De Moor B., Vandewalle J. *Oriented energy and oriented signal to signal ratios : The (Generalized) Singular Value Decomposition* in 'SVD and Signal Processing: Algorithms, Applications and Architectures.' Edited : E.F.Deprettere, North Holland, Elseviers Science Publishers B.V., 1988, pp. 209-232.

ORIENTED ENERGY AND ORIENTED SIGNAL-TO-SIGNAL RATIO CONCEPTS IN THE ANALYSIS OF VECTOR SEQUENCES AND TIME SERIES

Bart De Moor ; Jan Staar ; Joos Vandewalle
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

Keywords: Principal Component Analysis, Factor analysis, prewhitening, Mahalanobis transformation, Generalized Singular Value Decomposition, Total Linear Least Squares

Abstract

In this paper, two concepts are introduced : the oriented energy distribution of one vector sequence and the oriented signal-to-signal ratio concept of two vector sequences. It is shown how both concepts can be characterized numerically via the (generalized) singular value decomposition and how they allow to formalize factor-analysis-like (subspace-) methods. Several applications are described, including Total Linear Least Squares, realization theory and source separation.

1 Introduction

In a wide variety of systems and signal processing applications, vector sequences are measured or computed. Such a situation naturally arises whenever multivariable signals are measured in time at fixed locations in a measurement set up. For the analysis of such data sequences, a wide variety of multivariate analysis tools are available. The underlying theme of much multivariate analysis is simplification and explanation of the observed phenomena. In this paper, the problem of analysis of one or two $m \times n$ data matrices A and B is addressed. Usually $n \gg m$, where m denotes the number of measured channels while n denotes the number of measurements. It has occurred to some researchers in both signal processing and control systems that the singular value decomposition of matrices formed from observed data could be used to improve methods of signal parameter estimation and system identification. However, rationales for these methods have been very heuristic and in almost all cases are based upon the well-posedness of the algorithm, in casu the SVD. One purpose of this paper is to present a more convincing framework which is intended both to unify existing techniques and widen the area of applications.

The results obtained in this work, bear a lot of similarity with existing (statistical) techniques, such as principal component analysis, factor analysis, analysis of variance etc... Principal component analysis, originating in some work by Karl Pearson around the turn of the century and further developed in the 1930s by Harold Hotelling, consists of finding an orthogonal transformation of the original - stochastic - variables to a new set of uncorrelated variables, which are derived in non-increasing order of importance. These so-called principal components are linear combinations of the original variables and it is the analysist's hope that the first few components will account for most of the variation in the original data so that the effective dimensionality of the data can be reduced [3]. The concept of oriented energy defined and studied in this work, is closely related to principal component analysis. The concept of oriented signal-to-signal ratio is closely related to factor-analysis like methods (in 'modern' approaches so-called subspace methods) in which the used metric is imposed by the noise covariance matrix, acting as a prewhitener of the measurements via a so-called Mahalanobis transformation.

It will be shown how the framework of oriented signal-to-signal ratio provides a rationale for linear modeling problems where :

1. the complexity of the model is the rank of a certain matrix. Hence, the decision for the complexity essentially reduces to the meaningful determination of the rank of certain (prewhitened) matrices.

2. the model parameters are linked in one way or another to the subspaces and their properties, that are associated to the determined rank.

Several new aspects are emphasised throughout this paper. They constitute its main contribution:

- A general framework is derived, explicitly based upon the properties of the singular and generalized singular value decomposition. No statistical a priori assumptions (about e.g. probability distributions) are imposed. However, when such a priori information is available, it can be easily taken into account.

- The conceptual derivations based upon the (generalized) singular value decomposition are at the same time constructive: The very use of these factorizations delivers algorithms, that may be implemented in numerically robust and reliable software.

- Whereas the singular value decomposition is one of the tools in the analysis of a single vector sequence, it will be shown how the generalized singular value decomposition is the technique to be used in analyzing the mutual relation of two vector sequences.

Several examples will be presented that lend themselves to a translation and interpretation in the novel framework : total linear least squares with specified admissible complexity or tolerated misfit, high resolution location of narrowband sources, separation of maternal from fetal ECG and linear dynamical realization theory.

This paper is organized as follows: In section 2, the basic definitions and theorems of the concept of oriented energy and signal-to-signal ratio are defined and derived. The numerical tool to analyse the spatial activity of one vector sequence is the singular value decomposition as is demonstrated in section 3. When two vector sequences are to be studied relatively to each other, the generalized singular value decomposition applies. It is shown in section 4 how there exists a strong similarity between the singular value decomposition for the analysis of one vector sequence and the generalized singular value decomposition for the analysis of two vectorsequences. In section 5, the results are illustrated with some clarifying examples. The conclusions can be found in section 6.

