On the Use of the 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

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

1 Introduction

In recent years, the *'ordinary' singular value decomposition* (**OSVD**) and its generalizations, have become extremely valuable instruments in the analysis and the solution of problems in mathematical engineering. In most applications, the **OSVD** 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 **OSVD** 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.

NATO ASI Series, Vol. F 70
Numerical Linear Algebra, Digital Signal Processing and Parallel Algorithms
Edited by G. H. Golub and P. Van Dooren

In a first part the perspectives of SVD in engineering are situated. In section 3 the OSVD and Quotient SVD (QSVD) are described. In section 4 several fundamental concepts are defined and their properties discussed in terms of the (generalized) singular value decomposition. 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 OSVD and one of its generalizations, the QSVD, which is also an appropriate instrument to quantify the concept of canonical correlation analysis and the notion of angles between subspaces. We present 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 section 5 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 from input-output data, etc....

As will be illustrated, the benefits of using the (generalized) singular value decomposition are most pronounced in those signal processing and identification applications:

- where 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 caused by the squaring of matrices) is to be avoided.

- where a conceptual framework, such as the notion of oriented signal-to-signal ratio, may provide 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 OSVD and its generalizations may provide meaningful quantification. (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 singular value decomposition is the optimal bridge between limited measurement precision and robust modeling.

As to the computational requirements, the OSVD of large matrices poses no considerable difficulties when employing a mainframe computer (matrix size order of magnitude a few hundred). Moderately sized OSVDs (order of magnitude 50..70) are nowadays feasible on mini-computers and PC's. In many applications only part of the OSVD is needed (e.g. the singular values, the left singular vectors, the smallest singular vector, ...) which results in a reduction of the computational burden. Moreover, it can be expected that the intensive on-going research for parallelized and vectorized algorithms may result in real fast OSVD solvers, possibly exploiting the matrix structure which is present in a lot of engineering applications by so-called displacement rank concepts.

2 Opportunities offered by the technology and constraints imposed by the applications

In recent years one can witness an enormous expansion of computational power at all levels : supercomputers, workstations, mini and microcomputers, VLSI, ...In addition the measurements power is increasing by cheap sensors, transducers and data acquisition equipment. However in practice it is much more difficult to increase the accuracy of the measurements than to increase the volume of data. Hence the need for methods and software which can extract more accurate information from the measured data is more acute even at the expense of kflops of computations. In this situation we believe that the singular value decomposition is a very important tool which allows to take profit of the expansion of the computational power in order to improve the accuracy. The **OSVD** is well known in linear algebra for its solid numerical qualities [1]. However it is only recently explored as an important concept in digital signal processing. On the other hand its widespread use is still hampered by the computational burden. Hence it is one of the important targets of intensive research on parallel algorithms. The **OSVD** is thus an important focal point in the three research areas of this Advanced Study Institute.

From an application point of view one can distinguish at least three different environments where these areas meet.

- First of all in the number crunching environment supercomputers are used in order to solve massive problems like weather forecasts, seismic data, optimizations, simulations ... The volume of applications is limited, the value of each application is quite high and the equipment is fixed (e.g. hypercube) and shared with many.

- The design and production environment on the other hand is much more widespread and includes applications like design, laboratory, medicine, plant production, optimization and control. In these situations the equipment consists of a workstation or a personal computer and it can have a mathematical or a signal coprocessor for doing computation intensive operations like OSVD.

- The third environment is the consumer environment where dedicated software hardware products are used in sound, video, automotive or telecommunication applications. The high volume of these applications imposes the dedicated nature (e.g. application specific I.C.).

In each of these three environments different trade-offs have to be made between algorithms, parallellization, and architecture (multiprocessor, systolic, bitserial, bitparallel). We are mainly concerned with the design and production and the consumer environment, where the architecture is not

