Vandewalle J., De Moor B. *A variety of applications of singular value decomposition in identification and signal processing : A survey of the SVD research activities at the Katholieke Universiteit Leuven* in 'SVD and Signal Processing: Algorithms, Applications and Architectures.' E.F.Deprettere, North Holland, Elseviers Science Publishers B.V., 1988, pp.43-91.

A VARIETY OF APPLICATIONS OF SINGULAR VALUE DECOMPOSITION IN IDENTIFICATION AND SIGNAL PROCESSING

Joos Vandewalle and Bart De Moor
ESAT (Electronics, Systems, Automation and Technology)
Katholieke Universiteit Leuven
Kardinaal Mercierlaan 94
3030 Heverlee, Belgium
Tel: 32/16/220931 Telex 25941 elekul Fax :32/16/221855

Abstract

An overview is given of the research activities at the ESAT laboratory of the Katholieke Universiteit Leuven in Belgium, concerning (generalized) singular value decomposition and its applications. Besides original theoretical results and concepts, several examples of mathematical engineering applications are presented.

Keywords: generalized singular value decomposition, factor analysis, canonical correlations analysis, total linear least squares, realization and identification of state space models, subspace methods

Contents :

1 Introduction

In recent years, the (generalized) singular value decomposition has become an extremely valuable instrument in the analysis and the solution of problems in mathematical engineering. In most applications, the SVD provides a unifying framework, in which the conceptual formulation of the problem, the practical application and an explicit solution that is guaranteed to be numerically robust, are derived at once. In this way, the SVD has become a fundamental tool for the formulation and derivation of new concepts such as angles between subspaces, oriented signal-to-signal ratio's , canonical correlation analysis,... and for the reliable computation of the solutions to problems such as total linear least squares, realization and identification of linear state space models, source separation by subspace methods etc.. In this article, a survey will be presented containing the main results of still on-going research activities at the ESAT - laboratory of the Katholieke Universiteit Leuven during the past decennium. Several colleagues have contributed to the results reported in this paper. Besides the authors, let us mention Dr.Ir.Jan Staar, Dr.Ir.Sabine Van Huffel, Ir. Jan Vanderschoot, Ir.Dirk Callaerts, Ir. Marc Moonen, Ir. Jan Swevers, Ir. Lieven Vandenberghe and Ir. Piet Van Mieghem . References to the work of each can be found in a chronological list at the end.

In a first part of this article, several fundamental concepts are defined and their properties discussed in terms of the (generalized) singular value decomposition : Besides the technique of Total Linear Least Squares, a fundamental framework is summarized which allows to formalize 'factor-analysis-like' problems in terms of oriented signal-to-signal ratios. The computational tool is the generalized SVD which is also the appropriate instrument to quantify the concept of canonical correlation analysis and the notion of angles between subspaces. In this article, well known properties of the (G)SVD will only briefly be summarized, were it only for notational convenience. However, we present also a survey of several useful original insights in the singular value decomposition structure with respect to backward error analysis, condition numbers, sensitivity and the influence of noise.

In a second part of the survey, a wide variety of applications from a broad spectrum of scientific disciplines will be cited and illustrated : mechanics (Moments of inertia) , electrical network analysis (the conditioning of reference node choice), bio-medical engineering (signal source separation of maternal and fetal ECG), realization of systems from impulse response measurements, identification of industrial processes, etc....

Only the main concepts and insights will be emphasized throughout without paying attention to details, for which the interested reader is referred to a chronological list of references at the end.

Notations and abbreviations

R^m vector space of real m-tuples.

$A_{m \times n}$ matrix A with m rows and n columns

A^t matrix transpose

$\bar{A}$ matrix A with first row omitted

$\underline{A}$ matrix A with last row omitted

$\bar{A}^t$ first omit, then transpose

$A = diag(\alpha_1, \ldots, \alpha_p)$ If A is a $m \times n$ matrix and $p = min(m, n)$ then $A = (a_{ij})$ is diagonal and $a_{ii} = \alpha_i$ for $i = 1, ..., p$

$R(A)$ Range of A : $R(A) = \{y \mid y = A.x \text{ for some } x \}$

$N(A)$ Null space of A : $N(A) = \{x \mid A.x = 0\}$

$r(A)$ rank of A; $r(A) = dim[R(A)]$

$\text{span}\{v_1, \ldots, v_r\}$ vector space generated by linear combinations of the vectors $v_i, i = 1, .., r$

$\| A \|_F^2$ Frobeniusnorm of $A = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2$ where A is an $m \times n$ real matrix

$\| A \|_2$ 2-norm of $A = \max_{\|x\|=1} \| Ax \|$ where x is an $n \times 1$ real vector

$\| x \|_2^2 = \sum_{i=1}^{n} x_i^2$ where x is an $n \times 1$ real vector

$A \otimes B$ Kronecker product (section 5.6)

$A \square B$ Khatri-Rao product (section 5.6)

SVD singular value decomposition

GSVD generalized singular value decomposition

TLLS total linear least squares

LLS linear least squares

All numbers, vectors and matrices are real unless explicitly stated otherwise. Column vectors are denoted by (possibly indexed) small letters while row vectors are denoted as the transpose of a column vector. An $m \times n$ matrix U is called orthonormal ($m \geq n$) if $U^t . U = I_n$.

2 The (Generalized) Singular Value Decomposition

In this section, the theorems stating the existence of the singular value decomposition are presented for notational convenience only. For a proof and computational requirements, the reader is referred to literature [53]. Moreover some well known classical results that constitute part of the fundaments of the applications to be reported, are summarized.

2.1 The Singular Value Decomposition and its properties

Theorem 1 *The singular value decomposition for real matrices.*
If A is a $m \times n$ real matrix, then there exist real orthogonal matrices

$$U = [u_1 \;\; u_2 \;\; \cdots \;\; u_m]$$

and

$$V = [v_1 \;\; v_2 \;\; \cdots \;\; v_n]$$

such that

$$U^t \cdot A \cdot V = diag(\sigma_1, \sigma_2, \cdots, \sigma_p)$$

where

$$p = min(m, n)$$
$$\sigma_i \geq \sigma_{i+1} \geq 0 \;\; i = 1, .., p$$

The σ_i are the singular values of A and the vectors u_i and v_i are respectively the $i-th$ left and the $i-th$ right singular vector.

The set $\{u_i, \sigma_i, v_i\}$ is called the $i-th$ singular triplet. The singular vectors (triplets) corresponding to large (small) singular values are called large (small) singular vectors (triplets).

The SVD reveals a great deal about the structure of a matrix as evidenced by the following well known corollaries :

Corollary 1 *Let the SVD of A be given as in theorem 1 and*

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_p = 0$$

then

(1). *Rank property*

$$r(A) = r \;\; and \;\; \begin{array}{l} N(A) = span\{v_{r+1}, \cdots, v_n\} \\ R(A) = span\{u_1, \cdots, u_r\} \end{array}$$

(2). *Dyadic decomposition*

$$A = \sum_{i=1}^{r} u_i \cdot \sigma_i \cdot v_i^t$$

(3). *Norms*

$$\| A \|_F^2 = \sigma_1^2 + \cdots + \sigma_r^2$$
$$\| A \|_2 = \sigma_1$$

The rank property is one of the most valuable aspects of the SVD. The singular values can be considered as quantitative measures of the qualitative notion of rank : Algebraically, a matrix has a well-determined rank that is a nonnegative integer. However, in practice, the effects of rounding errors and fuzzy noisy data make numerical rank determination a non-trivial exercise. The dyadic decomposition provides a canonical description of a matrix as a sum of $r(A)$ rank-one matrices of decreasing importance, as measured by the singular values. The dyadic decomposition is in a lot of applications a rationale for data reduction. The three preceding results combine in the following (by now classical) quantification of near rank deficiency of a matrix, for a proof of which the reader is referred to [53] :

Corollary 2 *Let the SVD of the $m \times n$ real matrix A be as in theorem 1 with $r = r(A) \leq p = min(m,n)$ and define :*

$$A_k = \sum_{i=1}^{k} u_i \cdot \sigma_i \cdot v_i^t \quad with \quad k < r$$

then

$$\begin{array}{ll} min & \| A - B \|_2 = \| A - A_k \|_2 = \sigma_{k+1} \\ r(B) = k & \\ min & \| A - B \|_F^2 = \| A - A_k \|_F^2 = \sigma_{k+1}^2 + \cdots + \sigma_p^2 \\ r(B) = k & \end{array}$$

This important result is the basis of a lot of concepts and applications such as total linear least squares, data reduction, image enhancement, dynamical system realization theory and in all possible problems where the heart of the solution is the approximation, measured in 2-norm or Frobeniusnorm, of a matrix by one of a lower rank. Many more valuable properties of the SVD, including existence proofs, computational requirements and numerical considerations, sensitivity results, conditioning etc.. can be found in the modern bible of numerical analysis [53] and the references therein. We also refer to section 5 for some additional useful insights into the SVD structure.

2.2 The Generalized Singular Value Decomposition

Theorem 2 *The generalized singular value decomposition for real matrix pairs.*

If A is a $m \times n$ matrix with $n \geq m$ and B is a $m \times p$ matrix, then there exist orthogonal matrices U $(n \times n)$ and V $(p \times p)$ and an invertible X $(m \times m)$ such that :

$$X \cdot A \cdot U = D_A = diag(\alpha_i) \;\; \alpha_i \geq 0 \;\; i = 1,..,m$$

and

$$X \cdot B \cdot V = D_B = diag(\beta_i) \;\; \beta_i \geq 0 \;\; i = 1,..,q = min(m,p)$$

where

$$\beta_1 \geq \beta_2 \geq \cdots \geq \beta_r \geq \beta_{r+1} = \cdots = \beta_q = 0 \quad r = rank(B)$$

For an existence proof that is at the same time constructive, we refer to [53,p.319]. Obviously, the generalized singular value decomposition (GSVD) is for a pair of matrices what the SVD is for one matrix. Remark that the GSVD reduces to the SVD in the case that $B = I_m$. The elements of the set

$$\sigma(A,B) = \{\alpha_1/\beta_1, ..., \alpha_r/\beta_r\}$$

are referred to as the generalized singular values of A and B . In some applications, it is necessary to order the GSVD in a different way, namely in non-increasing order of the generalized singular values. The generalized singular values corresponding to the $\beta_i = 0$ are infinite. They are considered to be equal and come first. This alternative makes a lot of sense in the oriented signal-to-signal framework of section 4.1.

For more details concerning the GSVD, the reader is referred to [53] [54] and the references therein.

3 SVD and GSVD algorithm

Since there exists an intimate relationship between the SVD and the symmetric eigenvalue problem, at first sight, eigenvalue algorithms could be used in order to compute the SVD. However, the explicit squaring which is the first step when converting the SVD computation into a symmetric eigenvalue problem , can cause a considerable loss of numerical precision, (worst case half of the machine precision) especially when the matrices involved are ill-conditioned. An engineering example of this effect, which can have serious implications, is provided in section 6.1. The first fully reliable SVD routine to circumvent this numerical caveat , was the result of the pioneering work in the sixties of Gene Golub at Stanford University, who applied orthogonal transformations in a by now standard SVD algorithm, readily available in most numerical software packages. Other possible approaches include the Lanczos method, especially for large sparse matrices, all of which is described in the modern bible of numerically robust linear algebra [53] and the references therein. In recent years, quite some effort has been spent to one-sided and two sided Jacobi methods, because they are serious candidates for parallel implementation . Recently, there has been an increasing interest to the computation of the generalized singular value decomposition. As an analogy with the SVD, there is an intimate relation between the Generalized Singular Value Decomposition and the generalized symmetric eigenvalue problem. Numerically reliable software will avoid the explicit squaring of the matrices in converting the GSVD problem into a symmetric generalized eigenvector problem. As an example, consider the following small GSVD problem [51]

$$A = \begin{bmatrix} 1 & \mu & 0 \\ 1 & 0 & \mu \end{bmatrix} \quad B = \begin{bmatrix} \epsilon & 1 & -1 \\ \epsilon & 1 & 1 \end{bmatrix}$$

with $\mu \geq \epsilon_m \geq \mu^2$ and $\epsilon \geq \epsilon_m \geq \epsilon^2$ where ϵ_m is the machine precision. By some straightforward calculations, one can show that the generalized singular values of the matrix pair [A , B] are given by :

$$\begin{array}{ll} \alpha_1 = (\frac{2+\mu^2}{4+\mu^2+2\epsilon^2})^{1/2} & \beta_1 = (\frac{2(1+\epsilon^2)}{4+\mu^2+2\epsilon^2})^{1/2} \\ \alpha_2 = (\frac{\mu^2}{2+\mu^2})^{1/2} & \beta_2 = (\frac{2}{2+\mu^2})^{1/2} \end{array}$$

However, when the Grammian products $A.A^t$ and $B.B^t$ are explicitely computed as a first step in the computation of the GSVD, already a lot of information is lost, and the result will even depend upon the ordering of the terms in the computation of the inner product:

$$A \cdot A^t = \begin{bmatrix} 1+\mu^2 & 1 \\ 1 & 1+\mu^2 \end{bmatrix} \overset{\approx}{finite\ machine\ precision} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

For the inner products in $B.B^t$, one has to compute the inner product

$$\begin{aligned}[\epsilon \;\; 1 \;\; 1].[\epsilon \;\; 1 \;\; -1]^t &= (\epsilon^2 + 1) - 1 \approx 0 \quad (case1) \\ &= \epsilon^2 + (1 - 1) \approx \epsilon^2 \quad (case2)\end{aligned}$$

Then :

$$case\ 1 : B \cdot B^t \approx \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \quad case\ 2 : B \cdot B^t \approx \begin{bmatrix} 2 & \epsilon^2 \\ \epsilon^2 & 2 \end{bmatrix}$$

This leads to the generalized singular values, both for case 1 and case 2:

$$\begin{aligned}\alpha_1 &= \sqrt{2}/2 \quad \beta_1 = \sqrt{2}/2 \\ \alpha_2 &= 0 \qquad\quad \beta_2 = 1\end{aligned}$$

The use of orthogonal transformations is advocated everywhere as much as possible. However, the implementation of reliable software for GSVD is presently still under investigation although several results have already appeared in literature. [53] [61]

At the ESAT lab in Leuven, several alternatives to the above mentioned algorithms have been investigated. The main motivation is that in a lot of engineering applications, the general purpose SVD algorithm (complete decomposition, full machine precision) could be replaced by other algorithms with less stringent specifications.

- A lot of applications only require the computation of part of the singular spectrum (realization theory and data reduction : dominant singular triplets, source separation : intermediate triplet, total linear least squares, Pisarenko-type spectral estimation : smallest singular triplets). The storage reduction obtained by storing only the dominant triplets instead of the full matrix can be considerable e.g. in image processing despite the relatively heavy computational requirements.
- Moreover, very frequently the available data are noisy (industrial environment typically 10 % !), indicating that a full precision SVD makes no sense.
- A third specification is the adaptive computation of the SVD of matrices that are time-varying, either elementwise or by adding and deleting columns and/or rows.
- Finally, in a lot of applications the matrices are very structured ((block-) Hankel or - Toeplitz, circulant, ...) so that it could be hoped that structure expoiting algorithms perform better (faster) than the standard full size SVD algorithm. Moreover, considerable storage gain can be achieved if the elements of the structured matrix could be stored without redundancy, hence requiring matrix-vector multiplication based algortihms.

Descriptions of the following alternatives can be found in the indicated references:

1/ Results on the power method and the Chebyshev method for the SVD can be found in [15] [21]. One of the important observations (at least from didactical reasons) is that the Chebyshev method is nothing else than the eigenvalue power method applied to a certain matrix, constructed from the original one. This allows to translate all results that were derived for the SVD power method in [15] to the Chebyshev algorithm in [21], including convergence criteria, deflation strategies, accelerations algorithms and convergence level control rules that allow to compute the SVD up to the required numerical precision.

2/ An adaptation of Golub's full size SVD for the case that only the smallest singular triplets are needed, is described in [36] [49]. By a careful analysis the computational

requirements can be reduced in such a way that the PSVD algorithm (partial SVD) can be 3 times faster than the classical one, while obtaining the same accuracy. Moreover, in [49] a comparison can be found between 4 different methods to compute the null space of an approximation of a given matrix in the sense of total linear least squares (section 4.2) . Chebyshev iteration [21], inverse iteration , Rayleigh quotient iteration and the Lanczos method have been compared with respect to their computational efficiency in total linear least squares computations. The conclusion is that inverse iteration is the most promising iterative technique for solving generic TLLS problems of known rank. Moreover, the direct methods (Golub's SVD and PSVD) have also been compared with the iterative ones and several potential applications have been investigated [49] (parameter estimation, subset selection, discrete deconvolution).

3/ Finally, the one-sided Jacobi method has been studied in relation to the problem of adaptive computation of the SVD of a matrix containing measurements on the fetal and the mother ECG [22] [23] [29] [43]. As the measurements (6 to 9 channels) enter at a frequency of 250 Hz, the SVD is updated with orthogonal Given's rotations as to minimize the Frobeniusnorm of the off-diagonal elements. Convergence and speed of a possible implementation are studied . Also the implementation on a TMS-320 signal processor has been tested and has been shown to be feasible for real time applications.

4 Some fundamental geometric concepts based on (G)SVD

In this section, we pay some attention to results that are important from the conceptual as well as from the application point of view. First, we summarize the results on a framework that is suited to study factor-analysis-like (subspace) methods (section 4.1). In section 4.2., we summarize the rationale behind the total linear least squares approach for solving overdetermined sets of linear equations while in section 4.3. the relation between canonical correlation , principal angles between subspaces and the generalized SVD is established.

4.1 Oriented energy and oriented signal to signal ratios

In a wide variety of systems and signal processing applications, vector sequences are measured and analysed. Whenever *linear* models are used to 'explain' the measurements, one is interested in their fundamental characteristics, which are their complexity (the rank of certain matrices) and the parameters describing the model. These parameters can often be extracted from certain subspaces, of which the dimension is a measure for the complexity of the model. Hence in identifying linear models from noisy data, one is confronted with two basic non-trivial problems :

- the meaningful estimation of a rank
- the reliable computation of a corresponding subspace.