Notations and abbreviations:

All matrices and vectors are assumed to be real (although the concepts defined in this paper readily generalize to the complex field). Column vectors are denoted by small letters. Row vectors are denoted as the transpose of columnvectors. Capitals represent matrices. The letters i, j, k, l, m, n and r are integers (used e.g. for indexing) while real numbers are denoted by Greek symbols. No distinction is made between the notation of a linear transformation and its matrix representation nor between the notation of a vector and the column vector with its coordinates in some basis.

$\mathcal{R}^n$: n dimensional vectorspace of real n-tuples

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

A^t : transpose of the matrix A

A^+ : pseudo-inverse of the matrix A

$a_{n\times 1}$: real vector with n real components. Also called an n-vector

$a^t \cdot b$: Euclidean inner product of two n-vectors : $\sum_{i=1}^{n} a_i \cdot b_i$

e_i : unit vector with 1 as i-th component and other components 0

$\|a\|_2$: 2-norm of a vector $\sqrt{(a^t \cdot a)}$. A vector with norm 1 is called a unit vector

$\|A\|_F$: Frobenius norm of a matrix $\sqrt{(\sum\sum a_{ij}^2)}$ where the double sum extends over all rows and columns

$E_q[A]$: oriented energy of the column vector sequence of the matrix A in the direction of the vector q (definition 2)

$R_q[A, B]$: Oriented signal-to-signal ratio of the vectorsequences A and B (definition 3)

MmR$[A, B, r]$: The maximal minimal signal-to-signal ratio over all r-dimensional subspaces of the vector sequences $\{a_k\}, \{b_k\}$ (definition 5)

mMR$[A, B, r]$: The minimal maximal signal-to-signal ratio over all r-dimensional subspaces of the vector sequences $\{a_k\}, \{b_k\}$ (definition 5)

rank(A) : algebraic rank of the matrix A

$\text{span}_{\text{row}}(A)$: vectorspace generated by the rowvectors of the matrix A

$\text{span}_{\text{col}}(A)$: vectorspace generated by the columnvectors of the matrix A

$\text{span}_{\text{col}}[u_1, \ldots, u_r]$: if u_i are linearly independent m-vectors this denotes the r-dimensional subspace generated by the vectors $u_1, \ldots, u_r$

$\text{span}_{\text{col}}^{\perp}[u_1, \ldots, u_r]$: denotes the $m-r$-dimensional orthogonal complement

$Q^r \subset \mathcal{R}^m$: Q^r is a r-dimensional subspace of $\mathcal{R}^m$

int(α) : integer truncation of the real number α

SVD : Singular Value Decomposition (Section 3.1)

GSVD : Generalized Singular Value Decomposition (Section 4.1)

UB : The unit ball UB $= \{q \in \mathcal{R}^m \mid q^t \cdot q = 1\}$

2 Oriented energy and oriented signal-to-signal ratio concepts of a set of vectors

In this section, the basic definitions of oriented energy are given . The column vectors of an $m \times n$ matrix A are considered to form an indexed set of m-vectors, denoted by $\{a_k\}, k = 1, \ldots, n$. An m-vector q and the direction it represents in a vector space, are used as synonyms.

Definition 1 Energy of a vector sequence.
Consider a sequence of m-vectors $\{a_k\}, k = 1, \ldots, n$ and associated $m \times n$ matrix A. Its total energy $E[A]$ is defined via the Frobeniusnorm of the $m \times n$ matrix A:

$$E[A] = \|A\|_F^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \tag{1}$$

Definition 2 Oriented energy.
Let A be a $m \times n$ matrix and denote its n columnvectors as $a_k, k = 1, \ldots, n$. (n is possibly infinite) For the indexed vectorset $\{a_k\}$ of m-vectors $a_k \in \mathcal{R}^m$ and for any unit vector $q \in \mathcal{R}^m$ the energy E_q, measured in the direction q, is defined as:

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

A geometric visualisation is represented in fig.1.

More generally, the energy E_Q measured in a subspace $Q \subset \mathcal{R}^m$, is defined as:

$$E_Q[A] = \sum_{k=1}^{n} \|P_Q(a_k)\|^2 \tag{3}$$

where $P_Q(a_k)$ denotes the orthogonal projection of a_k into the subspace Q and $\|\cdot\|$ denotes the Euclidean norm.