fixed. With the advent of VLSI massive amounts of computational power will be available for the consumer and the professional in design and production. The question here is to exploit the computational power in order to perform a number of digital signal processing activities, command, monitoring, data compression, error correction, ... These digital signal processing activities exhibit a number of characteristics which have also an impact on the issues involved in the study of the singular value decomposition. First of all the data stream and the processing of the data should happen in real time. Moreover in recent years the sample rates are increasing from the bps to the kbps and higher rates. On the other hand many variables are continuously being measured or sensored (e.g. automatic control, medicine, ...). The accuracy range is quite different from the accuracy range usually considered in numerical algebra (single or double precision). For measured data 1% of full scale accuracy is usually considered to be very good, which corresponds with 40 dB signal to noise (S/N) ratio. Of course it cannot be tolerated that the algorithms would worsen the S/N ratio substantially. Hence the need for numerically reliable methods is even more acute than in numerical algebra. However the smaller singular triplets often are even less important here. Also it is clear that the **OSVD** is only a part, although often a very important one, of a global system. This global system includes many tasks which can be performed by hardware and software. Typical engineering trade-offs and compromises are made in order to design such systems. The volume of these applications and the dedicated nature of the tasks make these suited for VLSI implementation. It is expected that in this framework **OSVD** will play an important role, in much the same way that FFT has played an important role in digital signal processing and numerical analysis.

3 The (Quotient) Singular Value Decomposition

In this section, the theorems stating the existence and properties of the singular value decomposition are presented. For a proof and computational requirements, the reader is referred to literature [1].

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

$$U = [u_1 \ u_2 \ \cdots \ u_m], \quad V = [v_1 \ v_2 \ \cdots \ v_n]$$

such that

$$U^t \cdot A \cdot V = \begin{bmatrix} diag(\sigma_1, \sigma_2, \cdots, \sigma_r) & 0 \\ 0 & 0 \end{bmatrix}$$

where

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

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

Corollary 1 *Let the OSVD of A be given as in theorem 1 then*

(1) *Rank property*

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

(2) *Dyadic decomposition*

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

(3) *Norms*

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

(4) *Rank k approximation.*

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

then

$$\min_{r(B)=k} \| A - B \|_2 = \| A - A_k \|_2 = \sigma_{k+1}$$
$$\min_{r(B)=k} \| A - B \|_F^2 = \| A - A_k \|_F^2 = \sigma_{k+1}^2 + \cdots + \sigma_p^2$$

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 OSVD, including existence proofs, computational requirements and numerical considerations, sensitivity results, conditioning etc.. can be found in the modern bible of numerical analysis [1] and the references therein.

Theorem 2 *The quotient singular value decomposition for real matrix pairs* [1].
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

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

[1] As there are several other generalizations of the OSVD, such as the Product SVD (PSVD) and the Restricted SVD (RSVD), we propose to call this generalization the *Quotient SVD* QSVD because in most applications it are the ratios of the diagonal elements of D_A and D_B that are important. Moreover, under certain conditions, the QSVD delivers the OSVD of B^+A, which can be considered as a *matrix quotient*. For more details, the reader is referred to [45] and the references therein.

Observe that the **QSVD** reduces to the **OSVD** 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 quotient singular values of A and B. The quotient singular values corresponding to the $\beta_i = 0$ are infinite. They are considered to be equal and come first.

Only in the sixties numerically reliable methods were found to compute the **OSVD** [1]. However in many engineering applications, the general purpose **OSVD** algorithm (complete decomposition, full machine precision) can be replaced by other algorithms with less stringent specifications. This reduces the computational burden and/or storage requirements.

- 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 OSVD makes no sense.

- A third specification is the adaptive computation of the OSVD 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 is expected that structure exploiting algorithms perform better (faster) than the standard full size **OSVD** 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 algorithms.

Results on the *power method and the Chebyshev method* for the **OSVD** can be found in [21,24]. One of the important observations 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 **OSVD** power method in [21] to the Chebyshev algorithm in [24], including convergence criteria, deflation strategies, accelerations algorithms and convergence level control rules that allow to compute the **OSVD** up to the required numerical precision.
An adaptation of Golub's full size **OSVD** for the case that *only the smallest singular triplets are needed*, is described in [32]. By a careful analysis the computational requirements can be reduced in such a way that the partial **OSVD** algorithm can be 3 times faster than the classical one, while obtaining the same accuracy. Moreover, in [36] 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. Chebyshev iteration [24], 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 **OSVD** and partial **OSVD**) have also been compared with the iterative ones and several potential applications have been investigated [36] (parameter estimation, subset selection, discrete deconvolution).
Finally, the *one-sided Jacobi method* has been studied in relation to the problem of adaptive computation of the **OSVD** of a matrix containing measurements on the fetal and the maternal ECG [25,26,28,37]. As the measurements (6 to 9 channels) enter at a frequency of 250 Hz, the **OSVD** is updated with orthogonal Givens' rotations as to minimize the Frobenius-norm 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 up to 500 samples per second.

