{span}_{\text{row}}$[projection of $X(k,j)$ on observable subspace of state space]
— the row space of the row vectors of $H_{tl}.U_h$.
This geometrical visualization of the input-outputrelations immediately suggest the following:
— the dimension of the observable part of the state space can be estimated from the dimension of the projection of the output rowspace upon the orthogonal complement of the row space of the input block Hankel matrix and this under some (general) conditions yet to be determined (section 5.1) The part of the rowspace of $X(k,j)$ orthogonal to the row space of U_h can be considered as 'new' information, caused by the dynamical action of the system. This information is not contained in the input. Hence, the part of the rowspace of Y_h which is orthogonal to the rowspace of U_h contains information about the dynamics of the system.
-- the more orthogonal the rowspace of $X(k,j)$ to the rowspace of U_h, the more information that is available in the projection. Hence, it can be expected and will be confirmed furtheron, that the inputs are satisfactory for the identification whenever the rowspace of $X(k,j)$ is "sufficiently" orthogonal to $\text{span}_{\text{row}}(U_h)$.

3 The singular value decomposition (SVD).

While the geometrical visualization is crucial from a conceptual point of view, the singular value decomposition (SVD) is the key instrument in the explicit numerical solution of this identification scheme. In this paper, we will only give a brief summary of its main properties. Much more details can be found in [3].

The Autonne-Eckart-Young theorem *(restricted to real matrices)*
Every real $m \times n$ matrix A can be decomposed in three real matrices:

$$A_{m \times n} = U_{m \times m}.\Sigma_{m \times n}.V^t_{n \times n}$$

where U and V are orthogonal: $U.U^t = U^t.U = I_m$, $V.V^t = V^t.V = I_n$ and Σ is real, pseudo-diagonal with nonnegative diagonal elements:

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}$$

with $\Sigma_1 = diag(\sigma_i)$, $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ and r is the algebraic rank of the matrix A.
The elements σ_i are the singular values of the matrix A, the columns of U (V) are the left (right) singular vectors. They generate an orthonormal basis for the columnspace (rowspace) of the matrix A. The singular value decomposition is the most reliable way to estimate the rank of a matrix. It is very robust against perturbations and numerically reliable algorithms and efficient software are nowadays available in most software packages. The singular values can be considered as a kind of 'continuous rank measures': The smallest singular value measures in Frobenius norm the distance of the matrix from a matrix of lower rank .

4 The identification algorithm for noise-free data.

In this section, 2 versions of the SVD based identification approach are derived (section 4.2 and 4.3). They are both obtained starting from an important property of the so-called observability space of the system (section 4.1). For the time being, it is assumed that the input-output data are noisefree. However, in section 5, the effect of noise corrupting the input-output data will be investigated. An important difference in interpretation between the first and the second version of the identification approach will be studied there.

4.1 The shift structure of the observability space.

As will become clear furtheron, the observability matrix $\Gamma(i)$ or rather its columnspace, will play an important role. The space spanned by its columns will be called the observability space. The following theorem shows that this space only depends upon the observable poles of the system .First we introduce some notation:
If U is a $m \times n$ matrix, $\underline{U}$ ($\overline{U}$) will be a matrix with a reduced number of rows, obtained from U by omitting the last (first) blockrow. The precise number of rows to be omitted will be obvious from the context.

Theorem 1 *If a $p \times q$ matrix Y is of the form $Y = \Gamma(i).M$ with $\Gamma(i)$ the extended observability matrix and if $\text{rank}(Y) = \text{rank}(\Gamma(i)) = n < p$ and $n \leq q$ and if the columns the $p \times n$ matrix U form a basis for the columnspace of Y, then the matrix T that satisfies $\underline{U}.T = \overline{U}$ has the same eigenvalues as the matrix A in $\Gamma(i)$.*

Proof: Straightforward but omitted.