Of course, the summations in (1) and (2) require l^2-type convergence conditions on the set $\{a_k\}$ when n is infinite. In words, the oriented energy of a vector sequence $\{a_k\}$, measured in the direction q (subspace Q) is nothing else than the energy of the signal, orthogonally projected on the vector q (subspace Q).

In order to sharpen the geometric intuition on the above interpretations, the oriented energy distribution and the square root of this distribution is shown in fig.2 for a 3-vector sequence which is nearly singular. The physical significance of the definition becomes clear, when one considers the unit vector q (subspace Q) to be variable and to "sense" in all directions of the vectorspace $\mathcal{R}^m$. For a sequence of 3-vectors, we have plotted in each

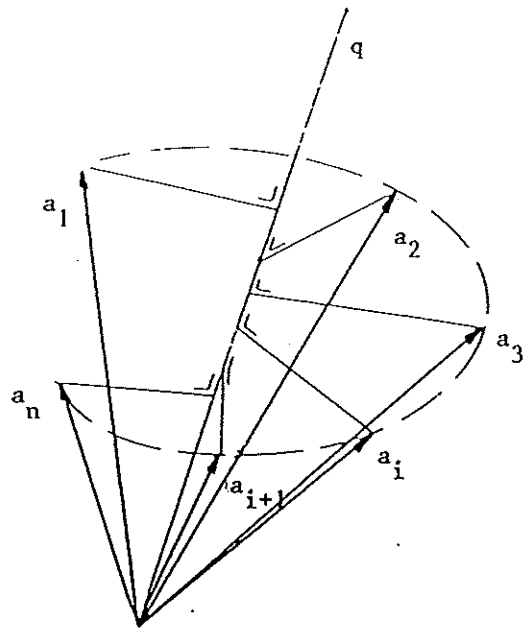

Figure 1: Illustration of oriented energy measurement

"sensing" direction q of the 3-dimensional vectorspace, a vector of length E_q. All points lie on a smooth surface (fig.2). In fig.3 a quarter of the surface has been cut out to show the typical ("sharp") narrowing that may occur in some directions. In three dimensions, the surfaces of oriented energy exhibit one direction of maximal oriented energy, one of minimal energy an a direction with a saddle point. This can readily be generalized to more dimensions.

Let us now consider two vector sequences $\{a_k\}$ and $\{b_k\}$

Definition 3 Oriented signal-to-signal ratio
The oriented signal-to-signal ratio $R_q[A,B]$ for two sets of m-vectors $\{a_k\}$ and $\{b_k\}, k = 1,\ldots,n$ (n possibly infinite), measured in the direction of a unit vector $q \in \mathcal{R}^m$, is defined by:

$$R_q[A,B] = \frac{E_q[A]}{E_q[B]} \tag{4}$$

More generally, the oriented signal-to-signal ratio $R_Q[A,B]$ for two vector sequences $\{a_k\}$ and $\{b_k\}$, measured in a subspace $Q \subset R^m$, is defined as:

$$R_Q[A,B] = \frac{E_Q[A]}{E_Q[B]} \tag{5}$$

In words, the signal-to-signal ratio of two vector sequences, measured in a direction q or subspace Q, is simply the ratio of the two oriented energies of the involved signal sequences in that direction or subspace. Remark that when B is rankdeficient, there exist directions in which the oriented energy $E_q[B] = 0$, possibly making the signal-to-signal ratio infinite if $E_q[A] \neq 0$.

The oriented energy distribution of course shows a more than coincidental relationship with the ellipsoid, described by the nonnegative definite quadratic form of $A \cdot A'$:

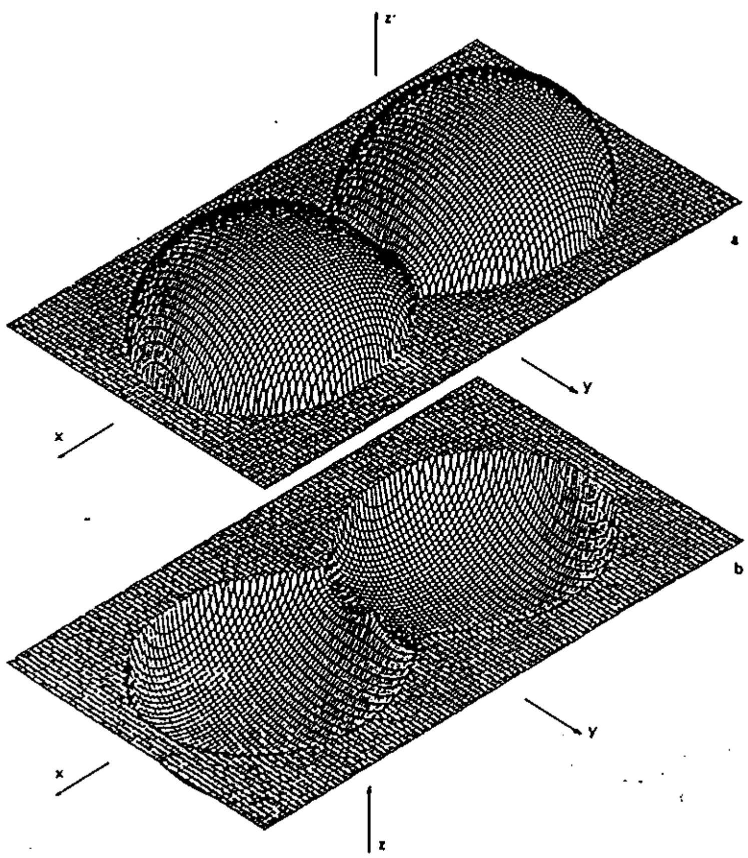

Figure 2: Oriented energy distribution in 3 dimensions

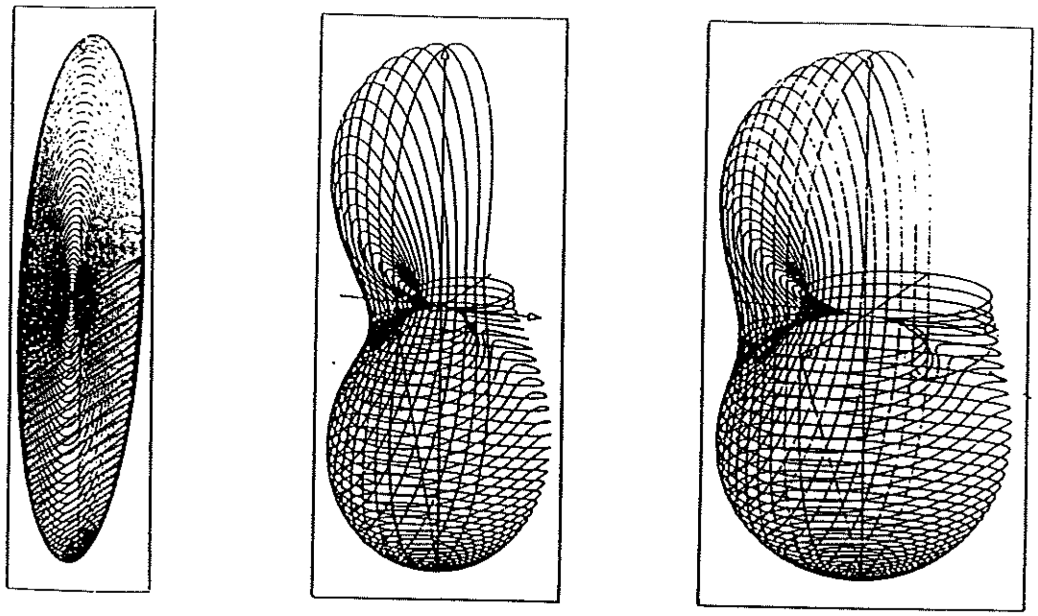

Figure 3: Oriented energy distribution in 3 dimensions

Theorem 1 *Consider the $m \times n$ matrix A and the vector sequence $\{a_k\}$ of its column vectors. Then, any m-vector r of the ellipsoid $\{r \mid r^t \cdot A \cdot A^t \cdot r = 1\}$, associated with the quadratic form of the matrix $A \cdot A^t$ and the oriented energy of the vector sequence $\{a_k\}$ in the direction $r/\|r\|$ are related by:*

$$\|r\|^2 \cdot E_{r/\|r\|}[A] = 1 \tag{6}$$

Proof: Trivial

□

In words, the energy distribution of a sequence of vectors $\{a_k\}$ can be constructed from the ellipsoid of the quadratic form of $A \cdot A^t$ by scaling any vector r on the ellipsoid until its length is $1/\|r\|^2$. Hence, the matrix $A \cdot A^t$ characterizes the quadratic form ellipsoid as well as the oriented energy distribution.