4 Fundamental geometric concepts based on the OSVD and the QSVD

In this section, we discuss several concepts based on the OSVD and QSVD that are useful in applications such as *oriented signal-to-signal ratios* (section 4.1), *canonical correlations* (section 4.2), *condition numbers* (section 4.3) and *an orthogonality principle for noise and exact data* (section 4.4).
In a wide variety of systems and signal processing applications, vector sequences are measured and analysed. Whenever *linear* models are used to describe 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 (quotient) 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 [38,1]), the identification of factor-analysis-like models (rotational invariance techniques [12]), separation of MECG/FECG [28,40,50,13], realization of linear state space models from impulse responses [15,22,35] and identification of state space models from noisy input-output measurements via canonical correlation analysis [41-43].

4.1 Oriented energy and signal-to-signal ratios

The recent introduction of the fundamental concepts of oriented energy and oriented signal - to -signal ratio [37] 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 [37] 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 quotient 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 OSVD $A = U \Sigma V^t$ *where $\Sigma = diag\{\sigma_i\}$. Then each direction of extremal oriented energy is generated by a left singular vector u_i with extremal energy equal to the corresponding singular value squared σ_i^2.*

Theorem 4 *Extremal directions of oriented signal-to-signal ratio.*
Let A and B be $m \times n$ matrices, with QSVD:

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

where the quotient 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 x_i^t of the matrix X and the corresponding extremal signal-to-signal ratio is the quotient singular value squared $(\alpha_i/\beta_i)^2$.

These two theorems are illustrated for two dimensions in figure 1. Observe that for the oriented energy the maximum and minimum corresponding to the largest resp. smallest singular vectors while a saddle point would correspond to the intermediate singular vector. 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 theorems is nothing else than the Courant-Fisher minimax characterization of the eigenvalues of symmetric operators [37,12]. 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 [37].

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, at least one

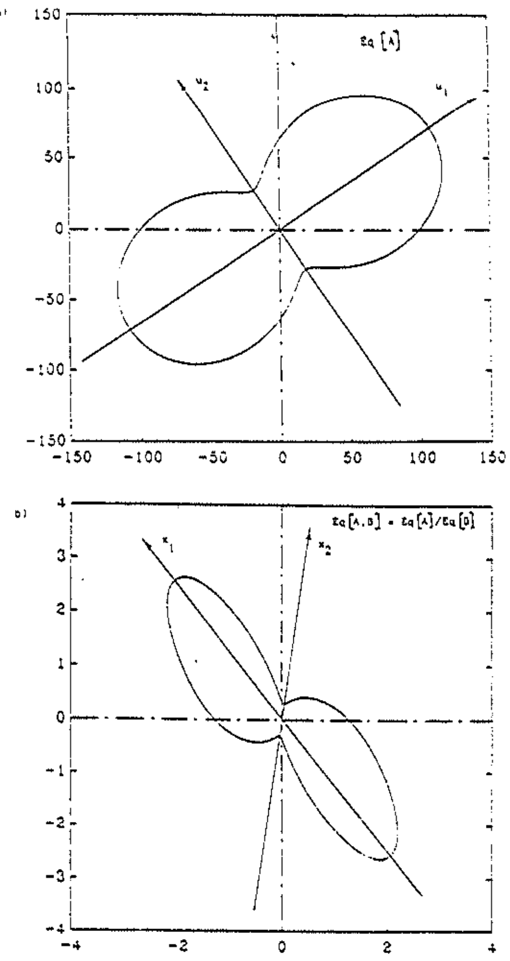

Figure 1: a) Oriented energy of a 2-vector sequence A and b) oriented signal-to-signal distribution of two 2-vector sequences A and B.

r-dimensional subspace has to exist where the worst case is better than 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 QSVD allows to find this subspace: It is the r-dimensional subspace generated by the first r rows of X, when the quotient singular values are ordered as in theorem 4.