It is obvious that the technique of the (generalized) singular value decomposition is very appropriate to describe and compute both ranks and subspaces. Examples can be found in a wide variety of applications : Sets of linear equations (total linear least squares [49] [53] (section 4.2)), the identification of factor-analysis-like models (rotational invariance techniques [63] , separation of MECG/FECG [29] [43] [50] [65] (section 6.3)), realization of linear state space models from impulse responses [1] [3] [17] [40] [48] (section 6.4) and identification of

state space models from noisy input-output measurements via canonical correlation analysis [47] (section 6.5.)

The recent introduction of the fundamental concepts of oriented energy and oriented signal - to -signal ratio [45] [48] has provided a rational framework in which both the estimation of ranks and subspaces can be formalized in a rigorous way.

Let A and B be two $m \times n$ matrices with $n \gg m$, both containing measurement vector sequences (typically n consecutive sample vectors from m measurements channels). The columns of A and B are denoted by $a_k, b_k \;\; k = 1, .., n$

Definition 1 *The oriented energy of the matrix A, measured in a direction q is defined as :*

$$E_q[A] = \sum_{k=1}^{n} (q^t a_k)^2$$

Definition 2 *The oriented signal-to-signal ratio of the two vector sequences A and B in the direction q is defined as :*

$$E_q[A, B] = E_q[A]/E_q[B]$$

There are straightforward generalizations of these definitions to oriented energy and signal-to-signal ratios in subspaces Q^r . In [45] it is shown that the analysis tool for the oriented energy distribution of a matrix A is the singular value decomposition, while the analysis tool of the oriented signal-to-signal ratio of two vector sequences A and B is the generalized singular value decomposition of the matrix pair $[A, B]$. These well understood matrix factorizations allow to characterize the directions of extremal oriented energy and oriented signal-to-signal ratio :

Theorem 3 *Extremal directions of oriented energy .*

Let A be a $m \times n$ matrix ($n \gg m$) with SVD $A = U\Sigma V^t$ where $\Sigma = diag\{\sigma_i\}$. Then each direction of extremal oriented energy is generated by a left singular vector with extremal energy equal to the corresponding singular value squared.

This theorem is illustrated for three dimensions in figure 1. Observe the maximum and minimum corresponding to the largest resp. smallest singular vectors while the saddle point corresponds to the intermediate singular vector.

Theorem 4 *Extremal directions of oriented signal-to-signal ratio.*

Let A and B be $m \times n$ matrices, with GSVD :

$$A = X^{-1}.D_a.U^t \;\; D_a = diag\{\alpha_i\}$$
$$B = X^{-1}.D_b.V^t \;\; D_b = diag\{\beta_i\}$$

where the generalized singular values (possible infinite) are ordered such that $(\alpha_1/\beta_1) \geq (\alpha_2/\beta_2) \geq ... \geq 0$. Then each direction of extremal signal - to -signal ratio is generated by a row of the matrix X and the corresponding extremal signal-to-signal ratio is the generalized singular value squared.

Observe that the extremal directions of oriented energy are orthogonal while this is not necessarily the case for the signal-to-signal ratio. The underlying tool for the proof of these

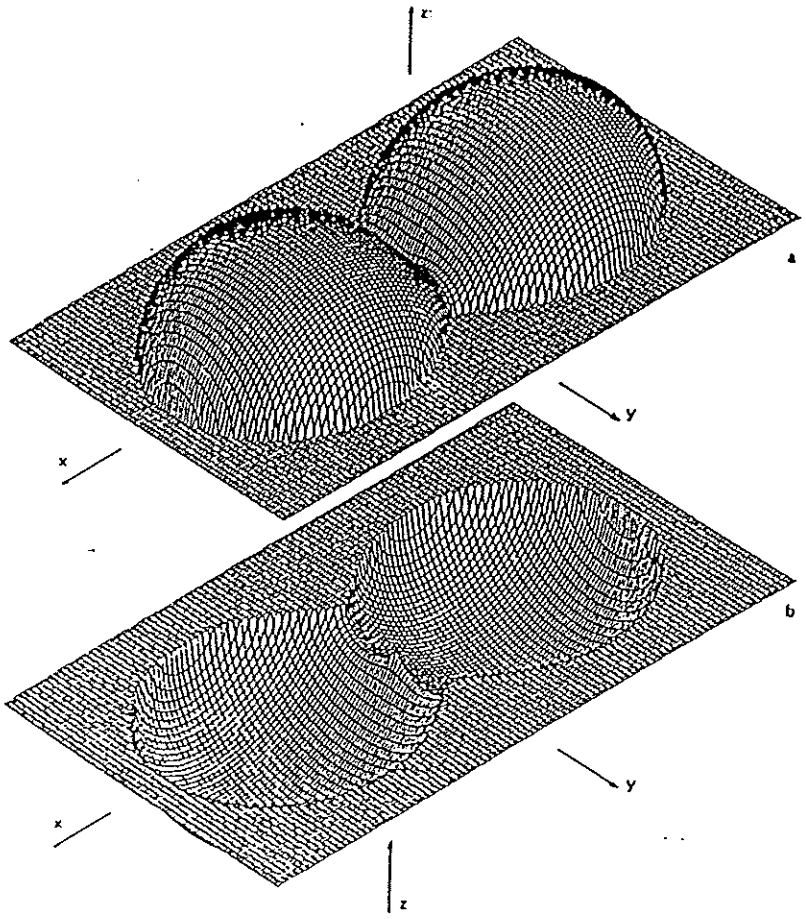

Figure 1: Oriented energy distribution of a 3-vector sequence . Remark that the extreme directions are orthogonal. In 3 dimensions, the extremes are a maximum, a minimum and one saddle-point.

theorems is nothing else than the Courant-Fisher minimax characterization of the eigenvalues of symmetric operators [45] [63].

Now one can proceed by investigating in which directions of the ambient space the vector signal in the matrix A can be best distinguished from the vector signal in the matrix B. This leads to the definition of maximal minimal and minimal maximal signal to signal ratios of two vector sequences [45] [48]

Definition 3 *Maximal minimal and minimal maximal signal-to-signal ratio.*

The maximal minimal signal-to-signal ratio of two m-vector sequences contained in the $m \times n$ matrices A and B over all possible r-dimensional subspaces ($r < m < n$) is defined as:

$$MmR[A, B, r] = \underset{Q^r \subset R^m}{Max} \quad \underset{q \in Q^r}{Min} \quad E_q[A, B]$$

Similarly, the minimal maximal signal-to-signal ratio is defined as:

$$mMR[A, B, r] = \underset{Q^r \subset R^m}{Min} \quad \underset{q \in Q^r}{Max} \quad E_q[A, B]$$

The idea behind these definitions is the following : For a given subspace Q^r of the m-dimensional ambient space ($r < m < n$) there is a certain direction $q \in Q^r$ for which the signal-to-signal ratio of the two vectorsequences A and B is minimal. This direction corresponds to the worst direction q in the sense that in this direction the energy of A is difficult to distinguish from the energy of B. This worst case of course depends upon the precise choice of the subspace Q^r. Among all r-dimensional subspaces, there must exist at least one r-dimensional subspace where the worst case is better that all other worst cases. This subspace is the r-dimensional subspace of maximal minimal signal-to-signal ratio. It comes as no surprise that the GSVD allows to quantify this subspace: It is the r-dimensional subspace generated by the first r rows of X, when the generalized singular values are ordered as in theorem 4.

Hence, the concept of oriented signal-to-signal ratio and the GSVD allow to formalize all model identification approaches, in which

- the determination of a suitable rank r provides the complexity of the model. This can be estimated from the generalized singular values and can be translated in reliable 'automatic' rank estimation procedures, based on a well motivated criterion.
- the model parameters follow from the corresponding subspace of maximal minimal signal-to-signal ratio.

Moreover, it can be shown that when the vector sequence B consists of an unobservable stochastic vector signal with known first and second order statistics (as is the case in most engineering applications), the GSVD solution corresponds precisely to the 'classical' Mahalanobis transformation that is commonly used in statistical estimators as a kind of prewhitening filter [45].

For more detail, the reader is referred to [45] [48]. Nice applications of the oriented signal-to-signal ratio concept include source separation techniques for fetal ECG extraction [14] [16] [20] [65], subspace methods to detect narrowband sources [63] and system identification [31] [46] [48] [51]. In general, it is expected that the framework of oriented signal-to-signal ratio will become a powerful analysis tool in what could be called factor-analysis-like modeling and identification methods of which a biomedical example will be presented in section 6.3.

4.2 Total linear least squares (TLLS)

Many problems in mathematical engineering, can be solved via a reconversion of the problem into the problem of solving a full rank overdetermined set of (conflicting) linear equations of the form $A \cdot x = b$. It is straightforward to derive the algebraic condition that guarantees the existence of at least one solution :

$$rank[A \;\; b] = rank[A]$$

However, often this condition is not satisfied due to measurement inaccuracies, limited machine precision , simplification or approximation of the original problem, non-linearities etc... The usual technique then is to extract a least squares solution (historically attributed to Gauss and Legendre) x' from the following minimization procedure :

$$x' : \; \| A \cdot x' - b \| = \min_{t} \; \| A \cdot t - b \|$$

Several efficient algorithms have been studied and are summarized in [53] [59] and the references therein. An important result is the well known orthogonal projection interpretation, in which x' is the exact solution of $A \cdot x' = b'$ where b' is the orthogonal projection of y into $R(A)$. So the least squares solution of the problem is nothing else than the solution to a different problem, in which the right hand side b is altered (as little as possible) in order to reduce the original (conflicting) problem into one that has an exact solution. The point that we want to make now is, that this modification is completely imposed on the right hand side b, while in a lot of engineering applications, not only this right hand side b is noisy, but the data collected in the matrix A are equally well perturbed by noise (mostly even by the same noise generating mechanism !). So, when there is no reason to assume the entries in the matrix A to be more accurate than the entries in the right hand side b, there is no motivation whatsoever to keep the columns in A unchanged. This reasoning is the origin of the so-called total linear least squares problem formulation [49][53]:

Find x' such that $A'x' = b'$ where

$$\| [A' \;\; b'] - [A \;\; b] \|_F = \min_{B,\, z;\, rank[B \;\; z] = rank[B]} \| [B \;\; z] - [A \;\; b] \|_F$$

Stated in this way, its solution is straightforward using the rank-reduction result of corollary 1.2. The total linear least squares solution is obtained from the SVD of the matrix $[A \;\; b]$:

Let $[A \;\; b] = U \cdot \Sigma \cdot V^t$, then the vector $[x^t \;\; -1]^t$ is parallel to the smallest right singular vector of $[A \;\; b]$ so that obtaining the solution x is merely a matter of scaling (in the generic case this poses no problem, however solutions to the non-generic case require some more attention and are discussed in [49]). Hence, once more the Singular Value Decomposition provides a fundamental tool in the analysis and numerically reliable solution of one of the most frequently occurring engineering problems. This approach was first discussed by Golub [53 and the references therein] though a completely independent problem formulation and solution was obtained at the same time by Staar [48] and VanHuffel [13] [49]. In statistics, the problem is called orthogonal regression and it was rediscovered many times, often independently . In addition to the evident reason of balance, there is also a very practical argument to spread the modification effort on both A and b, which is the following : The total effort to modify both A

and b into A' and b' corresponding to an exactly solvable set of equations can be shown to be never larger and usually much smaller than the effort necessary to perform this modification using b only [44] [49]. In [49] one can find a complete survey of the principle, algebraic and statistical properties, computational requirements and sensitivity results of the total linear least squares approach, including several practical examples and a survey of applications in which the total linear least squares approach has already proven its power [renal retention deconvolution, parameter estimation of the coefficients of the difference equation of a linear SISO system, modal structural analysis, etc...] These results are also reported in a series of papers [13][19][28][32][35][36][39][44].

In [26][27][33][42] one can find a unifying framework wherein it is shown that least squares and total least squares are both particular cases of a general identification scheme, that consists of a rank one modification of the sample covariance matrix of the data.

Let us now present an extreme example illustrating the difference between linear and total linear least squares. The example is taken from [48] :

Example :

$$\begin{bmatrix} N-1 & -1 & -1 & \cdot & \cdot & -1 \\ -1 & N-1 & -1 & \cdot & \cdot & -1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -1 & -1 & -1 & \cdot & \cdot & N-1 \\ -1 & -1 & -1 & \cdot & \cdot & -1 \\ -1 & -1 & -1 & \cdot & \cdot & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_{N-2} \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ \cdot \\ -1 \\ N-1 \\ -1 \end{bmatrix}$$

which is denoted as $A \cdot x = b$ where A is a $N \times (N-2)$matrix and b is a $N \times 1$ vector. It can be verified easily that the rank of the matrix $[A \;\; b]$ is $N-1$. This means that the vector b can not be written as a linear combination of the columns of the matrix A , indicating that the problem is not solvable exactly.

The preceding geometrical properties are depicted for the case $N = 4$ in fig. 2. It is an easy exercise to derive following results:

Linear Least Squares solution:

$$x' = -0.5[1...1..1]^t \;\; and \;\; b' = 0.5[-2 \;\; -2... -2 \;\; N-2 \;\; N-2]^t$$

$$b - b' = 1/2[0 \;\; 0...N \;\; -N]^t$$

and hence :

$$\| b - b' \| \; / \; \| b \| = \sqrt{N/(2(N-1))} \longrightarrow 1/\sqrt{2} \quad for N \longrightarrow \infty$$

Total Linear Least Squares solution:

$$([A \;\; b] - [\hat{A} \;\; \hat{b}]) = (\frac{1}{N-1})[-1 \;\; -1 \cdots -1 \;\; N-1]^t \cdot [1 \;\; 1 \cdots \;\; 1]$$

$\hat{x} = -[1 \;\; 1 \cdots \;\; 1]^t$ and hence :

$$\| b - \hat{b} \| \; / \; \| b \| = 1/(N-1)$$

which for $N \rightarrow \infty$ goes to 0 ! Comparing the required modification energies, results in :

$$\| b - b' \|^2 \; / \; \| [A \;\; b] - [\hat{A} \;\; \hat{b}] \|_F^2 = N/2$$

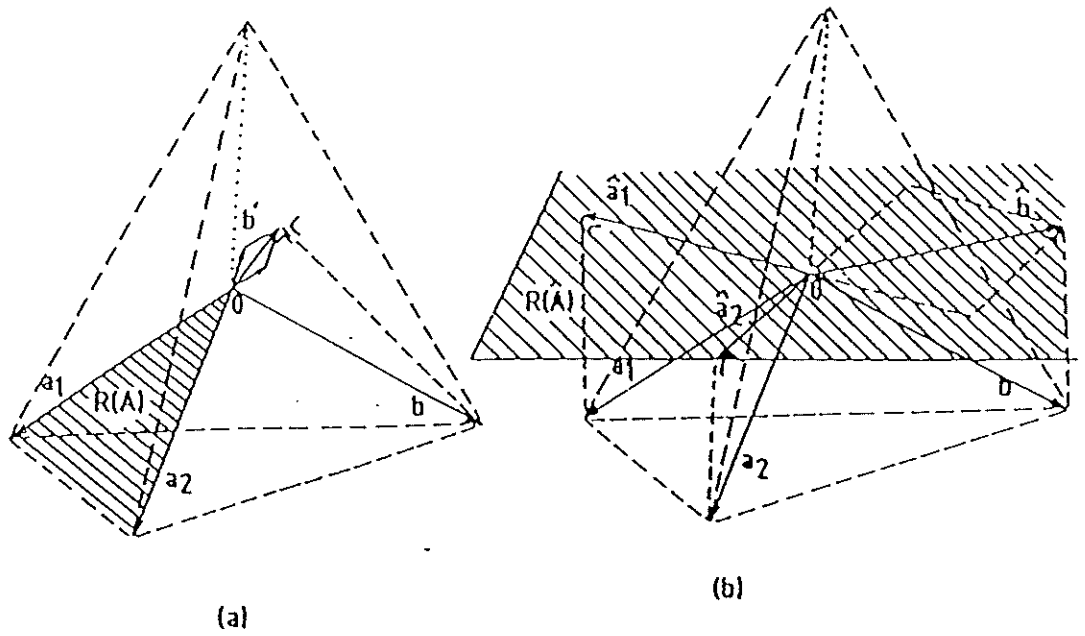

Figure 2: Geometric illustration of the difference between LLS and TLLS in the solution of $A.x = b$ where A has 2 columns a_1 and a_2. In the LLS solution the vector b is projected orthogonally onto R(A). In the TLLS solution, a_1 , a_2 and b are bent to each other, resulting in a new $\hat{a}_1, \hat{a}_2, \hat{b}$ such that rank $\left[\hat{a}_1\ \hat{a}_2\ \hat{b}\right] = 2$.

Hence, the conclusion is that this ratio of energies for the linear least squares approach with respect to the total least squares approach, increases linearly with N. The linear least squares approach always requires more modifcation energy than the total linear least squares approach. Moreover in this example it is clear that if only the right hand side is allowed to vary , it has to be modified over approximately 70 %. If the matrix A is allowed to vary also , a small change of $1/(N-1)$ is sufficient. If for instance the data in the example had been obtained by measurements and both the entries of A and b had been measured with a same absolute precision ϵ, then the right hand side and each of the columns of A are available with a relative precision $\epsilon/\sqrt{(N-1)}$ and a solution in the classical least squares sense will be merely a meaningless, purely formal approximation, whereas in the total linear least squares sense, it is still possible to find a quite accurate approximate solution in a well defined and meaningful sense.

4.3 Canonical correlation and angles between subspaces

The (generalized) SVD provides an important tool in the generalization and characterization of important geometrical concepts. One of these is the notion of angles between subspaces, which is a generalization of the angle between two vectors.