Corollary 1: Let the SVD of Y be : $Y_{p \times q} = P.S.Q^t$ with $\text{rank}(Y) = \text{rank}(\Gamma(i)) = n < p$ and $n \leq q$. Then P_1, the first $l \times n$ block of P and the $n \times n$ matrix T satisfying $\underline{P}.T = \overline{P}$ are matrices similar to C and A in $\Gamma(i)$, i.e. there exists a nonsingular $n \times n$ matrix R such that $P_1 = C.R$ and $T = R^{-1}.A.R$.
This means that matrices similar to A and C can be found if a basis for the columnspace of $\Gamma(i)$ is known. It will now be studied how such a basis can be found using input-output data only.

4.2 Identification algorithm : version 1

It is not difficult to see from the input - output matrix equation that the projection of the row space of Y_h upon the orthogonal complement of the row space of U_h is under general conditions a space that posesses the shift property of the observability matrix. This observation is exploited in the computation of the matrices A and C , starting from the input-output matrix equation (3):
— Choose the number of columns j of Y_h and U_h larger than $\max(mi, li)$. i must be an overestimation of i_0, the observability index of the system to be identified, while also $(j - mi) > n$
— Denote by $U_h^\perp$ the orthonormal matrix whose columnspace is the orthogonal rowcomplement of U_h i.e. $U_h.U_h^\perp = 0$. — Postmultiply the second input - output relation with $U_h^\perp.(U_h^\perp)^t$

$$Y_h.U_h^\perp.(U_h^\perp)^t = \Gamma(i).X(k,j).U_h^\perp.(U_h^\perp)^t \qquad (4)$$

This is of course equivalent with orthonormalizing (e.g. with some Gram-Schmidt procedure) the rows of Y_h against the rows of U_h.
— It is not difficult to see that the rank of the matrix $Y_h.U_h^\perp.(U_h^\perp)^t$ will be equal to the dimension of the projection of the observable part of the state sequence contained in $X(k,j)$ into the columnspace of $U_h^\perp$. This dimension is generically equal to the dimension of the observable part of the system. The precise condition for this result to be true is discussed in section 5.
— The rank n of $Y_h.U_h^\perp.(U_h^\perp)^t$ can be computed from the singular value decomposition . Let this SVD be :

$$Y_h.U_h^\perp.(U_h^\perp)^t = [P_1\ P_2].\begin{bmatrix} S_1 & 0 \\ 0 & 0 \end{bmatrix}.\begin{bmatrix} Q_1^t \\ Q_2^t \end{bmatrix} \qquad (5)$$

where P_1 is $li \times n$, P_2 $li \times (li - n)$, S_1 $n \times n$, Q_1 $j \times n$ and Q_2 $j \times (j - n)$.
It now follows from the corollary in section 4.1. that the matrices A_t satisfying $\underline{P_1}.A_t = \overline{P_1}$ and C_t = first $l \times n$ blockrow of P_1 are a realization of the matrices A and C of the model.
The computation of the matrices B and D is straightforward though less elegant :
Using the matrix P_2 defined in the SVD (5) and the pseudo-inverse U^+ of U_h, it can be found from the second input. -outputrelation that

$$P_2^t.Y_h.U^+ = P_2^t.H_{tl}$$

Define $K = P_2^t.H_{tl}$, then the matrix K can be partitioned in i blocks of dimension $(li - n) \times m$, satisfying

$$[K_1 \ldots K_i] = [P_1 \ldots P_i].\begin{bmatrix} D & 0 & 0 & \ldots & 0 \\ CB & D & 0 & \ldots & 0 \\ CAB & CB & D & \ldots & 0 \\ CA^2B & CAB & CB & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ CA^{i-2}B & CA^{i-3}B & CA^{i-4}B & \ldots & D \end{bmatrix}$$

where P_k are the $(li - n) \times (li)$ subblocks of P_2^t. This can be rewritten as an overdetermined set of linear equations in the unknowns B and D:

$$\begin{bmatrix} K_1 \\ K_2 \\ K_3 \\ \vdots \\ K_i \end{bmatrix} = \begin{bmatrix} P_1 & P_2 & P_3 & \ldots & P_i \\ P_2 & P_3 & P_4 & \ldots & 0 \\ P_3 & P_4 & P_5 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ P_i & 0 & 0 & \ldots & 0 \end{bmatrix}.\begin{bmatrix} I & 0 \\ 0 & \Gamma(i-1) \end{bmatrix}.\begin{bmatrix} D \\ B \end{bmatrix}$$

This can be solved for the unknown matrices B and D.
Although the computational requirements of the above identification scheme seem formidable, in reality, the computational complexity is moderate. This issue is discussed in section 5.3.