This correspondance implies that the oriented energy is everywhere continuous on the unit sphere and also everywhere differentiable. The relationship (5) directly implies that directions of extremal energy coincide with the principal axes of the ellipsoid, hence are orthogonal. This observation really invites to use the oriented energy concept in the analysis of the spatial activity of vector signals.

The importance of the new notion of oriented energy, with its close relation to the 'classical' quadratic form, will follow from both the conceptual as the numerical arguments developed in the remainder of this paper. It will be demonstrated how the oriented energy concept is indeed a powerful tool to separate signals from different sources, to 'filter' signals from noise and to select subspaces of maximal signal activity and integrity.

An important observation is that the range of values of the oriented energy distribution of a vector sequence, depends upon the choice of basis to which the vector coordinates refer. Moreover, the directions of extremal oriented energy are not preserved and the shape and values of the oriented energy distribution may totally change under non-orthonormal basis transformations. This indicates that the notion of oriented energy is a workable concept only if the choice of basis for the representation of the vector set $\{a_k\}$ is fixed by external or physical arguments. A similar observation holds for the oriented signal-to-signal ratio $R_q[A, B]$ of two vector sequences $\{a_k\}$ and $\{b_k\}$ in a fixed direction q. However, the range of the values of $R_q[A, B]$ considered over all unit directions q, is independent of the choice of basis. This implies that they have a wider physical significance. This important invariance property is stated more precisely as follows.

Theorem 2 Invariance property of the signal-to-signal ratio
Consider 2 sequences of m-vectors $\{a_k\}, \{b_k\}, k = 1, \ldots, n$. For every unit vector $q \in R^m$ and for every non-singular $m \times m$ matrix T that transforms $\{a_k\}, \{b_k\}$ into $\{T \cdot a_k\}, \{T \cdot b_k\}$ there exists an associated vector q' such that

$$R_q[A, B] = R_{q'}[T \cdot A, T \cdot B] \tag{7}$$

Proof: Verify that (7) is satisfied for $q' = \frac{(T^{-1})^t \cdot q}{\|(T^{-1})^t \cdot q\|}$. □

The message of theorem 2 is the following: Although the measurements of oriented energy ratios in a fixed direction depend on the choice of basis, the existence of a ratio with a certain value is independent of the chosen basis. Stated otherwise : if the oriented signal-to-signal ratio has a certain value with respect to a certain basis, it will have the same value in some direction for all possible choices of basis. Note that theorem 2 is constructive in the sense that it allows to compute this specific direction in the new basis.

3 The oriented energy concept and the singular value decomposition

In section 2, attention was paid to the basic concepts and properties of the oriented energy distribution . In this section, the tools will be studied which allow to characterize numerically the oriented energy concept. In section 3.1, results about the singular value decomposition are summarized. More conceptual relations between the SVD and the oriented energy properties of a vector sequence are established in section 3.2. In section 3.3, some numerical considerations are discussed.

3.1 The singular value decomposition (SVD)

For conceptual, numerical, algebraic and computational reasons, the singular value decomposition (SVD) is receiving more and more attention [6]. The SVD for real matrices is based upon the following theorem [6] which we name after its most important contributors:

Theorem 3 The Autonne-Eckart-Young theorem *(restricted to real matrices)*
For any real $m \times n$ matrix A, there exist a real factorization :

$$A = \underset{m \times m}{U} \cdot \underset{m \times n}{S} \cdot \underset{n \times n}{V^t} \tag{8}$$

in which the matrices U and V are real orthonormal, and the matrix S is real pseudo-diagonal with nonnegative diagonal elements.

The diagonal entries σ_i of S are called the singular values of the matrix A. It is assumed that they are sorted in non-increasing order of magnitude. The set of singular values $\{\sigma_i\}$ is called the singular spectrum of the matrix A. The columns $u_i(v_i)$ of $U(V)$ are called the left (right) singular vectors of the matrix A. The space $S_U^r = \text{span}_{\text{col}}[u_1, \ldots, u_r]$ is called the r-th left principal subspace. In a similar way, the r-th right singular subspace is defined. The triple (u_i, σ_i, v_i) is called the i-th singular triplet of the matrix A. Remark that the singular value decomposition of a real matrix is not unique. However, the singular values are uniquely determined. If the non-zero singular values are distinct, the corresponding singular vectors are unique up to the sign. If r singular values (zero or non-zero) coincide, the corresponding singular vectors are arbitrary as long as they generate an orthonormal basis for the corresponding r-dimensional subspace which is unique.