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

- the determination of a suitable rank r provides the complexity of the model.
- 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 QSVD solution corresponds precisely to the 'classical' Mahalanobis transformation that is commonly used in statistical estimators as a kind of prewhitening filter [37].

Nice applications of the oriented signal-to-signal ratio concept include source separation techniques for fetal ECG extraction [20,13], subspace methods to detect narrowband sources [12] and system identification [41-43]. 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 5.

4.2 Canonical correlation analysis

The OSVD 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

$$\| u \| = \| v \| = 1;$$
$$u^t.u_i = 0 \quad i = 1, ..., k-1$$
$$v^t.v_i = 0 \quad i = 1, ..., k-1$$

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 [1] that:

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

where the OSVD 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 $(m \times p)$ [1, 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 [41-43].

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 quotient singular value decomposition, all of which are followed by an **OSVD** 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 preferable with respect to the others. Finally, let us observe 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 short of a (generalized) total linear least squares problem.

4.3 Condition numbers

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.

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. One can show [39] that the property of 'being usually close' to either the maximum or the minimum significantly increases with increasing condition number. The minimum 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 many 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. 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. 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)

4.4 An orthogonality principle for noisy data

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, or dynamical realization of impulse responses from (block)-Hankel matrices or the identification of dynamical state space models from noisy input-output data with canonical correlation analysis. More specifically, we are confronted

with the following situation:

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

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 OSVD 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 [44]. An illustration can be found in the Fig.2, 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 identification 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.

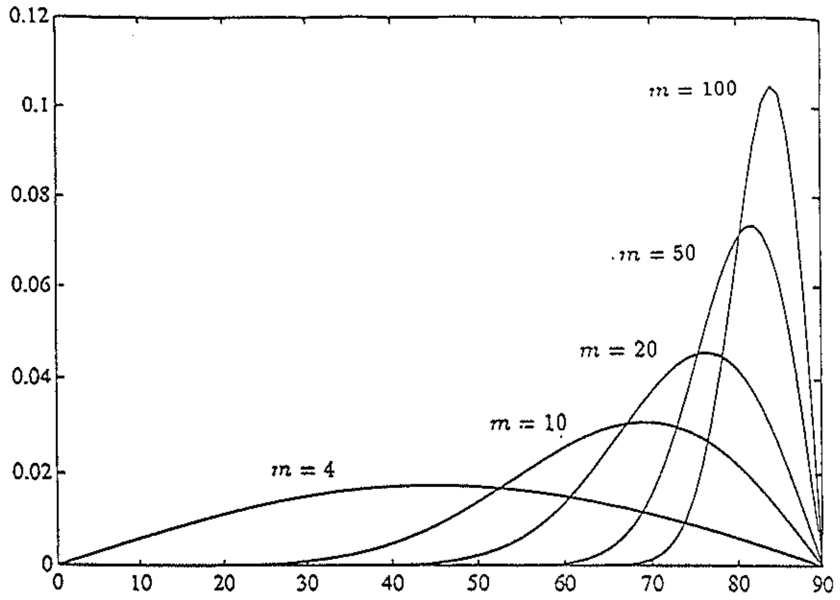

Figure 2: Probability distribution of the angle between an arbitrary direction in R^m (uniformly distributed) and a fixed 2-dimensional plane for m=4 (1), m=10 (2), m=20 (3), m=50 (4), m=100 (5). Observe the increasing probability that the direction is orthogonal to the plane.

In [41-44] 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 **OSVD** of A and C. We will derive the conditions under which the column space 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 **OSVDs**:

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

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

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:

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

Let P_1^t and P_2^t have the **OSVDs**:

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

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 reordering 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:

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

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

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

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 OSVD of C will approach the following 'limit' OSVD:

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

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 $E(B.B^t) = n\sigma^2.I_m$. Moreover, 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 singular values of C: As is obvious, a 'noise threshold' $\sqrt{n}\sigma$ will be recognized in the singular values (for sufficiently large overdetermination n/m).

For a detailed account of the application of this orthogonality principle in system identification, the interested reader is referred to [44].

5 Applications of the Quotient Singular Value Decomposition

In this section, we discuss some applications of the (quotient) singular value decomposition that are closely related to some of its well known algebraic, geometric and numerical properties. We have included a mechanical example (section 5.1), a biomedical signal processing application (section 5.2) and a summary of results on the application of the OSVD in realization theory and system identification (sections 5.3 and 5.4). Other applications can be found in the references.