4.3 Identification algorithm : version 2

While the identification approach derived in section 4.2. essentially makes computations on the input block Hankel matrix $U_h(k,i,j)$, the identification algorithm derived in this section will operate on the concatenated matrix $\begin{bmatrix} Y_h(k,i,j) \\ U_h(k,i,j) \end{bmatrix}$
Here U_h (Y_h) is a $mi \times j$ block Hankel matrix where m (l) is the number of system inputs (outputs) . It is assumed from now on that $j > (m + l).i$

Theorem 2 *If the matrices U_h ($mi \times j$) and Y_h ($li \times j$) with $j > \max(mi, li)$ contain input and output vectors of the system , then the concatenated matrix $\begin{bmatrix} Y_h \\ U_h \end{bmatrix}$ will satisfy the following properties:*

1/ rank $\begin{bmatrix} Y_h \\ U_h \end{bmatrix} = mi + n$ if the condition for rank-cancellation is satisfied (see section 5.1). n is the order of the observable system part.

2/ Partition the SVD of $\begin{bmatrix} Y_h \\ U_h \end{bmatrix}$ as :

$$\begin{bmatrix} Y_h \\ U_h \end{bmatrix} = [P_1\ P_2].\begin{bmatrix} S_1 & 0 \\ 0 & 0 \end{bmatrix}.\begin{bmatrix} Q_1^t \\ Q_2^t \end{bmatrix}$$

where

$$[P_1\ P_2] = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$$

The dimensions are : S_1 $(mi + n) \times (mi + n)$; P_{11} $(li) \times (mi + n)$; P_{12} $(li) \times (li - n)$; P_{21} $(mi) \times (mi + n)$; P_{22} $(li) \times (li - n)$; Q_1 $(mi + n) \times j$; Q_2 $(j - mi - n) \times j$. Let $P_{21}^\perp$ be such that $P_{21}.P_{21}^\perp = 0$ with $P_{21}^\perp$ a $(mi - n) \times n$ matrix then:

$$\begin{bmatrix} P_{11}.P_{21}^\perp & P_{12}.T \\ 0 & P_{22}.T \end{bmatrix} = \begin{bmatrix} \Gamma(i) & \Gamma(i)^\perp \\ 0 & -H_{tl}^t.\Gamma(i)^\perp \end{bmatrix}$$

where T is a $(li - n) \times (li - n)$ arbitrary non-singular matrix and $\Gamma(i)^\perp$ contains $(li - n)$ linear independent column vectors of the orthogonal complement of the columnspace of $\Gamma(i)$

Proof: Omitted. The proof essentially follows from the input- output matrix equation (3), from the properties of the singular value decomposition and theorem 1.

Theorem 2 is a key tool in the identification of a state space model A, B, C, D from input-output data:
For the estimation of the order n and the computation of the matrices A, B, C, D , all we need is the left singular vectors of the concatenated matrix $\begin{bmatrix} Y_h \\ U_h \end{bmatrix}$, which is a $(m + l)i \times j$ matrix with $j > (m + l)i$ The singular values allow to estimate the observable order of the system.
If the left singular matrix of $\begin{bmatrix} Y_h \\ U_h \end{bmatrix}$ is partitioned as:

$$\begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$$

as in theorem 2
then A and C can be computed from the shiftstructure of $P_{11}.P_{21}^\perp$ B and D follow from a similar observation as was used in section 4.2.:
— Partition $P_{12}^t = (\Gamma(i)^\perp)^t$ as $[K_1 \ldots K_i]$ where the K_k are $(li - n) \times (li)$ blocks.
— Partition $-P_{22} = (\Gamma(i)^\perp)^t.H_{tl}$ as $[Z_1 \ldots Z_i]$ where the blocks Z_k are each $(li - n) \times m$ matrices
then, a set of equations in the unknown matrices B and D can be written as:

$$\begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ \ldots \\ Z_i \end{bmatrix} = \begin{bmatrix} K_1 & K_2 & K_3 & \ldots & K_i \\ K_2 & K_3 & K_4 & \ldots & 0 \\ K_3 & K_4 & K_5 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ K_i & 0 & 0 & \ldots & 0 \end{bmatrix}.\begin{bmatrix} I & 0 \\ 0 & \Gamma(i-1) \end{bmatrix}.\begin{bmatrix} D \\ B \end{bmatrix}$$