Definition 4 *Let F and G be subspaces in R^m whose dimensions satisfy*

$$p = dim(F) \geq dim(G) = q \geq 1$$

The principal angles $\theta_1, \theta_2, .., \theta_q \in [0, \pi/2]$ *between* F *and* G *are defined recursively by:*

$$\cos(\theta_k) = \max_{u \in F} \max_{v \in G} u^t.v = u_k^t.v_k$$

subject to

$$\begin{aligned} &\| u \| = \| v \| = 1; \\ &u^t.u_i = 0 \quad i = 1, ..., k-1 \\ &v^t.v_i = 0 \quad i = 1, ..., k-1 \end{aligned}$$

The vectors $\{u_i\}, \{v_i\}, i = 1, .., q$ *are called the principal vectors of the subspace pair [F,G]*

If the columns of P $(m \times p)$ and Q $(m \times q)$ define orthonormal bases for the subspaces F and G respectively, then it follows from the minimax characterization of singular values [53] that :

$$\begin{aligned} &[u_1, \cdots, u_p] = P \cdot Y \\ &[v_1, \cdots, v_q] = Q \cdot Z \\ &\cos(\theta_k) = \sigma_k \quad k = 1, \cdots, q \end{aligned}$$

where the SVD of the ('generalized inner') product

$$P^t \cdot Q = Y \cdot diag(\sigma_1, .., \sigma_q) \cdot Z^t$$

From this, it is not difficult to devise an algorithm to compute the intersection of subspaces that are for instance the ranges of two given matrices A $(m \times p)$ and B $(m \times p)$ [53,p.430] This is precisely the idea behind the technique of canonical correlation, which appears to be very fruitful in the identification of linear dynamical state space models from noisy input-output measurements [46] [51] and will be summarized in section 6.5.

There are several ways to compute the canonical correlation structure of a matrix pair A,B (roughly all possible ways of computing an orthonormal basis for the row spaces) : two QR decompositions, two singular value decompositions, one generalized singular value decomposition, all of which are followed by an SVD of the generalized inner product of the orthonormal bases matrices. Another method is the computation of the right null space of the concatenated matrix

$$\begin{bmatrix} A \\ B \end{bmatrix}$$

However, it is expected that depending on the application at hand, one method could be prefarable with respect to the others. Experiences with these several alternatives will be reported in some future work.

Finally, let us remark that there exists an intimate relation between total linear least squares and canonical correlation analysis : Let A be an $m \times n$ $(m > n)$ and B an $m \times p$ $(m > p)$ matrix and consider the problem of solving X from $A.X = B$ One can then study the canonical correlation analysis applied to the column spaces of A and B and investigate the solutions to $A.X = B.Y$, which is nothing else than a (generalized) total linear least squares problem.

5 Important properties of SVD for identification and signal processing applications

In this section, an overview is presented of some useful insights and results obtained at ESAT in the last decennium. They form the cornerstone of a lot of applications, a survey of which will be presented in section 6.

5.1 Distance from non-genericity

A nice interpretation (discussed in [51]) of the SVD is that the smallest singular value of a full (column- or row-) rank matrix measures the distance of that matrix from non-genericity. Indeed, a randomly specified mxn matrix will generically be of full rank. In the $(m \cdot n)$-dimensional space of all possible elements of the matrix, the set of rank-deficient matrices are represented by hypersurfaces of which the equation is determined by zeroing conditions on the determinants of all possible minors of the matrix [67]. These hypersurfaces can look quite complicated. The union of these 'thin' sets is a 'thin' set which represents the non-generic situation. The smallest singular value measures nothing else than the distance of a matrix to this thin set of non-generic hypersurfaces (without computing their equation !) .

5.2 A backward error theorem

In the convergence analysis of SVD algorithms,like in [15] [21], the following little known 'backward error' theorem applies . It is a generalization of a similar result for the symmetric eigenvalue problem [62] and was obtained in [21] :

Theorem 5 *Let $[u, s, v]$ be an exact singular triplet of the matrix A and let $[x, \rho, y]$ be an approximation. Define the generalized Rayleigh quotient $\rho = x^t \cdot A \cdot y = y^t \cdot A^t x$ and the residus $r_x = A^t \cdot x - \rho \cdot y$ and $r_y = A \cdot y - \rho \cdot x$ Then, x and y are left and right singular vectors of a modified matrix of the form $A - M$ where M is defined via its SVD as :*

$$M = [x \;\; r_y / \parallel r_y \parallel] \begin{bmatrix} \parallel r_x \parallel & 0 \\ 0 & \parallel r_y \parallel \end{bmatrix} \begin{bmatrix} r_x^t / \parallel r_x \parallel \\ y^t \end{bmatrix}$$

This theorem can be applied in the computation of the SVD of noisy matrices by iterative algorithms (Power method [15], Chebyshev method [21], Lanczos [49] [53], Jacobi [30] [43] [62]) in the following way :

Assume that at a certain moment in the iteration an approximation $[x, \rho, y]$ for the k-th triplet $[u_k, \sigma_k, v_k]$ of a matrix A is available where $\rho = x^t.A.y = y^t.A^t.x$. We exploit the following lemma [53] :

Lemma 1 *Given two matrices $m \times n$ matrices A and B with singular values $\{a_i\}, \{b_i\}$ then $|a_i - b_i| \leq \parallel A - B \parallel$*

One can choose for B the matrix $A - M$ where M is given as in the theorem, so that :

$$B = A - M$$

Hence :

$$|\sigma_k - \rho| \leq \parallel A - (A - M) \parallel = \parallel M \parallel = max(\parallel r_x \parallel, \parallel r_y \parallel)$$

The error in the approximation ρ of the singular value σ_k is bounded by the norm of the backward error matrix M. If this norm of M becomes smaller than the uncertainty level of the noise that corrupts the data, further improvement of the estimate is useless. As an example, consider the uncertainty level in a matrix A of which the entries are perturbed by independent identically distributed noise of zero mean with variance σ^2 . Then it is easy to show (section 5.7) that the uncertainty level of every singular value equals $max(m, n).\sigma^2$. Hence the iteration can stop when $\parallel M \parallel < max(m, n).\sigma^2$.

5.3 A sensitivity theorem

Sensitivity results on the singular value decomposition are by now well known. Both the sensitivity of the singular spectrum and the sensitivity of the singular subspaces have been the subject of extensive study [53] [59] [62] [66]. A lot of these results are quite elaborate and at the same time not so easy for intuitive interpretation. A simpler but more suggestive approach is summarized in the following theorem, in which :

- the effects related to the singular spectrum are separated from those related to singular subspaces.
- it is described how the singular values and spaces can be forced away as far as possible while perturbing the matrix A as little as possible .

Theorem 6 *On rotations of maximal sensitivity. Let the SVD of the $m \times n$ matrix A be $A = U \cdot \Sigma \cdot V^t$.*

If $A + dA = U \cdot (\Sigma + d\Sigma) \cdot V^t$ then $\| dA \|_F = \| d\Sigma \|_F$

If $A + dA = (U \cdot G) \cdot \Sigma \cdot (H \cdot V)$ where i, j, G, H and z are related by :

$$G = \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccccccc} & & & i & & j & & & m \end{array} \\ \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cos(z) & \cdot & \sin(z) & 0 & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & -\sin(z) & \cdot & \cos(z) & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 1 & 0 \\ 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ \end{array} \begin{array}{c} \\ \\ \\ i \\ \\ j \\ \\ \\ m \end{array}$$

$$H = \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccccccc} & & & i & & j & & & n \end{array} \\ \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cos(z) & \cdot & -\sin(z) & 0 & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \sin(z) & \cdot & \cos(z) & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 1 & 0 \\ 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ \end{array} \begin{array}{c} \\ \\ \\ i \\ \\ j \\ \\ \\ n \end{array}$$

then $\| dA \|_F = (\sigma_i - \sigma_j)/\sqrt{2} \cdot \sin(z)$

For a proof , the reader is referred to [48]. The interpretation of this theorem is simple and elegant : For strictly spectrum related effects, variations in the singular spectrum and on the matrix A have exactly the same level. The proof of this classical result is based on the Frobeniusnorm preserving property of orthonormal matrices. When two singular vectors are rotated simultaneously over an angle z, within their own span, one obtains the maximal forward effect on the matrices U and V, for a minimal effect in A. The effort, necessary in

matrix A, to modify one singular vector, can be decomposed into elementary rotations in 2-dimensional singular planes, spanned by the considered singular vector and any remaining one. The effort, necessary to rotate a singular vector in such a plane, is proportional to the separation between the two involved singular values. Bending a singular vector in the direction of another one, is much easier when the involved singular values are close. The fact that the singular vectors become less well conditioned when their corresponding singular values are closer, is of course well known in numerical analysis [53], but it is very convenient indeed that this phenomenon can be established via the straightforward insight provided by the preceding theorem.

Let us conclude that a first order perturbation analysis for the singular value decomposition has been derived in [51], which is perfectly similar to the first order perturbation analysis of the eigenvalue decomposition in the pioneering work of Wilkinson [66]

5.4 Condition numbers

5.4.1 The 2-norm condition number

In many applications, appropriate measures of the sensitivity of the solution are useful. One of the matrix related measures is the so-called condition number of a matrix, which applies for linear system sensitivity problems among others [53]. The precise definition of the condition number depends on the used norm. The 2-norm condition number of a mxn matrix A plays an important role in a lot of applications. If the singular values of A are denoted by σ_i, then the condition number is defined as :

$$K_2(A) = \sigma_1/\sigma_p \text{ where } p = min(m,n)$$

If $K_2(A)$ is large, then A is said to be ill-conditioned, otherwise it is well-conditioned. In 2-norm, orthonormal matrices are perfectly conditioned : $K_2 = 1$, a property that guarantees numerical reliability in many 'modern' algorithms, including SVD methods. A geometrical interpretation of this condition number can be given via the hyper-ellipsoid that is associated with every matrix A [45] [48][53].

5.4.2 A restricted condition number

Condition numbers arise naturally in describing the sensitivity of solutions of sets of linear equations to inaccuracies in the data. Consider hereto the following two problems.

Problem 1 : For a given $m \times n$ matrix $A(m > n)$, how much is the maximal increase in relative inaccuracy, that may occur when solving x from $A \cdot x = y$, for the worst position of the right hand side y and the unluckiest position of the error dy. This is a question to which the condition number as derived above gives the answer :

$$K(A) = Max_{y,\, dy} \frac{\| dx \| / \| x \|}{\| dy \| / \| y \|} = \sigma_1/\sigma_n$$

Problem 2 : For a given $m \times n$ matrix A $(m > n)$ and fixed right hand side y , how much is the maximal increase in relative inaccuracy that may occur when solving x from $A.x = y$, for the unluckiest position of the error dy.

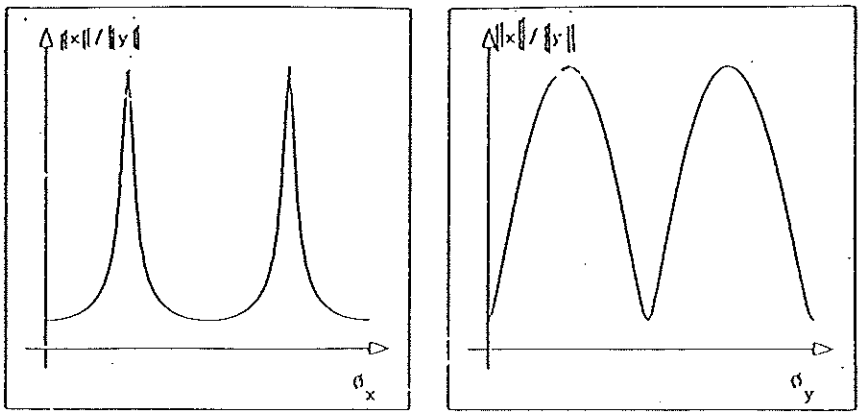

fig. 3.a.

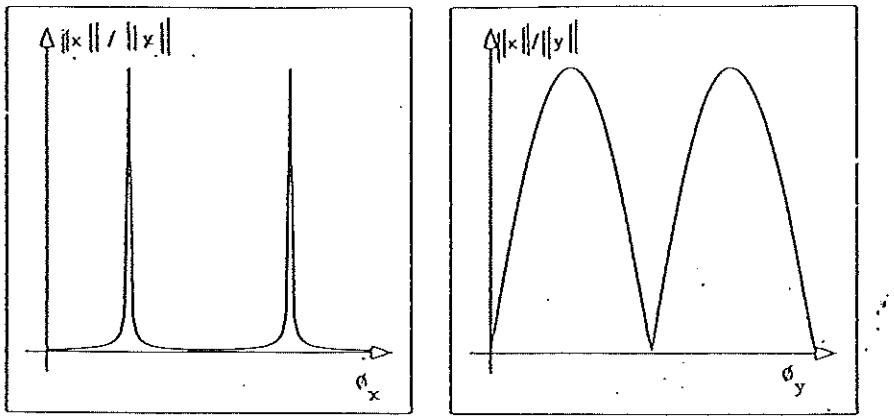

fig.3.b.

Figure 3: Norm amplification $\| x \| / \| y \|$ as a function of θ_x (angle between x and largest right singular vector) and θ_y (angle between y and largest left singular vector) for a 2 × 2 example. In fig. 3.a. the condition number is 10. In fig.3.b. it is 100.

The singular value decomposition allows for a direct answer via a kind of 'restricted' condition number :

$$K_y(A) = Max_{dy, A.x=y, A.dx=dy} \frac{\| dx \| / \| x \|}{\| dy \| / \| y \|} = \frac{\| y \|}{\sigma_1 \cdot \| x \|} \cdot K(A)$$

The first condition number only gives information about the matrix A, not taking into account the relative position of the right hand side y. The second condition number takes into account the actual position of y, which may drastically reduce the estimation of the solution sensitivity. First consider the norm amplification $\| x \| / \| y \|$ as a function of the angle θ_x between x and the largest right singular vector and the angle θ_y between y and the largest left singular vector for a simple 2 × 2 example. Two typical patterns of norm amplification are depicted in fig. 3 for a condition number K(A)=10 and for K(A)=100. One can observe that the property of 'being usually close' to either the maximum or the minimum significantly increases with increasing condition number. The minimum in all curves corresponds to the inverse of the largest singular value while the maximum equals the inverse of the smallest one.

Some important conclusions, which are general, can be drawn from these observations:

- In a lot of applications, the orientation of the error dy is unknown but one has an idea

about the magnitude of the norm $\| dy \|$. Hence, it is meaningful to assume the effect on x most probably to be of a level of $(1/\sigma_r).\ \| dy \|$, where σ_r is the smallest singular value of A.

- If each angle θ_x is equally probable, then x will have most probably a length equal to $1/\sigma_1.\ \| y \|$. One can verify that this is the case in discrete deconvolution problems, where the impulse response samples are computed from input-output measurements [48]. The error $\| dx \|$ will most probably be in these applications of the level $K(A)\ \| dy \|$, which can be rather bad.

- If each angle θ_y is equally probable, then x will have most probably a length close to $1/\sigma_r.\ \| y \|$. This is the case in applications such as polynomial fitting via Vandermonde matrices [48]. The error $\| dx \|$ will be probably of the level $K_y(A).\ \| dy \|$.

Now consider the sensitivity measure m_e , which is the error amplification

$$m_e = \frac{\| dx \| / \| x \|}{\| dy \| / \| y \|}$$

for all possible orientations of x, dx, y, dy . For a fixed y, m_e varies from $K_y(A)/K(A)$ to $K_y(A)$ depending on the orientation of dy. The worst case result for dy for fixed y is $m_e = K_y(A)$. Over all possible orientations of y, $K_y(A)$ varies from 1 to K(A). Moreover, if dy has a uniform distribution elementwise, then m_e is usually close to its upperbound $K_y(A)$. If dx has a uniform distribution elementwise, then m_e is usually much smaller than $K_y(A)$. One can conclude that the linear least squares solution x and the error dx tend to be:

- dy-insensitive if y is independent of A (polynomial smoothing)

- dy-sensitive if y is dependent on A (like in deconvolution)

5.5 On the condition of a linearly extended matrix

The following useful result was obtained in [7], in the analysis of the numerical conditioning of electrical network descriptions. However, it is believed that this result might also be useful in other applications. Moreover, the result is important enough in its own right to state it here as follows.

Theorem 7 *For any two pairs of matrices A and $B = [A\ \ A{\cdot}x]$ with finite condition numbers, where x is an arbitrary vector, the ratio of the condition numbers $K(A)/K(B)$ is bounded*

$$(1 + \|A.x\|_2)^{-1/2} \leq K(A)/K(B) \leq (1 + \| A.x \|_2)^{1/2}$$

For a proof , the reader is referred to [7]. Observe that an analoguous theorem and proof can be given for the addition or deletion of a linearly dependent row.