Proofs of the above classical existence and uniqueness theorems are found in [6] and the references therein. Some more properties of the singular value decomposition are mentioned here without proof.

Lemma 1 *The number of singular values, different from zero, equals the algebraic rank of the matrix A.*

In fact, the SVD is one of the most reliable tools to estimate in a numerically sound way the algebraic rank of a matrix.

Lemma 2 Dyadic decomposition
Via the SVD , any matrix A can be written as the sum of $r = \text{rank}(A)$ rank one matrices :

$$A = \sum_{i=1}^{r} u_i \cdot \sigma_i \cdot v_i^t \tag{9}$$

where (u_i, σ_i, v_i^t) is the i-th singular triplet of the matrix A.

Lemma 3 Frobenius norm of $m \times n$ matrix A of rank r

$$\|A\|_F^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2 = \sum_{k=1}^{r} \sigma_k^2 \tag{10}$$

where the σ_k are the singular values of A.

In words, the total energy in a vector sequence $\{a_k\}$ with associated matrix A as defined in definition 1, is equal to the energy in the singular spectrum.

The smallest non-zero singular value corresponds to the distance in Frobeniusnorm, of the matrix to the closest matrix of lower rank. This well known property makes the SVD attractive for approximation and data reduction purposes. There exists an important well-known relation between the singular value decomposition and the eigenvalue decomposition:

Lemma 4 *Let the $m \times n$ matrix A have an SVD as in (8). Then the columns of U are the eigenvectors of the Grammian $A \cdot A^t$ The rows of V are the eigenvectors of the Grammian $A^t \cdot A$. The positive real number σ_i is a non-zero singular value of A iff σ_i^2 is a non-zero eigenvalue of both $A \cdot A^t$ and $A^t \cdot A$.*

3.2 Conceptual relations between SVD and oriented energy

We are now in the position to establish the link between the singular value decomposition and the concept of oriented energy distribution.

Define the unit ball UB in $\mathcal{R}^m$ as $\text{UB} = \{q \in \mathcal{R}^m \mid \|q\|_2 = 1\}$

Theorem 4 *Consider a sequence of m-vectors $\{a_k\}, k = 1, \ldots, n$ and the associated $m \times n$ matrix A with SVD as defined in (8) with $n \geq m$. Then:*

$$E_{u_i}[A] = \sigma_i^2 \tag{11}$$

$\forall q \in \text{UB}$: *if* $q = \sum_{i=1}^m \gamma_i \cdot u_i$, *then*

$$E_q[A] = \sum_{i=1}^m \gamma_i^2 \cdot \sigma_i^2 \tag{12}$$

where UB *is the unit ball.*

Proof: Trivial from theorem 3. □

In words, the oriented energy measured in the direction of the i-th left singular vector of the matrix A, is equal to the i-th singular value squared. The energy in an arbitrary direction q can be reconstructed additively as a sum of 'orthogonal' oriented energies associated to the left singular directions, as soon as the coordinates γ_i of the vector q with respect to the left singular vectors are known. If the matrix A is rankdeficient, then there exist directions in $\mathcal{R}^m$ that contain no energy at all.

It should be observed that the singular values and vectors are generally critically dependent upon the scales used to measure the variables. This scaling could be the result of data acquisition requirements such as amplification, A/D conversion etc... Hence, additional physical motivitation is required to choose those scalings for the different measurement channels for which it is meaningful to 'compare' via the oriented energy concept. In the sequel it is assumed that this question has been resolved in advance.

With the aid of theorem 4, one can easily obtain, using the SVD, the directions and spaces of extremal energy, as follows.

Corollary 1 *Under the assumptions of theorem 4:*

1. $\max_{q \in \text{UB}} E_q[A] = E_{u_1}[A] = \sigma_1^2$ (13)
2. $\min_{q \in \text{UB}} E_q[A] = E_{u_m}[A] = \sigma_m^2$ (14)
3. $\max_{Q^r \subset \mathcal{R}^m} E_{Q^r}[A] = E_{S_U^r}[A] = \sum_{i=1}^r \sigma_i^2$ (15)
4. $\min_{Q^r \subset \mathcal{R}^m} E_{Q^r}[A] = E_{(S_U^{m-r})^\perp}[A] = \sum_{i=m-r+1}^m \sigma_i^2$ (16)
5. $\max_{Q^r \subset \mathcal{R}^m} \{\min_{q \in Q^r} E_q[A]\} = \min_{q \in S_U^r} E_q[A] = \sigma_r^2$ (17)
6. $\min_{Q^r \subset \mathcal{R}^m} \{\max_{q \in Q^r} E_q[A]\} = \max_{q \in (S_U^{m-r})^\perp} E_q[A] = \sigma_{m-r+1}^2$ (18)

where 'max' and 'min' denote operators, maximizing or minimizing over all r-dimensional subspaces Q^r of the ambient range space $\mathcal{R}^m$. S_U^r is the r-dimensional principal subspace of the matrix A while $(S_U^{m-r})^\perp$ denotes the r-dimensional orthogonal complement of S_U^{m-r}.