The computational requirements for this second version of the identification algorithm are even less than for the first one (section 5.3)

5 Properties of the new identification approach.

As will be explained in this section, the optimal choice for the matrix block dimensions i and j to be used in the construction of the block Hankel matrices $U_h(k,i,j)$ and $Y_h(k,i,j)$ is essential.
For the identification version 1 (section 4.2.) the following inequalities should be satisfied:
— $i > i_0$, where i_0 is the system observability index
— $j > \max(mi, li)$ (it will be assumed that $\text{rank}(U_h) = mi$, which can be interpreted as a persistant excitation condition for the input-sequence (but this will not be further pursued here)) — $(j - mi) > n$: $(j - mi)$ is the dimension of the orthogonal row complement of U_h.
— $li > n$: li is the number of rows of the matrix $Y_h.U_h^\perp.(U_h^\perp)^t$

For the identification version 2 (section 4.3.) the following inequalities should be satisfied:
— $i > i_0$, where i_0 is the system observability index.
— $j > (m + l).i$

In both cases, the conditions are such that i is sufficiently large, in order to allow enough 'space' to estimate the observable dynamical order n. Once i is fixed, j should be sufficiently large for orthogonal row complements to exist.
From hereon, it is assumed that these weak requirements for the dimensions i and j are satisfied. There are however strong additional reasons to require that j and i be chosen such that the block Hankel matrices $Y_h(k,i,j)$ and $U_h(k,i,j)$ are largely overdetermined, i.e. have much more columns than rows.
In this paper, we discuss the three most important without mentioning too much detail : the so-called condition of 'rank-cancellation', discussed in section 5.1., the robustness of the identification scheme with noisy observations n section 5.2. and the computational efficiency, which is discussed in section 5.3.

5.1 The condition of rank-cancellation.

The succes of the order estimation procedure via the singular value decomposition in both versions of the SVD identification approach, depends upon the geometrical situation of the row space of the matrix $\Gamma(i).X(k,j)$ with respect to the row space of the input block Hankel matrix $U_h(k,i,j)$, or better with respect to the ortogonal complement of the row space of U_h.
The dimension of the projection of the row space of $\Gamma(i).X(k,j)$ on the orthogonal complement $U_h^\perp$ of the row space of the matrix U_h, should be equal to the correct observable dimension n. Ideally, the row space of $\Gamma(i).X(k,j)$ (or that of $X(k,j)$) should be as orthogonal as possible to the row space of $U_h(k,i,j)$ Loosely speaking, no dimension may be 'lost' in the projection. The phenomenon where one or more dimensions of the observable part of the state space are lost in the projection on the orthogonal complement of the row space of U_h will be called 'rank-cancellation'.
The precise geometrical situation can rigorously be described via the concept of principal angles between subspaces, which is a generalization of the angle between two vectors [3]. Interesting enough, the computation of these principal angles also makes use of the singular value decomposition. The larger the principal angles between $\text{span}_{\text{row}}(X(k,j))$ and $\text{span}_{\text{row}}(U_h(k,i,j))$, the more information is contained in the dynamical behavior of the state sequence $X(k,j) = [x[k] \ldots x[k + j - 1]]$, which was not yet present in the row space of the input Hankel matrix U_h. For single input, single output systems, excited by an impulse, it can be demonstrated that these principal angles are always orthogonal. This corresponds precisely to the intuitive idea of an impulse as being the optimal identification excitation signal. The main conclusion is that both mathematically and via extensive simulations, it can be proven that rank-cancellation is not generic. Moreover, it can be influenced by choosing j sufficiently large. Using a qualitative statement, one can say that for a fixed input sequence, generically, the probability that rank cancellation occurs, decreases ,for fixed i, with increasing j. Since in a lot of identification experiments, the input sequence cannot be freely chosen, this result is important.
A critical example is now considered. In figure 1. the largest principal angle between the orthogonal complement $U_h^\perp$ of $\text{span}_{\text{row}}(U_h(k,i,j))$ and $\text{span}_{\text{row}}(X(k,j))$ is depicted for a SISO second order system as a function of the number of columns j and for a fixed number of blockrows $i = 5$. The smallest canonical angle is approximately constant (10^0 and is not shown).