5.6 Properties of the SVD of structured matrices

A $p \times q$ Hankel matrix H_{pq} constructed from the elements of an array of real numbers $h = \{h_k\}, k = 1, .., K\ (K > (p + q - 1))$ is a matrix with elements $H_{pq}(i, j) = h_{i+j-1}$. Rank deficient Hankel matrices of rank $r < min(p, q)$ play an important role in the analysis and realization of state space models from linear systems. Since the introduction of the by now

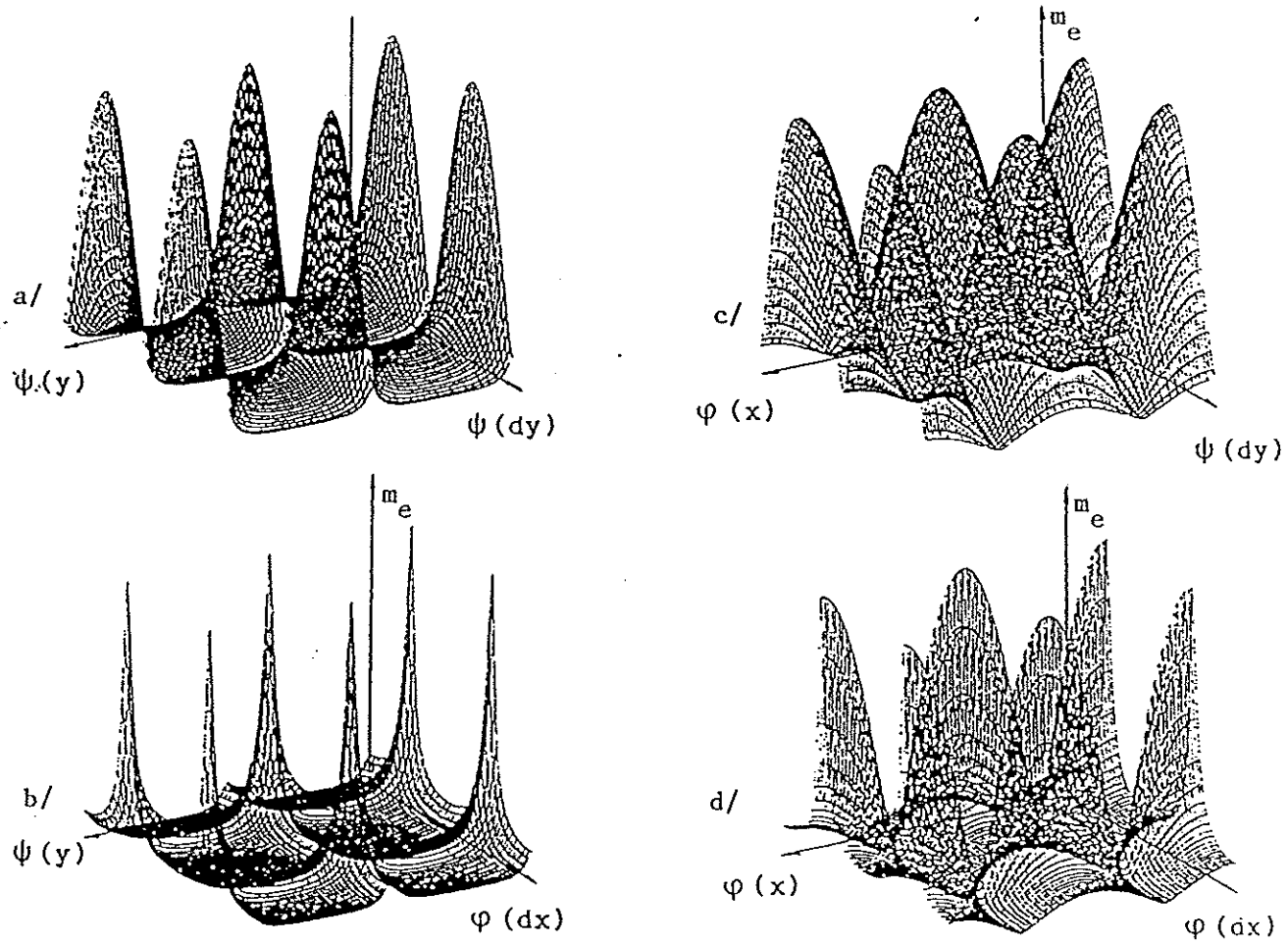

Figure 4: Error amplification ($\| dx \| / \| x \|$)/($\| dy \| / \| y \|$) for a 2 x 2 example for a uniform distribution in y and dy (4.a.), y and dx (4.b.), x and dy (4.c) and x and dx (4.d). The angles are with respect to the largest right (for x and dx) and largest left (for y and dy) singular vectors.

well understood realization theory of Ho-Kalman [55] , their system theoretic interpretation has been established. An important step was the introduction in the realization context of the singular value decomposition of the Hankel matrix, containing (possibly noise corrupted (section 6.4)) Markov parameters [57] [58]. Moreover there exists a close connection between the singular value decomposition and the concept of balanced realization [60]. It can be shown that the singular values can be used as quantitative measures of controllability and observability. In [38], some intriguing results have been obtained about the relation between the structure of rank deficient Hankel matrices and their singular value decomposition. Hereto, use is made of the matrix calculus of Kronecker and Khatri-Rao product. We will briefly summarize here the main results. For more information, the background material and the appropriate literature, the reader is referred to [38].

Definition 5 *Kronecker product.*

The Kronecker product of a $p \times q$ matrix A and a $m \times n$ matrix B , denoted by $A \otimes B$, is defined as the $(pm) \times (qn)$ matrix :

$$A \otimes B = \begin{bmatrix} a_{11}.B & a_{12}.B & \cdots & a_{1q}.B \\ a_{21}.B & a_{22}.B & \cdots & a_{2q}.B \\ \vdots & \vdots & & \vdots \\ a_{p1}.B & a_{p2}.B & \cdots & a_{pq}.B \end{bmatrix}$$

Definition 6 *Khatri-Rao product*

The Khatri-Rao product of 2 matrices A $(p \times q)$ and B $(m \times q)$ (same numbers of columns !), denoted by $A \square B$ is the $(pm) \times q$ matrix, defined by:

$$A \square B = [a_1 \otimes b_1 ... a_q \otimes b_q]$$

where a_i, b_i are the i-th column of A, resp. B.

So the Khatri-Rao product of two matrices is nothing else than their Kronecker product columnwise.

Now let us introduce some notation. The symbol vec(A) refers to a vector formed by stacking the columns of the matrix A in a long column vector. vecd(A) is also a column vector construction but only with the diagonal elements of A. If A is an mxn matrix , then $\bar{A}(\underline{A})$ is an $(m-1) \times n$ matrix obtained from A by omission of its first (last) row. $[\underline{A} \quad \bar{A}]$ is an $(m-1) \times 2n$ concatenated matrix.

By exploiting the properties of the Kronecker and Khatri-Rao products, one can prove very easily the following theorems:

Theorem 8 *If G_{pq} is a $p \times q$ matrix with SVD: $G_{pq} = U.\Sigma.V^t$ then G_{pq} is a Hankel matrix if and only if :*

$$\underline{U}.\Sigma.\bar{V}^t = \bar{U}.\Sigma.\underline{V}^t$$

This result is a special case of a more general factorisation result stated in [38]. The Khatri-Rao product provides an important new insight into the structure of the SVD of rank deficient Hankel matrices :

Theorem 9 *Let G_{pq} be a matrix of rank $n < min(p,q)$ with SVD :*

$$\begin{array}{ccccc} G_{pq} & = & U & .\Sigma. & V^t \\ p \times q & & p \times n & n \times n & n \times q \end{array}$$

Then G_{pq} is a Hankel matrix if and only if the singular values are the solution to the set of linear equations :

$$[\underline{V} \Box \bar{U} - \bar{V} \Box \underline{U}].vecd(\Sigma) = 0$$

$[\underline{V} \Box \bar{U} - \bar{V} \Box \underline{U}]$ is a $n^2 \times n$ matrix which is generically of *rank* $n^2 \times n - 1$. So generically the set of linear equations has only one solution (up to a scalar) ! Once the singular vectors of a rank deficient Hankel matrix are known, the structure of the Hankel matrix restricts the number of allowable singular values.

Theorem 10 *Let U $(p \times n)$ and V $(q \times n)$ contain the n left and right singular vectors of a $p \times q$ Hankelmatrix of rank $n < min(p,q)$. Then :*

$$rank[\underline{V} \otimes \bar{U} - \bar{V} \otimes \underline{U}] = n^2 - n$$

The matrix $[\underline{V} \otimes \bar{U} - \bar{V} \otimes \underline{U}]$ and vectors lying in its null space appear in the following :

Theorem 11 *Let U $(p \times n)$ and V $(q \times n)$ contain the n left and right singular vectors of a $p \times q$ Hankelmatrix H_{pq} of rank $n < min(p,q)$ Let vec(X) be a vector in the null space of $[\underline{V} \otimes \bar{U} - \bar{V} \otimes \underline{U}]$:*

$$[\underline{V} \otimes \bar{U} - \bar{V} \otimes \underline{U}].vec(X) = 0$$

where X is a $n \times n$ matrix. Then the matrix $H = U.X.V^t$ is a Hankel matrix. If the SVD of X is given by

$$X = U_x.\Sigma_x.V_x^t$$

then the SVD of that Hankel matrix H is given by :

$$H = (U.U_x).\Sigma_x.(V.V_x)^t$$

An additional result in [38] is that the system poles associated with H are a subset of those associated with H_{pq}. The last theorem provides a parametrization of all rank deficient Hankel matrices of dimensions $p \times q$ of which the system poles come from a specified set of system poles. It is conjectured that this theorem can be used in the optimization of the joint controllability and observability of a system over a finite time horizon.

Example : Consider a second order linear system with state space matrices

$$A = \begin{bmatrix} 0.9 & 0 \\ 0 & 0.8 \end{bmatrix} \quad B = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad C = [1 \;\; 1]$$

Its Markov parameters are the scalars $h_k = C.A^{k-1}.B$, $k = 1, 2, ..$ Construct any $p \times q$ Hankel matrix $(min(p,q) > 2)$ with these scalars, say a 4×3 Hankel matrix with elements $[2; 1.7; 1.45; 1.241; 1.0657; 0.9182]$ and compute its SVD.

$$\begin{array}{rccccl} H_{43} = & U & .\Sigma. & V^t & \text{with singular values} & \sigma_1 = 4.8808 \\ & 4 \times 2 & 2 \times 2 & 2 \times 3 & & \sigma_2 = 0.0142 \end{array}$$

The reader may now wish to verify that any 4×1 vector vec(X), where X is a 2×2 matrix, of the null space (which is two dimensional) of the 6×4 matrix $[V \otimes \bar{U} - \bar{V} \otimes U]$, gives rise to a Hankel matrix $U.X.V^t$.

5.7 The SVD of the sum of rectangular matrices with mutual orthogonal rowspaces

In a lot of mathematical engineering applications, matrices from measured data are constructed. Very often these matrices are rectangular, i.e. they have (say) many more columns than rows. For instance, there are as many rows as measurements channels and each time a measurement on all channels is done, an additional column is added to the matrix. The available data are very frequently rather noisy, i.e. they are perturbed by unobservable errors, of which (in the best case) the statistical properties are known. Moreover, almost always the 'exact data' mathematical model is such (at least in linear applications) that the data matrix would be rank-deficient if the data were noisefree, and that the crucial information about the desired model is contained in the null space of the exact data matrix. Think for instance of the solution of overdetermined linear equations (section 4.2), or dynamical realization of impulse responses from (block)-Hankel matrices (section 6.5) or the identification of dynamical state space models from noisy input-output data with canonical correlation analysis (section 6.6). More specifically, we are confrontated with the following situation :

$$\underset{m \times n}{C} = \underset{m \times n}{A} + \underset{m \times n}{B} \qquad \text{with } m \ll n \text{ where } r(A) = r < m$$

where A represents the exact data matrix of rank r and B represents the matrix containing the perturbations (measurement errors, model mismatch), which are assumed to be additive. A and B as such are unobservable but the entries of their sum C are the measured data. Generically, the matrix B will be of full (row) rank while the matrix A is assumed to be rank-deficient : The crucial model information (in all cited examples) is contained within the orthogonal complement of its columnspace.

The problem that will now be investigated is the following : How is the SVD of A modified by the perturbing influence of B and how can properties of A (rank and null spaces) be estimated from computations on C only. The crucial observation to make is contained in the following statement : **Under mild conditions, the canonical angles between the row space of the matrix A and the row space of the matrix B, approach 90^0 (orthogonality) as the overdetermination n / m increases .** A general discussion and proof of this statement and the conditions under which it is valid can be found in [51]. An illustration can be found in the following figures, where the probability distribution is computed of the angle between a vector and several subspaces in an ambient space of varying dimension. The elements of the vector are independently identically normally distributed with zero mean. It's the authors' belief that this observation is really at the heart of a lot of identication and estimation schemes such as (total) linear least squares, instrumental variables methods, dynamic identification, briefly in all those applications in which

- the estimation of the rank of an exact data matrix through computations on a noisy matrix makes sense.

- the information on the model is contained in the null space of the measurement matrix.

In [26] [27] [33] this statement is even taken to be an axiomatic basis for a new conceptual identification framework.

Let us demonstrate how the preceding statement can be applied to the modification analysis of the SVD of A and C . We will derive the conditions under which the column space

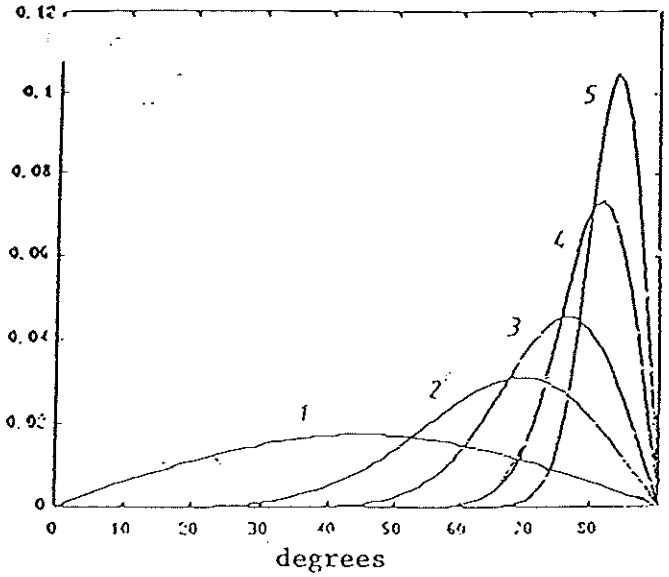

Fig.5.a.

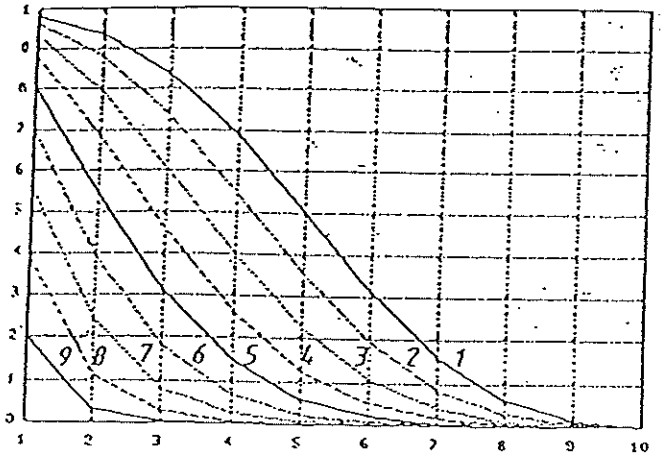

Fig.5.b.

Figure 5: 5.a. Probability distribution of the angle between an arbitrary direction in R^j (uniformly distributed) and a fixed 2-dimensional plane for j=4 (1), j=10 (2) , j=20 (3) , j=50 (4), j=100 (5) . Remark the increasing probability that the direction is orthogonal to the plane. 5.b. Probability that an arbitrary uniformly distributed direction in R^{10} makes an angle of more than 45 (1), 50 (2), ..., 85 (9) degrees with a fixed subspace S^i for i=1,2,...,10.

of the matrix A , including its dimension (rank) can exactly be recovered from a subspace of the columnspace of the matrix C .

Let A and B have the SVD's :

$$\underset{m \times n}{A} = \underset{m \times r}{U_a} \cdot \underset{r \times r}{\Sigma_a} \cdot \underset{r \times n}{V_a^t}$$

$$\underset{m \times n}{B} = \underset{m \times m}{U_b} \cdot \underset{m \times m}{\Sigma_b} \cdot \underset{m \times n}{V_b^t}$$

where $r < m < n$. Denote by $U_a^\perp$ any $m \times (m-r)$ orthogonal matrix satisfying

$$U_a^t . U_a^\perp = 0$$

and assume that the row spaces of A and B are orthogonal :

$$V_a^t . V_b = 0$$

Then, it is easy to show that the matrix C can be written as :

$$C = [U_a \;\; U_a^\perp] \begin{bmatrix} P_1^t \\ P_2^t \end{bmatrix} \quad with \quad \begin{array}{l} P_1^t = \Sigma_a V_a^t + U_a^t . U_b . \Sigma_b . V_b^t \\ P_2^t = (U_a^\perp)^t . U_b . \Sigma_b . V_b^t \end{array}$$

Let P_1^t and P_2^t have the SVD's :

$$P_1^t = X_1 \cdot S_1 \cdot Y_1^t$$
$$P_2^t = X_2 \cdot S_2 \cdot Y_2^t$$

then we have the result that C can be written as :

$$C = [U_a . X_1 \;\; U_a^\perp . X_2] \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix} \begin{bmatrix} Y_1^t \\ Y_2^t \end{bmatrix}$$

which is (up to a reordening of the singular values in S_1 and S_2) a singular value decomposition if :

$$Y_1^t . Y_2 = 0$$

or equivalently :

$$P_1^t . P_2 = 0$$

Because $V_a^t . V_b = 0$ this is true iff :

$$U_a^t . [U_b . (\Sigma_b . \Sigma_b^t) . U_b^t] . U_a^\perp = 0$$

The factor between square brackets can be recognized to be the sample covariance matrix of the vector signal contained in the matrix B. Now let Q_1 and Q_2 be such that :

$$U_a = U_b . Q_1 \;\; U_a^\perp = U_b . Q_2$$

then the following set of matrix equations is to be satisfied :

$$\begin{array}{l} Q_1^t . Q_1 = I_r \\ Q_2^t . Q_2 = I_{m-r} \\ Q_1^t . Q_2 = 0 \\ Q_1^t . (\Sigma_b . \Sigma_b^t) . Q_2 = 0 \end{array}$$

This poses no problem if $(\Sigma_b.\Sigma_b^t)$ is a multiple of the identity matrix. This is then equivalent with the fact that the oriented energy of the matrix B is isotropic. For general $(\Sigma_b.\Sigma_b^t)$ it is not so difficult to see that a solution looks as follows : Q_1 may contain any set of r different columns of the identity matrix I_m (or a linear combination of these) while Q_2 may contain the remaining m-r columns (or a linear combination of them) Hence, Q_1 and Q_2 must have a complementary zero pattern. The conclusion is that U_a must be generated by r singular vectors of U_b.