Proof: Property (13), (14), (15), (16) follow immediately from the SVD theorem 3 and from theorem 4 . Property (17) and (18) are nothing else then the classical Courant-Fischer minimax and maximin characterizations of the eigenvalues [6] □

In words, properties (13) and (14) relate the SVD to the mimima and maxima of the oriented energy distribution. In fact, it can be shown that extrema occur at each left singular direction. The r-th principal subspace S_U^r is, among all r-dimensional subspaces of $\mathcal{R}^m$, the one that senses a maximal oriented energy (property 15). Properties (15) and (16) show that the orthogonal decomposition of the energy via the singular value decomposition is canonical in the sense that it allows to find subspaces of dimension r where the sequence has minimal and maximal energy. This decomposition of the ambient space, as a direct sum of a space of maximal and minimal energy for a given vector sequence, leads to very interesting rank considerations, which will be exploited furtheron. Properties (17) and (18) characterize the min-max properties of the vectorsequence if this is restricted to p-dimensional subspaces. For a fixed subspace Q^r, the minimum energy is achieved for a certain direction q. When all minima are considered for all possible r-dimensional subspaces Q^r, then there is at least one maximum. This maximum of all minima can be interpreted as the best of all worst cases. Its algorithmic computation and the determination of the corresponding maximizing subspace, is closely related to the singular value decomposition, as is demonstrated in (17). A similar interpretation can be given for (18).

In signal processing, one often encounters long sequences of m-vectors a_k. This means that the corresponding $m \times n$ matrix will be largely overdetermined with much more columns n than rows m : $n \gg m$. The singular value decomposition allows to compact such sequences into sequences with equivalent oriented energy properties:

Theorem 5 *Consider a sequence of m-vectors $\{a_k\}, k = 1, \ldots, n$ and the associated $m \times n$ matrix A with SVD as defined in (8) with $n \geq m$.*

Then the sequence of m-vectors

$$\{u_1\sigma_1, u_2\sigma_2, \ldots, u_m\sigma_m\} \tag{19}$$

has the same oriented energy distribution as that of $\{a_k\}$.

Proof: Straightforward. □

One of the main applications of this theorem concerns vector stochastic processes: Ergodic vector stochastic processes can be characterized by an equivalent vector signal $U \cdot \Sigma$, closely related to the second-order joint moment matrix of the process. This implies that from the point of view of oriented energy,the sequence should not be known by an actual time realization. Only the knowledge of the equivalent sequence (19) with identical oriented energy distribution is required. An example will be considered in section 5.1.

Definition 4 Isotropy

The oriented energy distribution of a sequence of m-vectors $\{a_k\}$ will be called isotropic if the singular values of the corresponding matrix A are all equal to each other.

3.3 Numerical considerations

The practical value of theorem 5 is that it allows to compact a large amount of data without algebraic or numerical degeneracies. In this context, one should clearly distinguish theorem

4 from the compaction obtained by the computation of $A \cdot A^t$ described in theorem 1, and this especially concerning the numerical caveats. Indeed, all properties have been stated in terms of energies. This implies the summing of squares, which, in the presence of numerical round off errors caused by the limited machine precision, can lead to numerical disasters. The strength of the approach is that all computations can be performed without explicitly using these squares. The singular value decomposition obtains the decomposition of the vector sequence as a sum of orthogonal dyadic terms, weighted with singular values that can be computed within full machine precision. Numerically stable and reliable algorithms for the singular value decomposition are by now well known [6], fully tested and documented and available in reliable standard softwarepackages (e.g. Matlab, Eispack,NAG,...)