5.1 Moments of inertia

First the close connection between the OSVD and the Principal Axes and Moments of Inertia (PAMI) of a rigid body is established. Only the discretized case will be considered. Consider a rigid body consisting of K point masses $m(k)$, described 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:

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

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

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

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

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 [39].

However the 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.

Theorem 5 *On the relation between* OSVD *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 an 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 OSVD *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:*

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

The interested reader is encouraged to investigate the benefits of this relation between OSVD and the computation of the PAMI of a rigid body, by applying this result to the matrix M

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

Computing the **OSVD** of M in ϵ machine precision, delivers now the second singular value μ with a precision of ϵ/μ, whereas the computation via the tensor T only has precision ϵ/μ^2.

Corollary 2 *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 OSVD *of* $M = U \cdot \Sigma \cdot V^t$.

5.2 Fetal ECG extraction

The measurements of this biomedical application 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 [37]) could be applied. However, it has been verified [20,40] that for this specific application with an appropriate position 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 contribution of the mother heart is much stronger than that of the fetal heart. For all this reasons, one single **OSVD** 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 on this separation based on the strength of the signals we refer to [20,25,26,28,40].

Besides the single **OSVD** 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 **OSVD** based method, described in [13], for the same problem, lends itself very naturally to an interpretation and computation in terms of oriented signal-to-signal ratio and **QSVD**. 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 electro-magnetic model). The measurements are then projected into this subspace, which results in a maternal ECG filtering effect. This is a source separation based on the relative strength.

It is interesting to note that recently high resolution subspace methods have been introduced and analysed [12] to detect the number and the location of narrowband sources. Essentially, these methods reduce to the oriented signal-to-signal ratio framework.

5.3 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 H_k of a linear system. Find a minimal state space representation of the form

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

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 and the OSVD was introduced in its solution in [9,14]. 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 **OSVD**: $H = U.\Sigma.V^t$ 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 [9], in which it differs from [14] where an additional (block) Hankel matrix is to be constructed.

The use of the **OSVD** 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 [15,22,35]. 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. Several practical problems have been studied:

- the relation between the number of measurements to be used and the accuracy of the estimated parameters (poles, ..). It is obvious that the rank decision becomes easier and the accuracy improves considerably as the number of measurements is increased.

- In [22], one can find some results on the application of Hankel - **OSVD** 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 [35], 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 [9], since the constant in the order term does not contain the (rather large) norm of the pseudo-inverse of the Hankel matrix as in [9]. 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 [34]. When singular values get close, the corresponding singular vectors are no longer well conditioned and 'noise' and 'signal' subspaces get mixed. In [35] 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).

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

Starting from the mid-seventies, the **OSVD** and to a lesser extent, the **QSVD** 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 [7], controllability and observability questions [2], the concept of balanced realization, realization theory [9,14] and numerically reliable computation of the generalized eigenstructure of matrices [5].

The use of OSVD and QSVD in identification and modeling problems starting from noisy input-output measurements of linear systems, has been and still is the subject of intensive research at the ESAT laboratory. Considerable experience has been gained by applying the derived methods to several data sets from industrial processes. The first results concerned a 'brute force' approach [15,18,19] 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 OSVD based realization algorithm. Finally, a certain tail correction iteration procedure was applied in order to ameliorate the estimates. However, in [29,33,41,44], 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 and canonical correlation analysis.

If a linear system, with m inputs and l outputs is described by the above mentioned state space equations: 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 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 [29]. In a realistic identification environment, only 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 [29,41-44] 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 [29], 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).

Three different identification approaches can now be derived from this input-output equation: a linear least squares approach, a total linear least squares approach and a canonical correlation approach.

Linear Least Squares:

Let $U^{\perp}$ be any $j \times (mi - rank(U_h))$ matrix satisfying $U_h.U^{\perp} = 0$. Consider the OSVD of $Y_h.U^{\perp} = P.S.Q^t$. Under mild conditions [29], 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 [29,33]. 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.

Total Linear Least Squares:

Let the OSVD 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 [29,33]. 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 [41-44].