Hence, we have demonstrated that under the orthogonality condition of the row spaces of A and B, the subspace spanned by the vectors of U_a is not mixed with vectors of the subspace spanned by $U_a^\perp$ if either the oriented energy distribution of B is isotropic or if the subspace U_a is generated by r left singular vectors of the matrix B. However, it can only be recognized as such from the singular values, if one has a priori information about the relative magnitude of the singular values contained in S_1 and S_2. If one knows for instance a priori that all singular values in S_1 are larger than those in S_2 , then the first r left singular vectors of C generate the subspace spanned by the vectors of U_a . As an illustration consider the following:

Example : If the entries of B are identically normally distributed with zero mean and variance σ^2, then with increasing probability for increasing overdetermination n/m the SVD of C will approach the following 'limit' SVD :

$$[U_a \;\; U_a^\perp].\begin{bmatrix} (\Sigma_a^2 + n.\sigma^2 I_r)^{1/2} & 0 \\ 0 & \sqrt{n}.\sigma.I_{m-r} \end{bmatrix}\begin{bmatrix} (\Sigma_a^2 + n.\sigma^2 I_r)^{-1/2}(\Sigma_a V_a^t + \sqrt{n}.\sigma.V_1^t) \\ V_2^t \end{bmatrix}$$

where $V_1 = B^t.U_a/(\sqrt{n}\sigma)$ and $V_2 = B^t.U_a^\perp/(\sqrt{n}\sigma)$.

Note that the noise covariance matrix is a multiple of the identity matrix $E(B.B^t) = n\sigma^2.I_m$. Moreoever, remark the Pythagoras-like squaring of the scaled orthogonal matrices in the upper part of the right singular matrix. This implies that one may not hope to recover approximately the row space of the matrix A from the analysis of the matrix C. The columnspace of A however can be recovered, by inspecting the the singular values of C : As is obvious, a 'noise treshold' $\sqrt{n}\sigma$ will be recognized in the singular values (for sufficiently large overdetermination n/m).

6 The (Generalized) Singular Value Decomposition in mathematical engineering applications

6.1 Moments of inertia

This section establishes the close connection between the SVD and the Principal Axes and Moments of Inertia (PAMI) of a rigid body as described in [48, p.81-90]. Only the discretized case will be considered. Consider a rigid body consisiting of K point masses $m(k)$, descibed with coordinates $x_i(k)$ in a N-dimensional Euclidean space. With respect to a certain reference O (often the center of gravity), one defines the moment of rotational inertia I_t around an axis t as :

$$I_t = \sum_{k=1}^{K} r_{t,m(k)}^2 m(k)$$

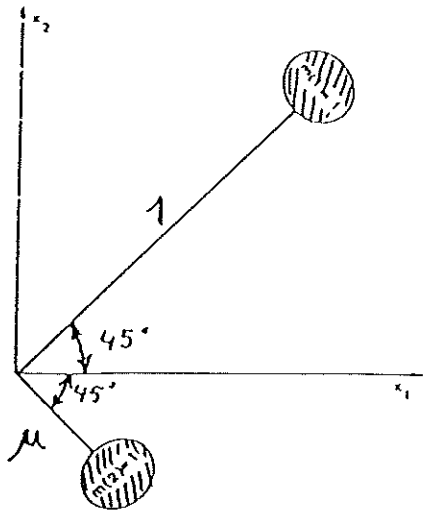

Figure 6: An example of a numerically critical PAMI problem

where $r_{t,m(k)}$ is the distance from the k-th point mass to the axis . I_t is a positive function of the orientation t, with one minimum I_{t1}, one maximum I_{tN} and $N-2$ saddle points $I_{t2},\cdots,I_{t(N-1)}$. The orientations $t_1,..,t_N$ of extremal rotational inertia form a set of orthogonal directions in the N-dimensional Euclidean space and are called the principal axis (PA) of the rigid body and the associated moments of inertia are called the principal moments of inertia (MI) with respect to O. Classically, the PAMI are obtained as the right eigenstructure of a so-called inertia tensor T, i.e. the PA are the right eigenvectors and the MI are the eigenvalues of T, where :

$$T = \begin{bmatrix} t_{11} & -t_{12} & -t13 & \cdots & -t_{1N} \\ -t_{21} & t_{22} & -t23 & \cdots & -t_{2N} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ -t_{N1} & -t_{N2} & -t_{N3} & \cdots & t_{NN} \end{bmatrix}$$

with

$$t_{ii} = \sum_{k=1}^{K} m(k) \cdot [\sum_{\substack{n=1 \\ n\neq i}}^{N} x_n^2(k)] \quad i = 1,\cdots,N$$

$$t_{ij} = \sum_{k=1}^{K} m(k) \cdot x_i(k) \cdot x_j(k) \qquad \begin{matrix} i,j=1,\cdots,N \\ i \neq j \end{matrix}$$

T is easy to obtain from the mass distribution of the considered body, described with respect to a chosen basis. It is a symmetric, positive semi-definite matrix. However, the crucial point we want to emphasize is that in forming T explicitly, important numerical accuracy may be lost! Computing the PAMI via the eigenstructure of the inertia tensor T requires ϵ^2 precision computations in order to handle properly ϵ-precision data. This statement can be illustrated with the following example :

Example: Consider a rigid body with 2 point masses $m(1) = 1$ at $[x_1(1), x_2(1)] = [1/\sqrt{2}, 1/\sqrt{2}]$ and $m(2) = 1$ at $[x_1(2), x_2(2)] = [\mu/\sqrt{2}, -\mu/\sqrt{2}]$(Fig.6). The associated inertia tensor is

$$T = 1/2 \begin{bmatrix} 1+\mu^2 & -1+\mu^2 \\ -1+\mu^2 & 1+\mu^2 \end{bmatrix}$$

The PAMI follow from the eigenstructure of T: principal axes : $[1 \quad -1]^t$ and $[1 \quad 1]^t$, associated moments : 1 and μ^2

Now let us investigate how this classical approach behaves numerically when only finite precision arithmetic is possible and/or only finite precision measurements are available.

- If the measurements are exact, but only finite precision computation is possible, then the inertia tensor T is represented in the machine by a matrix in which $m(2)$ does not even appear, if the machine precision ϵ is larger than μ^2 which means that one will find the PAMI of a body in which the mass $m(2)$ is not present anymore.

- Second consider inaccuracies γ_i on the measurements of the coordinates with an absolute level γ (all measurements are performed with the same absolute precision), satisfying the following bounds :

$$1 \gg \mu \gg \gamma \gg \mu^2$$

Such measurement errors however will affect the inertia tensor T more than the effect μ^2 of mass $m(2)$:

$$T = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \begin{bmatrix} \mu^2 & \mu^2 \\ \mu^2 & \mu^2 \end{bmatrix} + \begin{bmatrix} 2 \cdot \gamma_2 & -\gamma_1 - \gamma_2 \\ -\gamma_1 - \gamma_2 & 2 \cdot \gamma_1 \end{bmatrix} \approx \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \begin{bmatrix} 2 \cdot \gamma_2 & -\gamma_1 - \gamma_2 \\ -\gamma_1 - \gamma_2 & 2 \cdot \gamma_1 \end{bmatrix}$$

In the general case, both types of inaccuracy will combine and possibly affect the results of the computation in a severe way. However, the following observation is crucial : The observed degeneration of precision is not essential to the PAMI problem itself, but to its formulation as an eigenstructure problem. The loss of accuracy is caused by the mere fact of using the tensor T. It is again the SVD that allows to solve the problem in such a way that guarantees $\epsilon + \gamma$ accurate results with ϵ precision arithmetic on γ accurate data.

Theorem 12 *On the relation between SVD and PAMI .*

Consider a rigid body B of K discrete point masses $m(k)$, located at $x(k) = [x_1(k), \cdots, x_n(k)]^t$ in a N-dimensional coordinate system with reference O. For such a rigid body, a $N \times K$ matrix M is constructed with elements $M(i,j) = \sqrt{m(j)} \cdot x_i(j)$.

- *Then a rigid body B' can be constructed with identical PAMI with N unit pointmasses, located at*

$$\sigma_i \cdot u_i \quad i = 1, \cdots, N$$

where σ_i and u_i are the i-th singular value and i-th left singular vector in the SVD of M.

- *The principal axis of inertia of B are the left singular vectors of M*
- *The principal moments of inertia are obtained from the singular values of M as :*

$$I_{t_i} = \sum_{\substack{n=1 \\ n \neq i}}^{N} \sigma_n^2$$

For a simple proof , see [48 ,p.84].

Remark 1: The interested reader is encouraged to investigate the benefits of this relation

between SVD and the computation of the PAMI of a rigid body, by applying this result to the previous discussion where the tensor T was used instead of the matrix M. For this example, the matrix M equals:

$$M = 1/\sqrt{2} \begin{bmatrix} 1 & \mu \\ 1 & -\mu \end{bmatrix}$$

Computing the SVD of M in ϵ machine precision, delivers now the second singular value μ with a precision of ϵ/μ.

Remark 2 : A useful interpretation of this theorem is that any rigid body consisting of K point masses in N-dimensions, can be replaced by a rigid body with at most N point masses and moments of inertia as in the theorem.

Remark 3 : As a useful intermediate result, it is not difficult to verify that the M and T matrices of any rigid body are related by :

$$M \cdot M^t = \| M \|_F^2 \cdot I_{NN} - T$$

Moreover, the computation of the PAMI of a composite object consisting of two sub-bodies B_1 and B_2 is facilitated by the following corollary :

Corollary 3 *If a body is composed of two subbodies B_1 and B_2 with equivalent PAMI decompositions $U_1 \cdot \Sigma_1$ and $U_2 \cdot \Sigma_2$, then an M-matrix for B is obtained as:*

$$M_B = [U_1 \cdot \Sigma_1 \quad U_2 \cdot \Sigma_2]$$

and its associated equivalent PAMI structure is $U \cdot \Sigma$ where the SVD of $M = U \cdot \Sigma \cdot V^t$.

Remark 2 and the corollary are illustrated in figure 7:

6.2 The conditioning of the reference node choice in an electrical network

The existence and algebraic properties of singular matrix representations for N-terminal circuits have been described during nearly 60 years. Surprisingly enough, almost no attention was focussed on their numerical properties, which will be the subject of this section. It will be shown how from a practical point of view certain network descriptions are better suited when the accuracy of the computations, the tolerances of the components or the precision on the variables is pushed to extreme values. The subsequent statements apply to admittance, impedance and hybrid matrix representations but to fix ideas, consider the case of an admittance matrix representation of an N-terminal circuit :

$$Y^*[v_1^* \cdots v_t^* \cdots v_N^*]^t = [i_1^* \cdots i_s^* \cdots i_N^*]^t \quad or \quad Y^* \cdot v^* = i^*$$
$$Y[v_1 \cdots v_{t-1} \;\; v_{t+1} \cdots v_N]^t = [i_1 \cdots i_{s-1} \;\; i_{s+1} \cdots i_N]^t \quad or \quad Y \cdot v = i$$

The first equation is the indefinite representation, in which all terminal potentials v_k^* and currents i_k^* occur. The second equation is the definite representation for which node voltage v_t has been chosen as a reference voltage and the Kirchoff Current Law equation at node s has been dropped. Usually, $s = t$. Y is obtained by simply deleting row s and column t in Y^* . Also Y^* can be constructed from Y by invoking the so-called zero row sum and

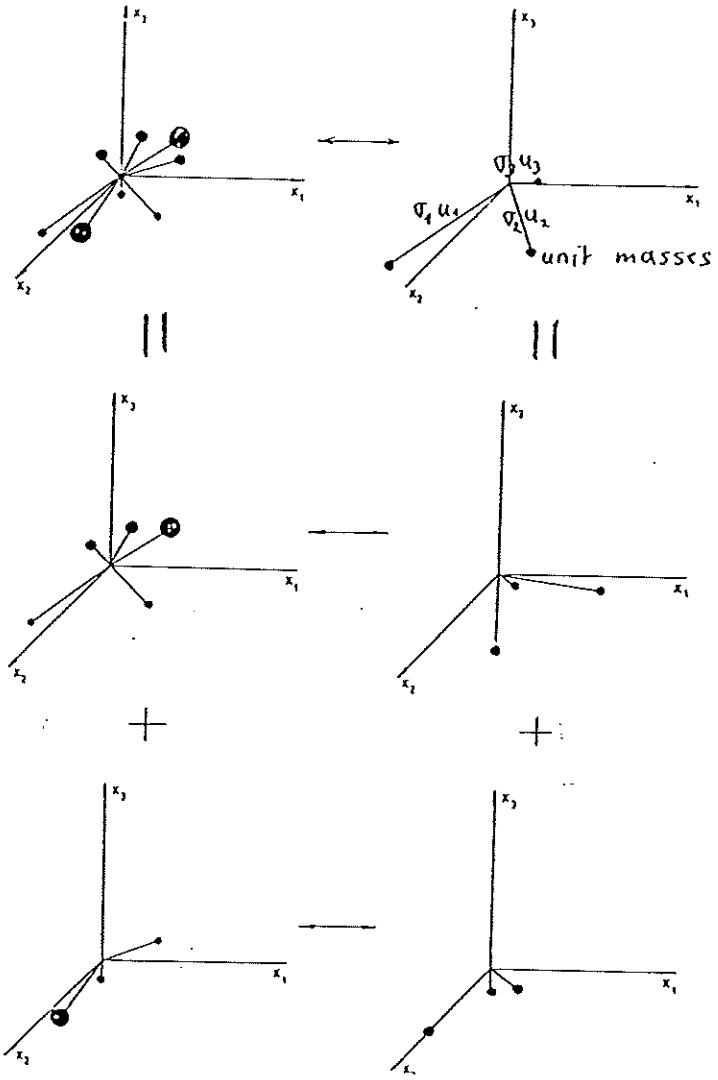

Figure 7: The discrete mass rigid body of the left upper corner is the sum of two rigid bodies. The right hand side depicts the equivalent 3 point mass configurations.

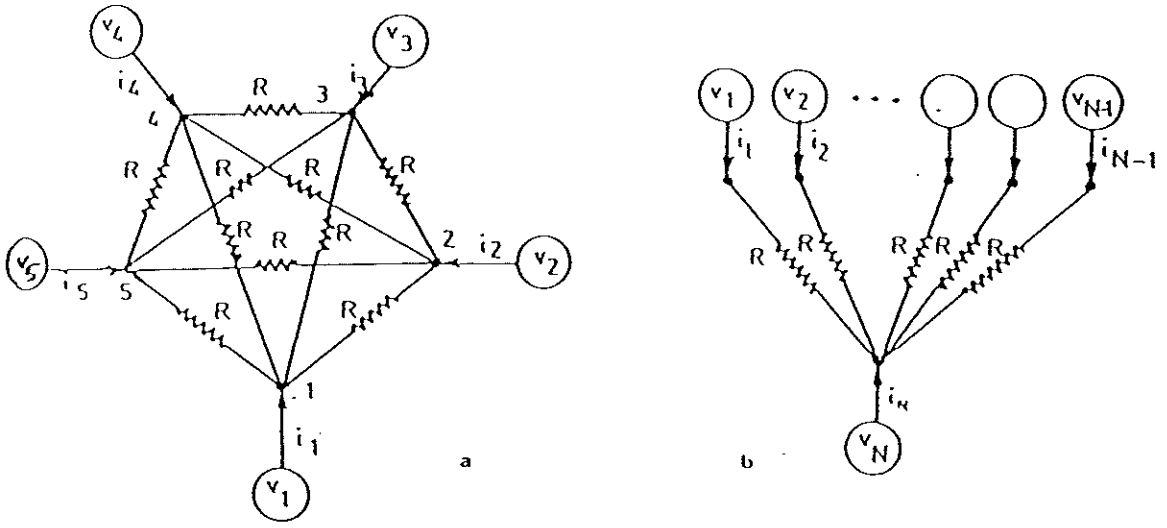

Figure 8: 8.a and 8.b Resistive circuits as extreme examples

zero column sum properties of Y^* [48] These trivial algebraic relationships can be invoked to qualify both descriptions as equivalent. However, we will now demonstrate that numerically, this statement is incorrect. It turns out that , even for surprisingly simple circuits, one of the matrices Y or Y^* may be ill-conditionned, whereas the other is well conditionned. This is stated formally in the following

Theorem 13 *On the conditioning of reference node choice*

The condition number ratio of the definite and indefinite admittance matrices Y and Y^* satisfies :

$$1/N \leq K(Y)/K(Y^*) \leq N$$

The proof is an immediate consequence of the theorem on the condition number of linearly extended matrices, presented in section 4. and can be found in [7] and [48] . This theorem generate bounds on the variation of the condition number by adding or deleting a row or a column. These bounds are however tight as is shown in the following examples :

Example : Consider the resistive network of figure 8a with a complete graph, with N nodes ($N > 2$) and all resistances equal to R .The indefinite matrix description is given by:

$$1/R \begin{bmatrix} N-1 & -1 & -1 & \cdots & -1 \\ -1 & N-1 & -1 & \cdots & -1 \\ \vdots & \vdots & \vdots & & \vdots \\ -1 & \cdots & \cdots & N-1 & -1 \\ -1 & -1 & \cdots & -1 & N-1 \end{bmatrix} \begin{bmatrix} v_1^* \\ v_2^* \\ \vdots \\ \\ v_N^* \end{bmatrix} = \begin{bmatrix} i_1^* \\ i_2^* \\ \vdots \\ \\ i_N^* \end{bmatrix}$$

The condition number of Y^* here is one : all singular values are equal to $N\sqrt{(N-1)}/R$. Hence this description is perfectly well conditioned. Now in order to obtain a definite description one might choose v_k as a reference voltage and delete the associated k-th equation. The matrix Y, which is the matrix Y^* with any row k and column k deleted, has a condition number equal to N, hence the upperbound of the theorem is reached. For large N, obviously

it is better to use the indefinite circuit description for this network topology.