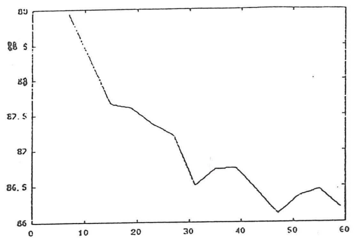

Figure 1

Observe that the largest canonical angle decreases with increasing overdetermination j/i. Extensive simulations on several systems have shown that this behavior is generic. It constitutes one of the reasons to choose j/i large.

5.2 Overdetermination j/i decreases noise sensitivity.

Till now, it was assumed that the available input-output data are noise-free. In most practical circumstances however, measurement noise corrupts the data. It is assumed in this paper that this can be modeled as unobservable additive perturbations of the 'exact' data. Denoting by ˆ exact, by ˜ pure noise quantities, the measured input-output data are stored in their respective block Hankel matrices that can be written as:

$$U_h = U_h(k,i,j) = \widehat{U_h}(k,i,j) + \widetilde{U_h}(k,i,j) = \widehat{U_h} + \widetilde{U_h}$$

$$Y_h = Y_h(k,i,j) = \widehat{Y_h}(k,i,j) + \widetilde{Y_h}(k,i,j) = \widehat{Y_h} + \widetilde{Y_h}$$

Three effects for this noisy situation can be observed as a function of the overdetermination j/i:
— 1. It can rigorously be proven [14], that , the probability that the pure noise row space $\text{span}_{\text{row}}(\widetilde{U_h})$ is orthogonal to the exact data row space $\text{span}_{\text{row}}(\widehat{U_h})$, increases with increasing overdetermination j/i. (This is by the way the same reason why in the solution of inconsistent linear equations a large overdetermination is advocated, be it in least squares or in total linear least squares approaches [2] [13], because the orthogonality of noise and exact spaces then allows to separate them).
— 2. For increasing overdetermination j/i, the singular values that are used to determine the order of the observable part of the system, grow towards certain 'asymptotic' levels, which allows to distinguish clearly the system singular values from those caused by only the noise. A full explanation and mathematical demonstration of this observation is now available [14]. As an illustration, consider the following simple example: In fig.2. the smallest canonical angles between $\text{span}_{\text{row}}(Y_h)$ and $\text{span}_{\text{row}}(U_h)$ are shown as a function of the number of columns j for a second order SISO system, where the output measurements are corrupted by 1% noise. The number of blockrows $i = 5$. The largest singular value is not shown. In the noiseless case (under the conditions described in 5.1), only 2 canonical angles would differ from 0. Observe that the three canonical angles caused by the 'pure' noise reach a 'saturation' level, while those coming from the 'true' system keep increasing. Hence it becomes more easy to determine the true order of the system for increasing overdetermination j/i.

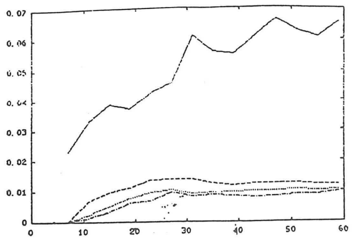

Figure 2

— 3. Strong consistency results can be proven that allow to conclude that in both identification algorithms (version 1 and 2), the correct columnspace of the extended observability matrix $\Gamma(i)$ is identified with increasing probability for increasing overdetermination j/i . By theorem 1, the shiftstructure of this columnspace allows a realization of the model matrices A and C.