Canonical Correlations:

The canonical correlation approach to the identification of a state space model, is based upon the following fundamental observation [41-44]:

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:

- $dim(row\ space\ Z) = rank(Z) = n$
- 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 OSVD steps, that may be implemented in a very streamlined identification algorithm. Adaptive versions for updating and downdating the QR- and OSVD factorizations via a gliding window approach are actually being implemented, taking into account the specific structure of the matrices. For more detail, the reader may wish to consult [41-44].

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, the results of which can be found in [29] [41] [42] [43] [44].

6 Conclusions

In this paper, it is claimed that the singular value decomposition and the quotient singular value decomposition have a great potential for signal processing in much the same way as the FFT had a great impact on digital signal processing in the seventies and eighties. Several applications were presented or referred to. The benefits of using the (quotient) 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 crucial importance and the potential loss of numerical accuracy 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 directly in terms of the (quotient) 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 **OSVD** and the **QSVD** 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 (quotient) singular value decomposition is the optimal bridge between limited measurement precision and robust modeling.
As to the computational requirements, the OSVD of large matrices poses no considerable difficulties when employing a mainframe computer (matrix size order of magnitude a few hundred). Moderately sized OSVDs (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 OSVD solvers, possibly exploiting the matrix structure which is present in many engineering applications.

Acknowledgement: The authors greatly appreciate the stimulating interaction between all the members of the research team at ESAT.

References

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

[2] 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.

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

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

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

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

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

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

[9] 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.

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

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

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

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] De Moor B., Vandewalle J., "An adaptive singular value decomposition algorithm based on generalized Chebyshev recursions", in: Mathematics in signal processing, T.S. Durrani, J.B. Abbiss, J.E. Hudson, R.N. Madan, J.G. McWhirter and T.A. Moore (ed.), Clarendon Press, Oxfordm 1987, pp. 607-635.

[25] 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.

[26] Callaerts D., Vanderschoot J., Vandewalle J., Sansen W., "An online 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] 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, Vol. 9, pp. 401-410.

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

[36] Van Huffel, S., Vandewalle, J., "The partial Total Least Squares Algorithm", *Journal Comput. and Applied Math.*, Vol. 21, pp. 333-341, 1988.

[37] De Moor, B., Staar, J., Vandewalle, J., "Oriented energy and oriented signal-to-signal ratio concepts in the analysis of vector sequences and time series", in *SVD and Signal Processing* E. Deprettere (ed.), North Holland, 1988, pp. 209-232.

[38] Vanhuffel, S., Vandewalle, J., "The total least squares technique: computation, properties and applications", in *SVD and Signal Processing* E. Deprettere (ed.), North Holland, 1988, pp. 189-207.

[39] Vandewalle, J., De Moor, B., "A variety of applications of singular value decomposition in identification and signal processing"", in *SVD and Signal Processing* E. Deprettere (ed.), North Holland, 1988, pp. 43-91.

[40] Callaerts, D., Vandewalle, J., Sansen, W., Moonen, M., "On-line algorithm for signal separation based on SVD", in *SVD and Signal Processing* E. Deprettere (ed.), North Holland, 1988, pp. 269-276.

[41] De Moor, B., Moonen, M., Vandenberghe, L., Vandewalle, J., "Identification of linear state space models with singular value decomposition using canonical correlation concepts", in *SVD and Signal Processing* E. Deprettere (ed.), North Holland, 1988, pp. 161-169.

[42] De Moor, B., Moonen, M., Vandenberghe, L., Vandewalle, J., "The application of the canonical correlation concept to the identification of linear state space models", in A. Bensousan, J.L. Lions, (Eds.) *Analysis and Optimization of Systems* Springer Verlag, 1988, pp. 1103-1114.

[43] De Moor, B., Moonen, M., Vandenberghe, L., Vandewalle, J., "A geometrical approach for the identification of state space models with singular value decomposition", *Proceedings of the International Conference on Acoustics, Speech and Signal Processing 1988*, New york, April 1988, pp. 2244-2247.

[44] De Moor B., *Mathematical concepts and techniques for modelling of static and dynamic systems*, Ph. D. Thesis, Katholieke Universiteit Leuven, Electrical Eng. Dept., 1988.

[45] De Moor B., Golub G.H. *Generalized Singular Value Decompositions: A proposal for a standardized nomenclature*. Report, Dept. Computer Science, January 1989, Stanford University.