Example : Consider now the network of figure 8b. Its indefinite matrix description as given by :

$$1/R \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & -1 \\ 0 & 1 & 0 & \cdots & 0 & -1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & . & . & \cdots & 1 & -1 \\ -1 & -1 & -1 & \cdots & -1 & N-1 \end{bmatrix} \begin{bmatrix} v_1^* \\ v_2^* \\ \vdots \\ \\ v_N^* \end{bmatrix} = \begin{bmatrix} i_1^* \\ i_2^* \\ \vdots \\ \\ i_N^* \end{bmatrix}$$

The condition number of Y^* is N . Hence in this case, the indefinite description is ill-conditioned for large N. Obviously, by chosing v_N as the reference voltage, and dropping the N-th equation, the resulting definite matrix Y is perfectly well conditioned (It is the unit matrix !) . This example shows that the lower bound of the theorem is tight

6.3 Fetal ECG extraction

The examples reported in this section are nice practical applications of the factor-analysis-like formalism that was explained in section 4.1.

The measurements are obtained from cutaneous electrodes placed at the heart and the abdomen of the mother. If there are p measurement channels (typically 6 to 8), the sampled data are stored in a pxq matrix M_{pq} where q denotes the number of consecutive samples that are processed. The p observed signals $m_i(t)$ (the rows of M_{pq}) are modeled as unknown linear combinations (modeled by a static $p \times r$ matrix T) of r source signals $s_j(t)$, corrupted by additive noise signals $n_i(t)$ with known (or experimentally verified) second order statistics. Hence the model has the well known factor-analysis-like structure :

$$M_{pq} = T_{pr}.S_{rq} + N_{pq}$$

where the rows of S_{rq} are the source signals . The problem now consists of a rank decision to estimate r and of a subspace determination problem to determine the subspace generated by the columns of the matrix T, which are the so-called lead vectors. Since the second order statistics are assumed known, the conceptual framework of oriented signal-to-signal ratio (Mahalanobis transformation [45]) could be applied. However, it has been verified [14] [50] (and there exists a physical explanation) for this specific application that with an appropriate positioning of the electrodes, the subspace spanned by the lead vectors of the mother heart is three dimensional and orthogonal to the three-dimensional subspace generated by the lead vectors of the fetal heart transfer. Moreover, the source signals of mother and fetal are orthogonal vectors if considered over a sufficiently long time wherein the energy of the mother heart signals exceeds considerably that of the fetal heart. For all this reasons, one single SVD suffices to identify the subspace corresponding to the fetal ECG and by projecting the measurements on this subspace, the MECG can be eliminated almost completely. For more details, the reader is referred to a series of publications [14] [16] [18] [20] [22] [23] [29] [30] [43] [50]

Besides the single SVD approach for FECG/MECG separation, which is based on some restrictive source orthogonality requirements (though fulfilled under mild conditions) it is interesting to note that another SVD based method , described in [65], for the same problem,

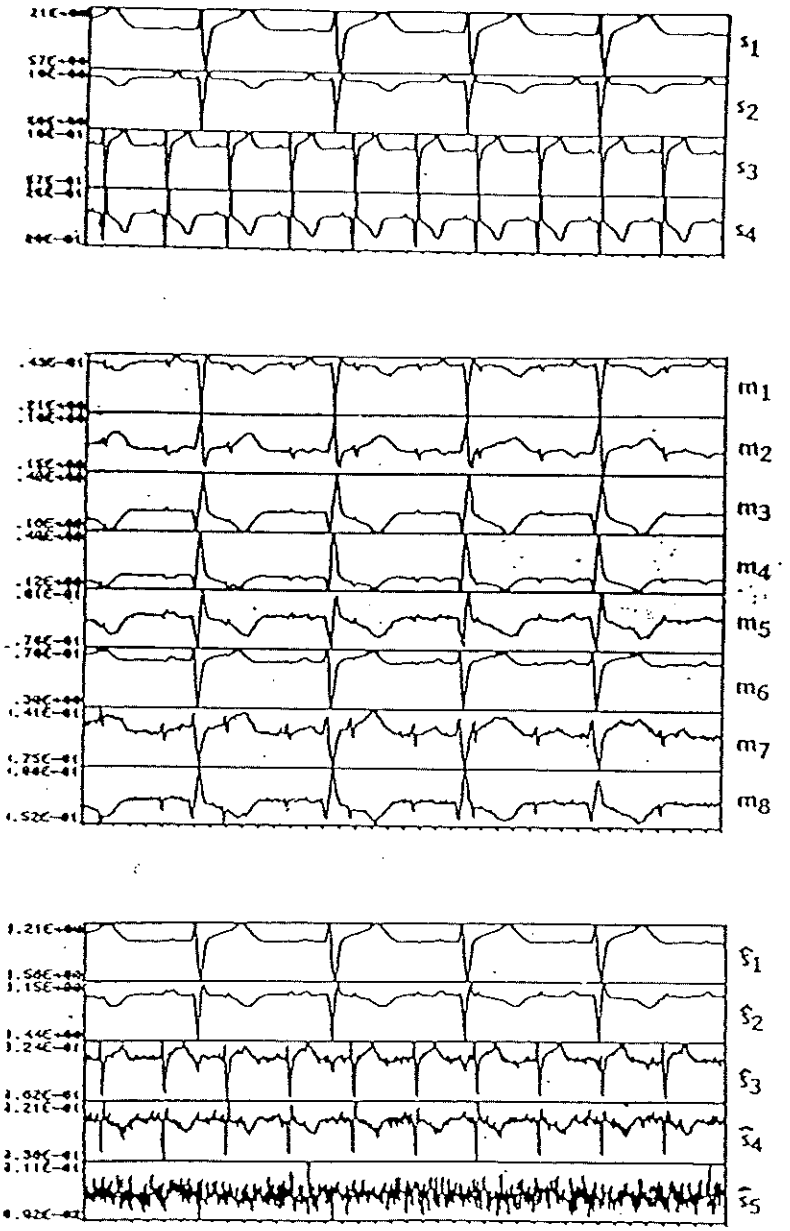

Figure 9: Simulation results for FECG separation The source signals $s_1, \ldots, s_4$ are from the mother and from the fetus. In the measurements $m_1, \ldots, m_8$, the sources appear to be mixed. One can notice the relatively weak FECG contribution. Applying SVD to the measurements matrix allows under mild condition to retrieve the subspace of maximal minimal signal-to-signal ratio in which the fetus can best be distinguished from the mother. When the measurements are projected into this subspace, a nice separation of Maternal ($\hat{s}_1$, $\hat{s}_2$) and fetal ECG ($\hat{s}_3$, $\hat{s}_4$) and noise ($\hat{s}_5$) is achieved.

lends itself very naturally to an interpretation and computation in terms of oriented signal-to-signal ratio and GSVD, as described in section 4.1. Basically, the method consists of constructing (by visual inspection) 2 matrices. The first one contains only FECG complexes while the second one is built from maternal ECG complexes. The method then reduces to the determination of that subspace in which the FECG signal can best be distinguished (from the point of view of oriented energy), from the mother ECG. This is equivalent with the determination of the maximal minimal signal-to-signal ratio subspace of a fixed dimension (for instance dimension 3 for the FECG, a choice which can be based upon a physical electromagnetic model). The measurements are then projected into this subspace, which results in a maternal ECG filtering effect.

It is interesting to note that recently high resolution subspace methods have been introduced and analysed [63]to detect the number and the location of narrowband sources . Essentially, these methods reduce to the oriented signal-to-signal ratio framework of section 4.1.

6.4 Realization and exponential fitting

The problem of realization of state space models is the following (stated here for discrete time systems) :

Given (possibly noise corrupted) Markov parameters of a linear system. Find a minimal state space representation of the form

$$x_{k+1} = A.x_k + B.u_k$$
$$y_k = C.x_k$$

As is well known, the Markov parameters satisfy $H_k = C.A^{k-1}.B$.

The problem was solved in its full generality by Ho-Kalman [55] and the SVD was introduced in its solution in [68] in 1974 and in [58] in 1978 . The algorithm is by now almost classical :

- construct the (block) Hankel matrix H of Markov parameters H_k and choose its dimensions sufficiently large.
- Factorize it using SVD : $H = U.\Sigma.V^t$ This allows to estimate the order of the system from inspection of the rank. Moreover, C and B can be read off immediately from certain (block) rows and (block) columns respectively.
- Exploit the so-called shift structure of the (block) Hankel matrix in order to estimate the state transition matrix A. This reduces to the solution of an overdetermined set of linear equations in [58], in which it differs from [68] where an additional (block) Hankel matrix is to be constructed.

The use of the SVD in these has the almost 'classical' advantages : robust rank estimation, noise insensitivity, high resolution and accuracy . This solution is applied to the estimation of amplitudes, dampings, frequencies and phases in a series of papers [1] [3] [6] [17] [40] Furthermore, it has been applied for high resolution spectral analysis in the separation and localisation of narrow-band sources , the analysis and classification of electromyograms and as a second step in a two step identification procedure, in which first the impulse response is estimated using TLLS deconvolution . These results can be found in a series of master theses at the ESAT laboratory . Several practical problems have been studied :

- the relation between the number of measurements to be used and the accuracy of the estimated paramaters (poles, ..) is studied in [48] . As an example, consider fig. 10

  It is obvious that the rank decision becomes easier and the accuracy ameliorates considerably as the number of measurements is increased.

- the behavior of the singular values of noise free Hankel matrices for first order degenerated impulse responses of the form $h_k = \alpha.k^{\mu-1}.a^k$ as a function of the pole multiplicity and the pole a is depicted in fig 11 for a two different sample sizes. More details can be found in [1] [48]

- In [17], one can find some results on the application of Hankel - SVD based methods for high resolution spectral analysis. It has been verified with some simple experiments that the method is extremely robust and performs as well as algorithms that are claimed to be optimal.

- In [40] , the sensitivity of the method is investigated for the estimation of coefficients c_i and exponents b_i from noisy observations f(t):

$$f(t) = \sum_{i=1}^{n} c_i.exp(b_i.t) + n(t)$$

  As a result, it is shown that, under mild conditions, the error in the computed exponents is of the order :

$$O(\sigma_{n+1}/(\sigma_n - \sigma_{n+1}))$$

  the quotient of the largest 'noise' singular value and the gap between the smallest 'signal' and the largest 'noise' singular value. This error estimate is much better than Kung's [58], since the constant in the order term does not contain the (rather large) norm of the pseudo-inverse of the Hankel matrix as in [58] . The result is a rigorous demonstration of some facts that are already intuitively obvious: The more the measurements are corrupted by noise, the smaller will be the gap between σ_n and σ_{n+1} and hence the less accurate will be the estimates. Moreover, there is a one-to-one relation between the exponents and the subspace that is spanned by 'corresponding' singular vectors [38] . When singular values get close, the corresponding singular vectors are no longer well conditioned and 'noise' and 'signal' subspaces get mixed. In [40] one can find also some experimental verification of the fact that 'square' Hankel matrix are best suited for the estimation of the rank (the number of exponentials).

6.5 Identification of state space models from noisy input-output measurements

Starting from the mid-seventies, the SVD and GSVD have made their appearance in the systems and control literature , where they have become the cornerstone of numerically reliable implementation of algorithms for Kalman decomposition [52] , controllability and observability questions [57] , the concept of balanced realization [66], realization theory [58] [68] and numerically reliable computation of the generalized eigenstructure of matrices [64].

The use of SVD and GSVD in identification and modeling problems starting from noisy input-output measurements of linear systems, has been and still is the subject of intensive

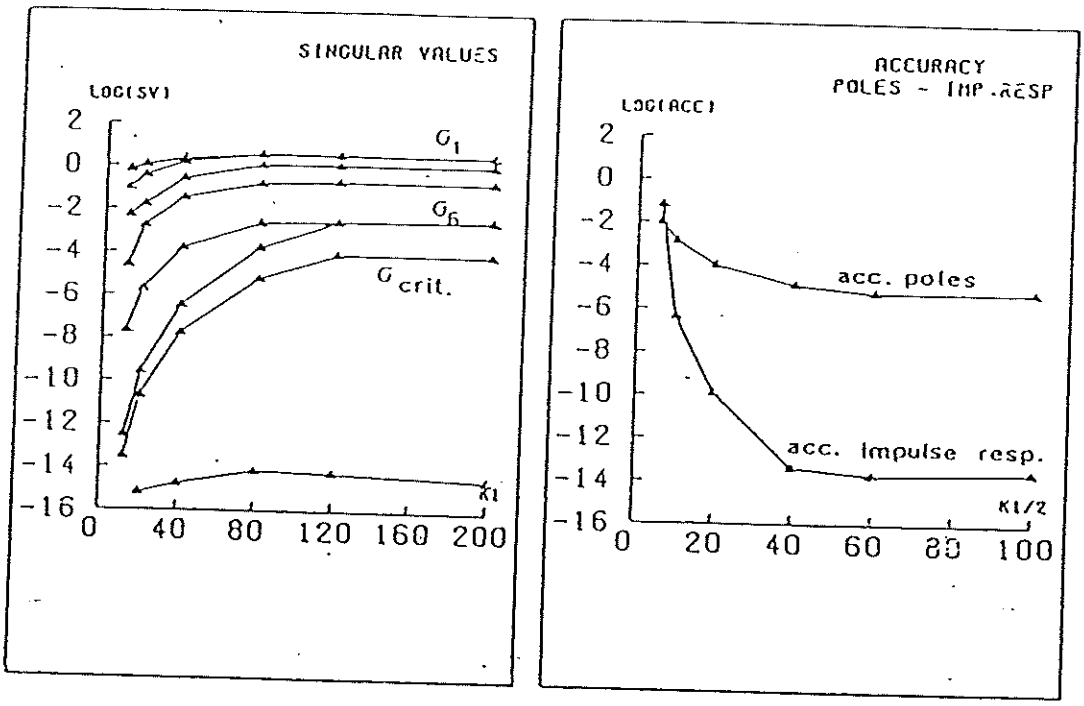

fig.10.a.

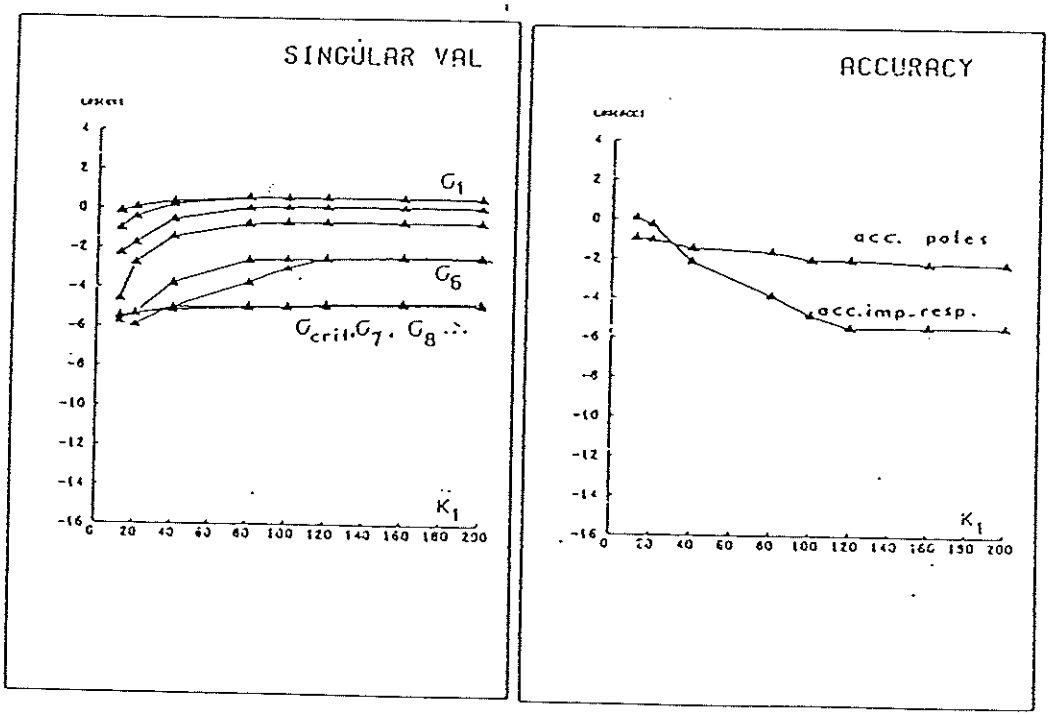

fig.10.b.

Figure 10: Influence of the number of measurements K_1 to be used on the rank decision and the acuracy of poles and realized impulse response. Fig.10.a. Exact data Fig.10.b. Noisy data : Notice that the smallest 'signal' singular value emerges from the noise from 40 used samples only.