In the case where the data are corrupted by additive noise, one can now make a meaningful distinction between the two identification versions (section 4.2 versus section 4.3).
If the input data are noise free, but the output data are considered to be noisy, the first identification approach, which projects the row space of the output Hankel matrix upon the orthogonal complement of the row space of the input Hankel matrix should be preferred. This correspond to a linear least squares projection interpretation of the identification [2].
When both input and output measurements are noisy (corrupted by the same absolute amount of noise), the second version of section 4.3. should be used. This corresponds to a total linear least squares interpretation of identification [2].

5.3 The computational requirements.

Although computational details will be reported elsewhere, we briefly summarize in this paper some important observations.

Identification approach version 1 (section 4.2)
— The orthogonalization of the row space of the block Hankel matrix Y_h with respect to that of the block Hankel matrix U_h requires a Gram-Schmidt orthogonalization procedure. Preferably, the Hankel structure could be exploited.
— The resulting matrix $Y_h.U_h^{\perp}.U_h^{\perp t}$ is (for reasons to be discussed in section 5) a largely overdetermined matrix, with much more columns than rows. Hence, its singular value decomposition can be computed very efficiently by first performing a $R.Q$ decomposition (R lower triangular, Q orthonormal) and then computing the SVD of the matrix R
Identification approach version 2 (section 4.3)
— The singular value decomposition of a largely overdetermined concatenated matrix $\begin{bmatrix} Y_h \\ U_h \end{bmatrix}$ is required. This can again be achieved by first computing the $R.Q$ factorization, followed by the SVD of R.
— An adaptive version of the presented identification algorithms, that may be used for the identification of time-varying linear systems, is actually being implemented and tested. It exploits both the Hankel structure and a rank one updating mechanism of the SVD via the $R.Q$ factorization. Preliminary results are promising.

6 Some real life examples.

The performance of the algorithms has been evaluated on both simulated and industrial data sets. In this section, an example is presented that is adopted from [6] [7] [8].
The model has been identified from sequences regarding normal operating conditions of an ethane - ethylene distillation column. White Gaussian noise with standard deviation equal to 10 % of the standard deviation observed on the single (exact) input-output components was added before identification was carried out. A 6-th order model was selected from the singular spectrum of $Y_h.U_h^{\perp}.U_h^{\perp t}$. The identification was carried out using the second version, presented in section 4.2. The identified model was then validated by comparing original and simulated outputs, using the exact original inputs. Fig.3 shows the exact inputs while fig.4 shows both the exact and simulated outputs.

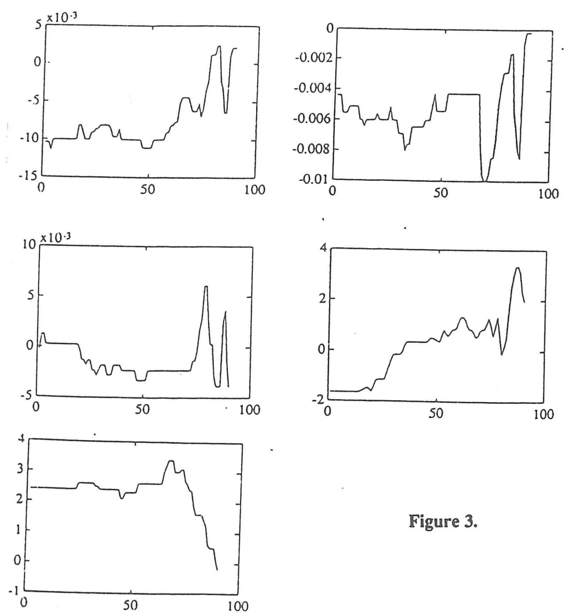

Figure 3.

7 Conclusions

In this paper, a survey was given of geometrical concepts for a new identification strategy. The properties of the singular value decomposition are exploited to compute a state space model from noisy input -output observations. Two versions have been derived: One which allows for a linear least squares interpretation, a second which has more the character of total linear least squares. Future work will be directed to a complete geometrical treatment of the identification scheme and to exploit efficient numerical implementation of adaptive versions of the singular value decomposition for structured matrices.

8 Acknowledgement.

The authors would like to thank Prof. R. Guidorzi from Bologna University for the input-output measurements from industrial plants, used in section 6 of this paper.