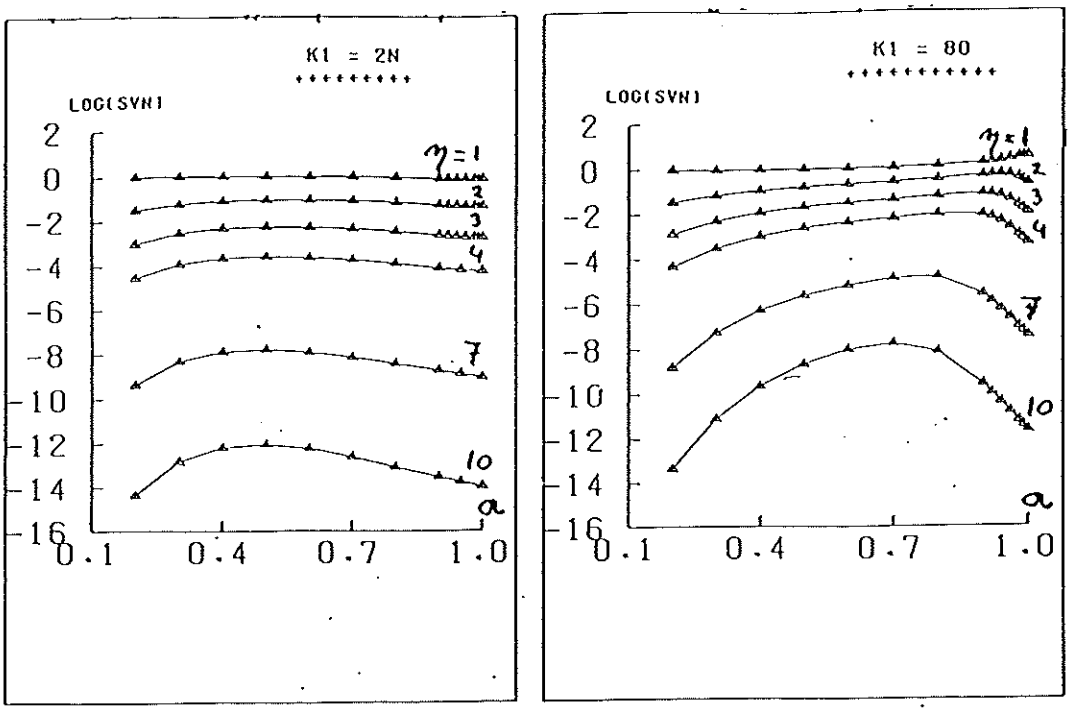

Figure 11: Smallest singular value of the Hankel matrix of Markov parameters of an elementary signal of the form $h_k = \alpha k^{\eta-1} a^k$ as a function of the multiplicity η and the damping a . α is a scaling factor that scales the signal so that it has unit energy. Fig.11.a. Minimal number of samples Fig.11.b. 80 samples

research at the ESAT laboratory. Considerable experience has been gained by applying the derived methods to several data sets from industrial processes [2] [8] [10] [11] [24] [25] [31] [34] [37] [41] [48] [51].

The first results concerned a 'brute force' approach, that consisted of a deconvolution algorithm (TLLS) for the computation of the multivariable impulse response. This was then realized into a state space model using an SVD based realization algorithm. Finally, a certain tail correction iteration procedure was applied in order to ameliorate the estimates. Results are reported in ref. [1] - [6] , [8] [10] [11] [48] . Some examples of the deconvolution of abruptly changing linear systems, are shown in fig 13.

However, in [31] [37] [41] [46] [51], a fundamental structured matrix input - output equation is derived, which provides a much more elegant framework for the formulation and solution of the multivariable identification problem. Moreover, the new approach fits perfectly well into the conceptual framework of oriented signal-to-signal ratio (section 4.1) [45] and canonical correlation analysis.

If a linear system , with m inputs and l outputs is described by the state space equations:

$$x_{k+1} = A.x_k + B.u_k$$
$$y_k = C.x_k + D.u_k$$

then by straightforward substitutions, the following input-output matrix equation can be derived:

$$Y_h = \Gamma_i.X + H.U_h$$

Here, Y_h (U_h) is a block Hankel matrix with block dimensions $i \times j$, containing i+j-1 consecutive output (input) vectors . There are several good reasons to choose these dimensions

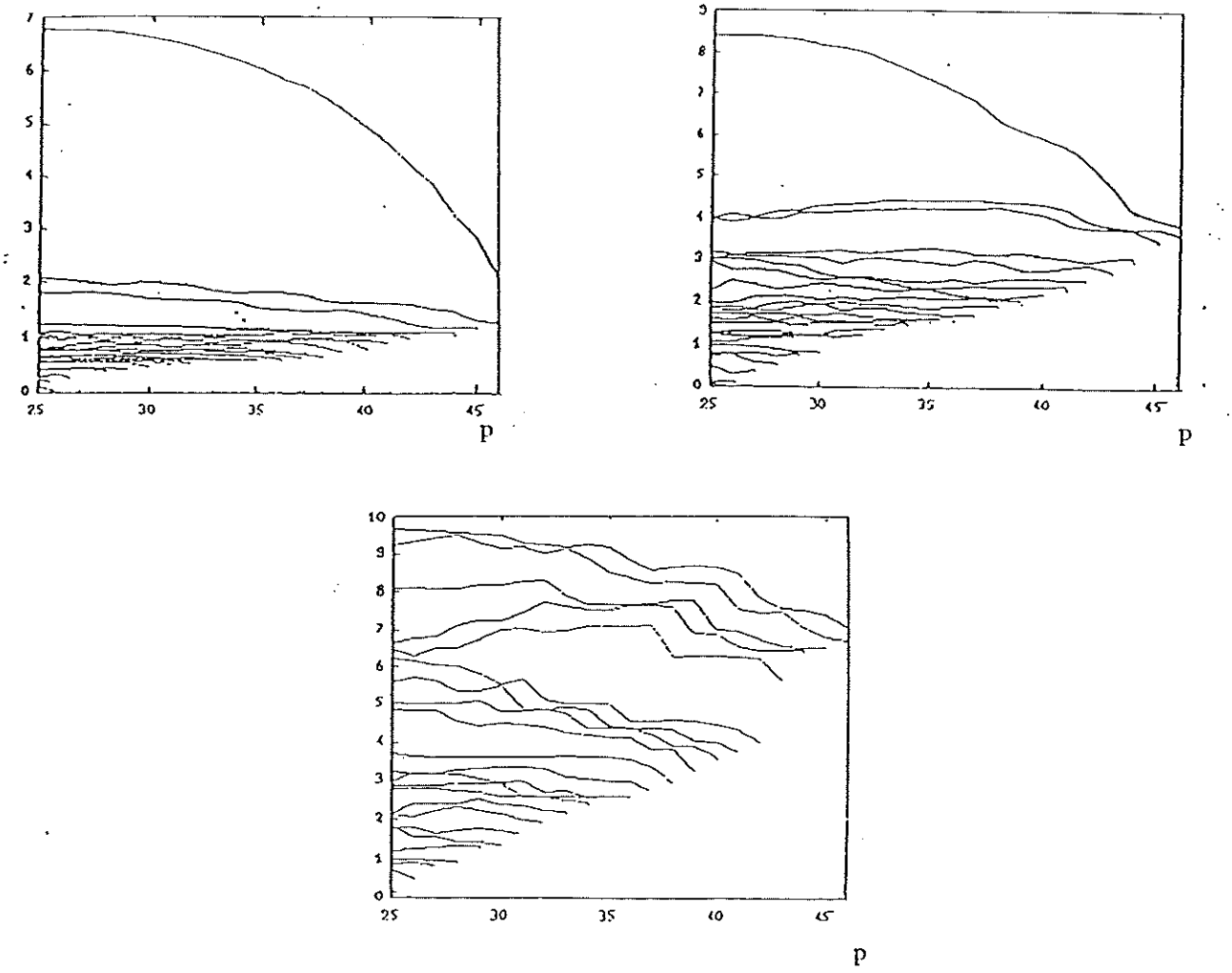

Figure 12: Singular values (except largest 2) as a function of p of a $p \times (50 - p)$ Hankel matrix of a noisy signal consisting of 3 real exponentials and noise. The results with 3 different increasing noise levels are shown. The more rectangular the Hankel matrices are, the more difficult it is to distinguish the smallest 'signal' singular value (which is the largest shown) from the largest 'noise' singular value (the second shown).

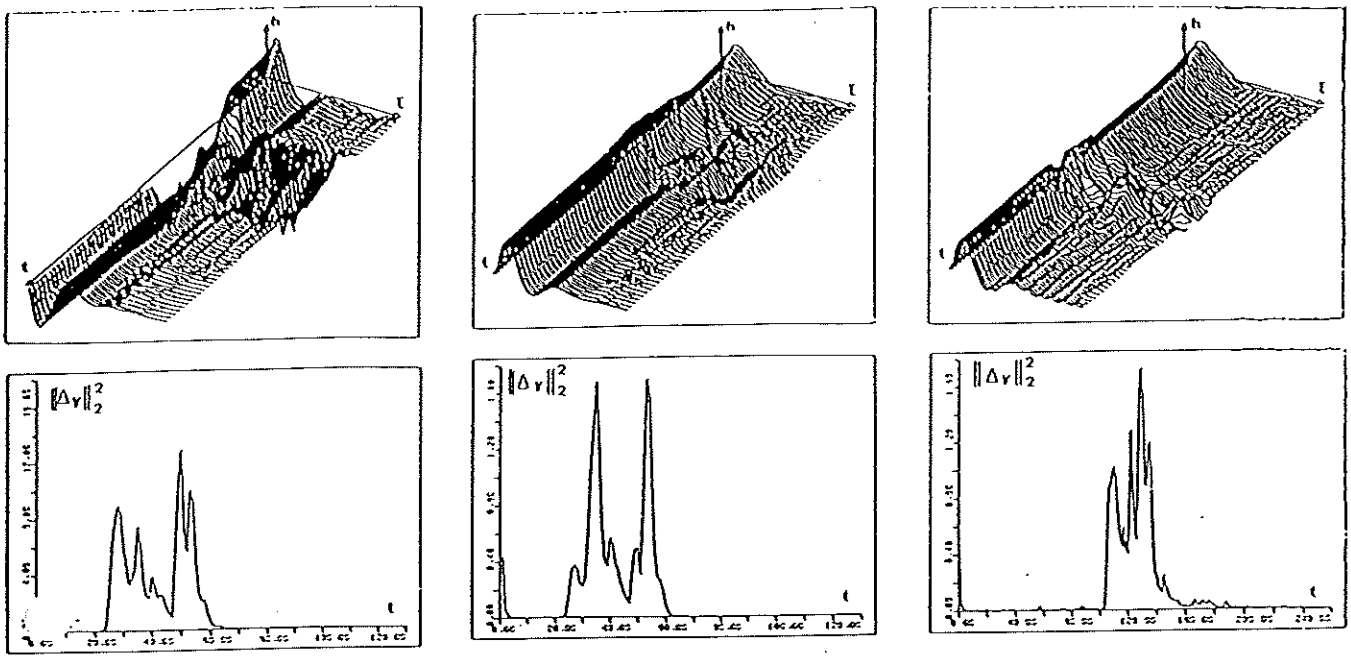

Figure 13: Adaptive identification of the impulse response of an abruptly changing system: Fig.13.a. Inversion of the impulse response; Fig.13.b. Doubled eigenfrequency; Fig.13.c. Same as 13.b. but with 5 % noise added .

in such a way that $max(li, mi) \ll j$. Γ_i is the extended observability matrix. X contains j consecutive state space vectors and H is a lower triangular block Toeplitz matrix containing Markov parameters [31] . In a realistic identification environment, only (possibly noise corrupted) input-output observations are available. Hence, only the matrices Y_h and U_h are known, up to additive noise. In a series of papers [31] [37] [41] [46] , the following results have been obtained from the geometrical representation of this input output equation :

$$rank \begin{bmatrix} Y_h \\ U_h \end{bmatrix} = rank(U_h) + n$$

where n is the dimension of the (excited) observable part of state space. Hence, under mild conditions [31] [41] , one can estimate n from the singular value decomposition of the concatenation of the input and output block Hankel matrix.

- The singular values of U_h serve as quantitative measures for the degree of persistency of excitation of the input sequence. Loosely speaking, the input sequence has to be persistently exciting in order to 'excite' all modes of the systems. When the matrix U_h is (nearly) rank deficient (some singular values are small) the input sequence is 'poor' in that it (almost) consists of a finite number of complex exponentials. When the singular values are all (almost) equal, the input sequence tends to be 'white' noise . Also for an impulsive input, the singular values are all equal (SISO). For more details of the relation between singular values and suitable input sequences, the reader is referred to [48] [51]

Three different identification approaches can now be derived from this input-output equation : a linear least squares (section 6.6.1) and a total linear least squares (section 6.6.2) approach and finally a canonical correlation approach (section 6.6.3).

6.5.1 Linear least squares identification

Let $U^{\perp}$ be any $j \times (mi - rank(U_h))$ matrix satisfying $U_h.U^{\perp} = 0$. Consider the SVD of $Y_h.U^{\perp} = P.S.Q^t$. Under mild conditions [31], rank(S) = n and there exists a non-singular $n \times n$ matrix R such that:

$$P = \Gamma_i.R$$

This implies that a realization of the state transition matrix and the output matrix of the form $R^{-1}.A.R, C.R$ can be performed in a similar way as in Kung's realization algorithm. The matrices $R^{-1}.B$ and D follow from a set of linear equations [31] [37]. In [31] [51] it is shown that this identification approach corresponds to a linear least squares version for identification problems where the input is noisefree while the output is noisy. The row space of the output block Hankel matrix is orthogonalized with respect to the input block Hankel row space . Examples can be found in [31] [37] [41] [51]

6.5.2 Total linear least squares identification

Let the SVD of

$$\begin{bmatrix} Y_h \\ U_h \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} S_1 & 0 \\ 0 & 0 \end{bmatrix} Q^t$$

where $rank(S_1) = rank(U_h) + n$ and the partitioning of the left singular matrix is such that P_{11} is a $(li) \times (mi + n)$ matrix. Then, there exists a non-singular nxn matrix T such that :

$$P_{11}.P_{21}^{\perp} = \Gamma_i.T$$

where $P_{21}^{\perp}$ is any $(mi+n) \times n$ matrix satisfying $P_{21}.P_{21}^{\perp} = 0$. This implies that a realization of the state transition matrix and the output matrix of the form $T^{-1}.A.T, C.T$ can be performed in a similar way as in Kung's realization algorithm. The matrices $T^{-1}.B$ and D follow from a set of linear equations [31][37] Contrary to the previous versions, this corresponds to a total linear least squares approximation of the multivariable identification problem , which applies when both input and output are corrupted by the same amount of noise. Considerable insight has been gained into the behavior of the algorithm in noisy industrial applications. More details are found in [31] [37] [41] [51].

6.5.3 Canonical correlation identification

The canonical correlation approach to the identification of a state space model, is based upon the following fundamental observation [46] [51]:

Let Y_1, U_1 be a output - input block Hankel pair (block dimensions $i \times j$) containing output- input measurements on a linear dynamical system up to time k and let Y_2, U_2 be another output input block Hankel pair of block dimensions $i \times j$, containing measurements from time k+1 on. If the rows of the matrix Z (with j columns) form a basis for the intersection of the row spaces of

$$\begin{bmatrix} Y_1 \\ U_1 \end{bmatrix} \quad and \quad \begin{bmatrix} Y_2 \\ U_2 \end{bmatrix}$$

then:

- $dim(row\ space\ Z) = rank(Z) = n$

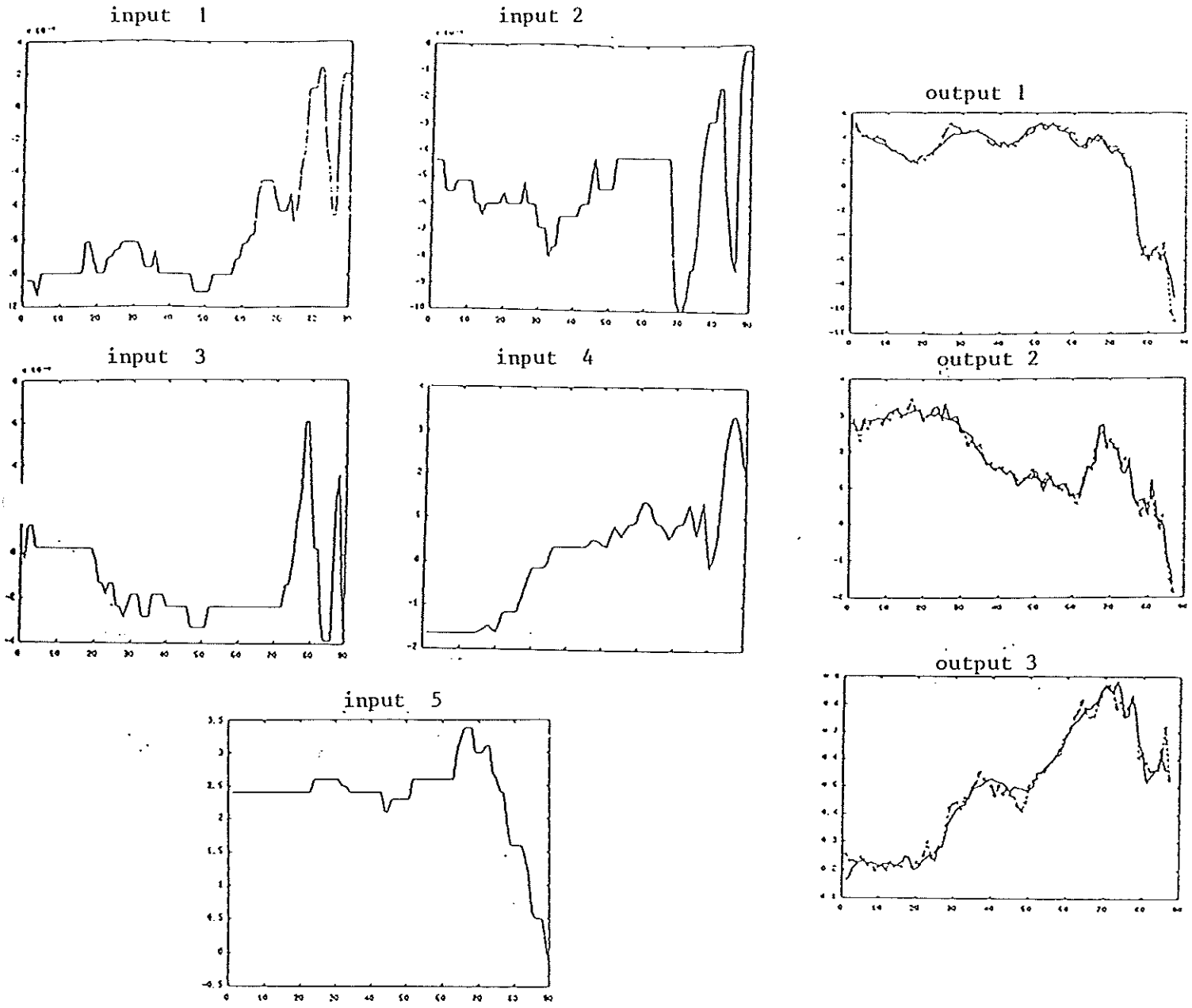

Figure 14: Canonical correlation identification of a chemical destillation column with 5 inputs and 3 outputs. Simulation with the derived 6th order state space model is shown.