References

[1] Budin M.A. Minimal realization of discrete linear systems input-output observations. IEEE Trans. on Automatic Control, Vol. AC-16, no.5, october, p.395-401, 1971.

[2] De Moor B., Vandewalle J. A unifying theorem for linear and total linear least squares identification techniques. Submitted.

[3] Golub. Matrix computations. North Oxford Academic Publishing Co., John Hopkins University Press, 1983.

[4] Gopinath B. On the identification of linear time invariant systems from input-outputdata. The Bell System Technical Journal, Vol.48, May-June, no.5.,p.1101-1113, 1969.

[5] Guidorzi R.P. Invariants and canonical forms for systems structural and parametric identification. Automatica, Vol.17, no.1., p.117-133, 1981.

[6] R.P. Guidorzi, M.P. Losito and T.Muratori. The range error test in the structural identification of linear multivariable systems. IEEE Trans. on Automatica Control, Vol.AC-27, no.5, October 1982.

[7] R.P.Guidorzi, M.P.Losito and T.Muratori. On the last eigenvalue test in the structural identification of linear multivariable systems. Proc. of the 5th European Meeting on Cybernetics and Systems Research, Vienna, April 1980.

[8] R.P.Guidorzi, P.Rossi. Identification of a power plant from normal operating records. Automatic Control Theory and Applications, Vol.2., no.3., September 1974.

[9] Ho B.L., Kalman R.E. Effective construction of linear state variable models from input output functions. Regelungstechnik, Vol.14,pp545-548, 1966.

[10] Kung S.Y. A new identification and model reduction algorithm via singular value decomposition. Proc. of the 12-th Asilomar Conf. Circ. Syst. Comp., Nov.,1978.

[11] Liu R., Suen L.C. Minimal dimension realization and identifiability of input-outputsequences. IEEE Trans.on Automatic Control, Vol.AC-,no. ,april p.227-232, 1977.

[12] Ljung L. and T.Soderstrom. Theory and practice of recursive identification. MIT Press, Cambridge, Nassachusetts, 1983.

[13] Van Huffel S. Analysis of the Total Least Sqaures problem and its use in parameter estimation. Doctoral Thesis, ESAT, Dept. of Electrical Engineering, Katholieke Universiteit Leuven, 1987.

[14] Van Mieghem Piet, Vandenberghe Lieven. Een nieuw algoritme voor de identifikatie van lineaire multivariable systemen. Master Thesis (in Dutch), ESAT, Electrical Engineering Department, Katholieke Universiteit Leuven ,1987.

[15] Zeiger H.P., McEwen A.J. Approximate linear realizations of given dimension via Ho's algorithm.IEEE Trans. on Automatic Control, Vol.AC-19, p153, 1974.

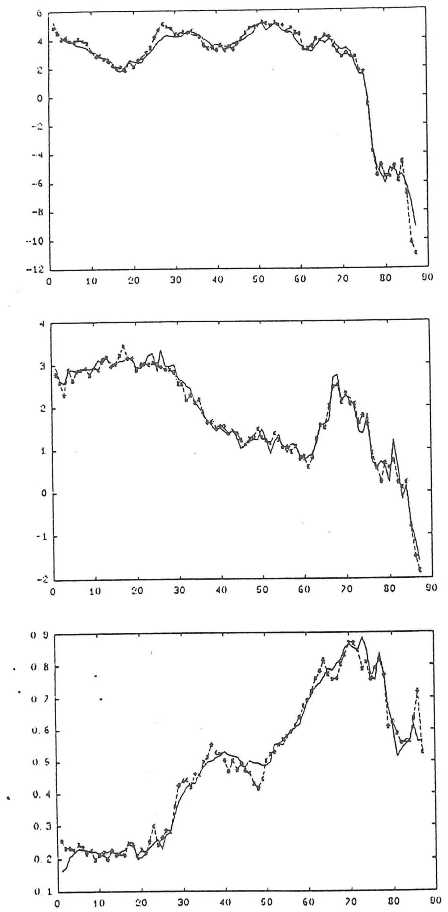

Figure 4.