- there exists a non-singular $n \times n$ matrix R such that

$$Z = R.[x_{k+1} \;\; x_{k+2} \ldots x_{k+j}]$$

Hence, the matrix Z is nothing but a state vector sequence realization. Once such a sequence is available, the model matrices A, B ,C ,D follow from the set of linear equations :

$$\begin{bmatrix} x_{k+1} \\ y_k \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x_k \\ u_k \end{bmatrix}$$

that can be solved with TLLS or one of its variations.

Hence, the 'difficult' problem of identification of a linear state space model has now been reduced to 2 SVD steps, that may be implemented in a very streamlined identification algorithm.Adaptive versions for updating and downdating the QR- and SVD factorizations via a gliding window appraoch are actually being implemented, taking into account the specific structure of the matrices. For more detail, the reader may wish to consult [46] [51]

The above summarized identification algorithms have been tested with success on a lot of industrial processes including glass furnaces, power plants, chemical reactors, biological systems, heating and ventilation of confined spaces, and the identification of a flexible arm. These are described in [24][25][31][34][37][41][46][51]. An example of the identification of a chemical destillation column with 5 inputs (left) and three outputs (right) is depicted in fig 14.

7 Conclusions

In this paper, a survey was presented of typical engineering applications of the singular value decomposition and the generalized singular value decomposition. Several examples were presented or referred to. The benefits of using the (generalized) singular value decomposition are most pronounced in those applications:

- where essentially rank decisions and the computation of the corresponding subspaces determine the complexity and parameters of the model
- where numerical reliability is of primordial importance and the potential loss of numerical accuracy (as the squaring of matrices) is to be avoided.
- where a conceptual framework, such as the notion of oriented signal-to-signal ratio, may provide unrevealed additional insight, such as in factor-analysis-like problems.
- where the problem can be stated in terms of the (generalized) singular value decomposition, which leads immediately to a reliable and robust solution, such as in a canonical correlation analysis environment.
- where robustness analysis, conditioning and sensitivity optimization are crucial, linked together with geometrical insight and interpretation , for which the GSVD may provide meaningful quantifications. (condition numbers, principal angles,...)

Moreover, in most engineering applications the number of measurements or the data acquisition poses only minor organisational problems (although the design of a measurement set up causes considerable efforts). The cost of the sensors however increases with higher accuracy and signal-to-noise requirements. In this environment the (generalized) singular value decomposition is the optimal bridge between limited measurement precision and robust modeling.

As to the computational requirements, the SVD of large matrices poses no considerable difficulties when employing a mainframe computer (matrix size order of magnitude a few hundred). Moderately sized SVD's (order of magnitude 50..70) are nowadays feasible on mini-computers and PC's. However, it can be expected that the intensive on-going research for parallellized and vectorized algorithms may result in real fast SVD solvers, possibly exploiting the matrix structure which is present in a lot of engineering applications by so-called displacement rank concepts.

Acknowlegement : The authors greatly appreciated the efforts of Dr. Shaohua Tan and Lieven Vandenberghe in typesetting the manuscript, and the stimulating interaction between all the members of the research team mentioned in the introduction.

References

[1] Staar J., Vandewalle J., "Comparison of multivariable MBH realization algorithms in the presence of multiple poles and noise disturbing the Markov sequence",*4th Int. Conf. on Anal. and Optimization of Systems.* A.Bensoussan and J.L.Lions (Ed.), Springer, pp.141-160, Berlin, 1980.

[2] Staar J., Vandewalle J., Wemans M., "Realization of measured multivariable input-output sequences with prescribed uncertainty." *Proc. of the 1st IASTED Intern. Symposium Modeling Identification and Control*, Davos, Acta Press, pp.62-66, 1981.

[3] Staar J., "A study of MBH realization algorithms", *Automatica*, Vol.17, no.3, pp.523-534, May 1981.

[4] Staar J., Vandewalle J., Wemans M., "Reliable modeling algorithms for adaptive controllers", *Proc. 13th CIRP Int. Seminar on manufacturing systems*(Faculty press Ljubljana, Yugoslavia), pp.119-134, Leuven, May 1981.

[5] Wemans M., Staar J., Vandewalle J., "ADAPT : an interactive program for the analysis and design of adaptive controllers", *Proc. 13th CIRP Int. Seminar on Manufacturing systems.* (Faculty Press Ljubljana, Yugoslavia) pp.135-150, May 1981.

[6] Staar J., Vandewalle J., Wemans M.,"A general class of numerically reliable algorithms for the realization of truncated impulse response", *Preprints IFAC World Congress*, Kyoto, Vol.1, pp.13-18, 1981.

[7] Staar J., Vandewalle J., "Numerical implications of the choice of a reference node in the nodal analysis of large circuits", *Int. J. of Circuit Theory and Applications*, Vol.9,pp.488-492,1981.

[8] Staar J. , Vandewalle J., "On the feasibility of model identification", *Proc. ECCTD*, pp.832-837, The Hague, Aug. 1981.

[9] Staar J., Vandewalle J., "Singular value decomposition: A reliable tool in the algorithmic analysis of systems", *Journal A*, Vol.23, pp.69-74, 1982.

[10] Vandewalle J., Staar J., "Modeling of linear systems : critical examples, problems of numerically reliable approaches", *Proc. IEEE Int. Symp. on Circuits and Systems*, ISCAS-82, Rome , pp.915-918, 1982.

[11] Staar J., Vandewalle J., Claeys F., De Bock H., "On the recursive identification of multivariable state space models", *Symposium on Application of multivariable systems theory*, The Institute of Measurements and Control, Plymouth , Great Britain, 26-28 October 1982, pp.49-57 .

[12] Staar J., Vandewalle J., Van Den Meersch E., "An efficient numerically stable algorithm for the computer aided analysis of circuits and systems",*Proc. IEEE Conf. ISCAS 1983*, Newport Beach, May 1983, pp.218-221.

[13] Van Huffel S., Vandewalle J., Staar J., "The total linear least squares problem : formulation, algorithm and applications", *Proc. IEEE ISCAS Conf. Montreal*, May 1984, pp.328-331.

[14] Vanderschoot J., Vandewalle J., Janssens J., Sansen W., Vantrappen G., "Extraction of weak bioelectrical signals by means of singular value decomposition", *Proc. of Sixth Int. Conf. on An. and Opt. of Systems*, Nice, June 1984, Springer Verlag, Berlin, pp.434-448.

[15] Vandewalle J., Staar J., De Moor B., Lauwers J., "An adaptive singular value decomposition algorithm and its application to adaptive realization", *Proc. of Sixth Int. Conf. on An. and Optimization of Systems*, Nice, June 1984, Springer Verlag, Berlin, pp.33-47.

[16] Vandewalle J., Vanderschoot J., De Moor B., "Source separation by adaptive singular value decomposition", *Proc. IEEE ISCAS Conf.*, Kyoto 5-7 June 1985, pp.1351 - 1354 .

[17] De Moor B., Vandewalle J., "A numerically reliable algorithm for fitting a sum of exponentials or sinusoids to noisy data", *Proc. IFAC Symp. on Comp. Aided Control Systems Design.* Copenhague, CADCE '85, pp.434-439., June 1985.

[18] Vanderschoot J., Vantrappen G., Janssens J., Sansen W., Vandewalle J., Callaerts D., "The use of singular value decomposition in multilead abdominal FECG. Processing of non-invasive signals from the fetal heart", *Europ. J.Obstet. Gynec. repr.Biol*, Vol.20, pp.74-75, 1985.

[19] Van Huffel S., Vandewalle J., "The use of total least squares techniques for identification and parameter estimation", *7th IFAC Conference on Identification and Parameter Estimation.* York 3-7 July 1985. Vol.2, pp.1167-1172.

[20] Vandewalle J., Vanderschoot J, De Moor B., "Filtering of vector signals based on the singular value decomposition", *7th European Conference on Circuit Theory and Design.* Prague 2-6 September 1985, pp.458-461.

[21] De Moor B., Vandewalle J., "An adaptive singular value decomposition algorithm based on generalized Chebyshev recursions", *Proc. of the Conference on Mathematics in Signal Processing.* 17-19 September 1985, University of Bath, UK.

[22] Callaerts D., Vanderschoot J., Vandewalle J., Sansen W., Vantrappen G., Janssens J., "An adaptive on-line method for the extraction of the complete fetal electrocardiogram from abdominal multilead recording", *Journal of Perinatal Medicine*, Vol.14, pp.421-433, 1986.

[23] Callaerts D., Vanderschoot J., Vandewalle J., Sansen W., "An on-line adaptive algorithm for signal processing applications using SVD", *EUSIPCO-86*, The Hague, 2-5 September 1986, Signal Processing III : Theories and Applications Eurasip, pp.953-956.

[24] Berckmans D., De Moor B., Vandewalle J., Goedseels V., "Modeling the design of ventilation controllers for pig houses", *Proc. of the Int. Conf. on Pig, Rabbit and Small Bird Species Housing CIGR*, Sept 8-11 1986 Section 2, Rennes, France (3 pp.).

[25] Berckmans D., De Moor B., Vandewalle J., Goedseels V., "Modeling strategy for performance estimation of control systems in confined spaces", *Second International Conference on System Simulation in Buildings*, Liege , Belgium, Dec. 1-3, Section 6.3., 1986 (16 pp.).

[26] De Moor B., Vandewalle J., "A geometrical approach to the maximal corank problem in the analysis of linear relations", *Proc. of the 25th IEEE Conf. Decision and Control*, Athens, Greece, December 1986.

[27] De Moor B., Vandewalle J., "The uniqueness versus the non-uniqueness principle in the identification of linear relations from noisy data", *Proc. of the 25th IEEE Conf. on Decision and Control*, Athens, Greece, December 1986.

[28] Van Huffel S., Vandewalle J., De Roo M., Willems J., "Reliable and efficient deconvolution technique based on total least squares for calculating the renal retention function", *Medical and Biological Engineering and Computing.* Vol. 25, pp.26-33 January 1987.

[29] Vanderschoot Jan, Callaerts Dirk, Sansen W., Vandewalle J., Vantrappen G., Janssens J., "Two methods for optimal MECG elimination and FECG detection from skin electrode signals", *IEEE Transactions on Biomedical Engineering*, Vol. BME-34, no.3, March 1987.

[30] Callaerts D., Vanderschoot J., Vandewalle J., Sansen W., Vantrappen G., Janssens J., Van Cutsem E., "Fetal electrocardiogram measuring method and equipment (FEMME)", *Proc. First World Symposium on Computers in the Care of Mother, Fetus and Newborn.* Vienna, Austria 8-11 March 1987, J.of Perinatal Medicine, Vol.15 (suppl1), p.33, 1987.

[31] De Moor B., Vandewalle J., "A geometrical strategy for the identification of state space models of linear multivariable systems with singular value decomposition", *Proc. of the 3rd Int.Symp. on Appl. of Multivariable System Techniques*, Inst. of Measurements and Control, pp.59-68, 1987.

[32] Van Huffel S. , Vandewalle J., "Subset selection using the total least squares approach in Collinearity problems with errors in the variables", *Lin. Alg. and its Applications*, Vol.88/89, pp.695-714, April 1987.

[33] De Moor B., Vandewalle J., "The uncertainty principle of mathematical modeling", *8th Symposium of Information Theory in the Benelux*, Deventer, The Netherlands, 21-22 May 1987.

[34] Berckmans D., De Moor B., Vandewalle J., Goedseels V., "On line identification and control of heating, ventilating and air-conditioning in livestock buildings," *Workshop on Identification Techniques in Practice : Possibilities and realizations,* Belgisch Instituut voor Regeltechniek en Automatisatie, Leuven, 27 May 1987.

[35] Van Huffel S., Vandewalle J., "Algebraic relationships between classical regression and total least squares estimation", *Lin. Alg and its Applications*, Vol. 93, pp.149-160, August 1987.

[36] Van Huffel S., Vandewalle J., Haegemans A., "An efficient and reliable algorithm for computing the singular subspace of a matrix, associated with its smallest singular values", *J.of Comput. and Applied Mathematics*, Vol.19,pp.313-330, 1987.

[37] Swevers J., Adams M., De Moor B., "A new direct time domain identification method for linear systems using singular value decomposition", *Proc. of the 12th International Seminar on Modal Analysis*, Katholieke Universiteit Leuven, Leuven 21-24 September 1987.

[38] De Moor B., Vandewalle J., "Non-conventional matrix calculus in the analysis of rank deficient Hankel matrices of finite dimensions", *Systems and Control Letters*, 1987 (in Press).

[39] Van Huffel S., Vandewalle J., "The application of the total least squares technique in linear parameter estimation", *Journal A*, 1987 (in Press).

[40] De Groen P., De Moor B., "The fit of a sum of exponential to noisy data", *Journal of Computational and Applied Mathematics*, 1987 (In Press) .

[41] De Moor B., Vandewalle J., Moonen M., Van Mieghem P., Vandenberghe L., "A geometrical strategy for the identification of state space models of linear multivariable systems with singular value decomposition", (subm. IFAC Identification Beijing 1988).

[42] De Moor B., Vandewalle J., "A unifying theorem for linear and total linear least squares identification schemes", Submitted for publication, June, 1987.

[43] Callaerts D., Vandewalle J., Sansen W., Moonen M., "On-line algorithm for signal separation based on SVD", This volume.

[44] Van Huffel S., Vandewalle J., "The total least squares technique : Computation, Properties and applications", This volume.

[45] De Moor B., Staar J., Vandewalle J., "The concept of oriented signal-to-signal ratio and the generalized singular value decomposition", This volume.

[46] De Moor B., Moonen M., Vandenberghe L., Vandewalle J., "Identification of linear state space models with singular value decomposition using canonical correlation concepts", This volume.

[47] De Moor B., Moonen M., Vandenberghe L., Vandewalle J., "Identification of linear state space models with singular value decomposition using canonical correlation", To appear in *Proc. International Conf. on Acoustics, Speech and Signal Processing*, April 1988, New York.

[48] Staar J., "Concepts for reliable modeling of linear systems with application to on-line identification of multivariable state space descriptions", Doct. Diss., ESAT, Electrical Engineering Dept., Kath.Univ.Leuven, 1982.

[49] Van Huffel S., "Analysis of the total least squares problem and its use in parameter estimation", Doct. Diss., ESAT, Electrical Eng. Dept. , Kath . Univ.Leuven, 1987.

[50] Vanderschoot J., "Analysis of biomedical signals by numerically reliable techniques", expected for 1988.

[51] De Moor B. Contents of doct. diss. : Identification with SVD and GSVD : concepts and algorithms, the uncertainty principle of mathematical modeling. (expected for Spring 1988)

[52] Fernando K.V., Hammarling S.J., "A product induced singular value deomposition for two matrices and balanced realisation", *NAG Technical report TR8/87*, The Numerical Algorithms Group Limited, 1987

[53] Golub G.H., Van Loan C.F., *"Matrix Computations"*. North Oxford Academic / Johns Hopkins University Press. 1983 .

[54] Hammarling S., "The singular value decomposition in multivariate statistics", *AGM Signum Newsletter*, Vol.20, no.3, pp.2-25, July 1985.

[55] Ho B.L., Kalman R.E., "Effective construction of linear state variable models from input/output functions", *Proc. 3-rd Allerton Conf. Circuits and Systems Theory*, pp.449-459, 1965.

[56] Friedlander B., Morf M., Kailath T., Ljung L., "New inversion formulas for matrices classified in their distance from Toeplitz matrices", *Linear Algebra and its applications*, 27:31-60, 1979.

[57] Klema V.C., Laub A.J., "The singular value decomposition:its computation and some applications", *IEEE Trans. Automatic Control*, Vol.AC-25, no.2,pp.164-167, 1980.

[58] Kung S.Y., "A new identification and model reduction algorithm via singular value decomposition", *Proc. 12-th Asilomar Conf. Circ. Syst. comp.*, Nov. 1978.

[59] Lawson C.L., Hanson R.J., *"Solving least squares problems"*, Prentice Hall Series in Automatic Computation, Englewood Cliffs 1974.

[60] Moore B.C., "Principal component analysis in linear systems : Controllability, Observability and Model reduction", *IEEE Trans. Automatic Control*, Vol.AC-26, pp.17-32, Feb. 1981.

[61] Paige C.C., "Computing the generalized singular value decomposition", *SIAM J. Sci.Statist.Comput.*, 7, pp.1126-1146,1986.

[62] Parlett B.N., *"The symmetric eigenvalue problem"* Prentice Hall Series in Computational Mathematics, Prentice Hall Inc., Englewood Cliffs N.J. 1980 .

[63] Roy R.H., "Estimation of signal parameters via rotational invariance techniques", Ph.D. thesis, Stanford University, August 1987.

[64] Van Dooren P., "The generalized eigenstructure in linear systems theory", *IEEE Trans. Automatic Control*, Vol.AC-26, no.1, pp.111-130, 1981.

[65] van Oosterom A., Alsters J., "Removing the maternal component in the fetal ECG using singular value decomposition", In *Electrocardiography '83*, I.Ruttkay-Nedecky and P.MacFarlane, Eds. Amsterdam, The Netherlands: Excerpta Medica, 1984,pp.171-176.

[66] Wilkinson J.H., *"The algebraic eigenvalue problem"*, London, Oxford Univ. Press, 1965.

[67] Wonham W.M., *"Linear multivariable control : A geometric approach"*, New York, Springer 1974.

[68] Zeiger H.P., Mc Ewen A.J., "Approximate linear realizations of given dimensions via Ho's algorithm", *IEEE Trans. Automatic Control*, Vol.AC-19, pp.153, 1974.