Another crucial issue concerns the computational cost of the singular value decomposition. Typically, in signal processing applications, the number of vectors n in a m-vectorsequence $\{a_k\}$ is much larger than the number of components m. This implies that the associated $m \times n$ matrix A is largely overdetermined ($n \gg m$) and hence the SVD of a 'very rectangular' matrix is required. Fortunately, there exists a simple 'trick' which allows to stably compute such SVD's. Without going into technical detail, it suffices to mention that first the R-Q factorization of the overdetermined matrix is computed , followed by the SVD of the lower triangular matrix R. This results an a considerable computational saving. Moreoever, since it is 'easy' and 'cheap' to compute a rank one update of the R-Q factorization (necessary when one extra column is added), this opens interesting perspectives for an algorithm that is adaptive in the number of measurements (updating and downdating strategies).

4 Signal-to-signal ratios and the generalized singular value decomposition

While in the previous section, the link between the oriented energy distribution of one vector sequence with the SVD of the associated matrix was established, in this section the relation of the signal-to-signal ratio of two vector sequences with the generalized singular value decomposition will be studied. The use of the generalized singular value decomposition allows to develop a highly instrumental parallellism between the concept of oriented energy distribution and the signal-to-signal ratio of two vector sequences. In section 4.1, the theorem stating the existence of the generalized singular value decomposition is given together with its main properties. In section 4.2., it is shown how to apply the GSVD in order to compute the maximal minimal signal-to-signal ratio of two vectorsequences. In section 4.3., attention is paid to the numerical implications of the GSVD.

4.1 The Generalized Singular Value Decomposition [GSVD]

Theorem 6 GSVD
Let A be a $m \times n (n \geq m)$ and B a $m \times p$ matrix, then there exist orthogonal matrices U $(n \times n)$ and V $(p \times p)$ and a non-singular $m \times m$ matrix X such that

$$A = X^{-1} \cdot D_A \cdot U^t \tag{20}$$
$$B = X^{-1} \cdot D_B \cdot V^t \tag{21}$$

where

$$D_A = \mathrm{diag}(\alpha_1, \ldots, \alpha_m), \quad \alpha_i \geq 0,$$

is a rectangular diagonal $m \times n$ matrix, and

$$D_B = \mathrm{diag}(\beta_1, \ldots, \beta_q) \quad, \beta_i \geq 0, \quad q = \min(m, p),$$

is a rectangular diagonal $m \times p$ matrix and

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

Proof: see e.g. [6] □

The elements of the set $\sigma(A, B) = \{\alpha_1/\beta_1, \ldots, \alpha_r/\beta_r\}$ are referred to as the generalized singular values of A and B. The theorem is a generalization of the SVD since $\sigma(A, B)$ equals the singular spectrum of the matrix A if $B = I_m$. In this paper, the case where $n > m$ and $p > m$ will be of interest. In that case, remark that if U and V are partitioned as $U = [U_1 \;\; U_2]$ and $V = [V_1 \;\; V_2]$ where U_1, U_2, V_1, V_2 are $m \times n, m \times (n - m), p \times m, p \times (p - m)$ matrices, the GSVD of A and B can be written as

$$\begin{aligned} A &= X^{-1} \cdot D_A \cdot U_1^t \\ B &= X^{-1} \cdot D_B \cdot V_1^t \end{aligned} \qquad (22)$$

where D_a and D_b are now square diagonal. This notation will be used from here on.

There exists an intimate theoretical link between the generalized singular value decomposition of the matrix pair $[A, B]$ and the generalized symmetric eigenvalue problem:

Lemma 5 : *Let A and B be as in theorem 6. Then the generalized singular values are the square roots of the generalized eigenvalues γ of the symmetric eigenvalue problem :*

$$A \cdot A^t \cdot x = \gamma B \cdot B^t \cdot x.$$

The matrix X of theorem 6 contains the generalized eigenvectors and diagonalizes simultaneously $A.A^t$ and $B.B^t$.

For more interesting properties, the reader is referred to [6].

4.2 Conceptual relations between the signal-to-signal ratio and GSVD

In this section, it will be demonstrated how the Generalized Singular Value Decomposition allows to characterize the signal-to-signal ratio of two given sequences of m-vectors $\{a_k\}, \{b_k\}, k = 1, \ldots, n$ with associated $m \times n$ matrices A and B. It is assumed that $n > m$ as is the case in signal processing applications. Often, one of the signals, say $\{a_k\}$ can be considered as the desired signal while the second one $\{b_k\}$ represents the undesired signal that in some sense corrupts the desired one. The problem of interest then clearly consists in separating the desired part from the undesired signal. In quite some applications, examples of which will be presented in section 5, the associated $m \times n$ matrix A is rankdeficient: $\mathrm{rank}(A) < m$. In this case, one has a physical motivation to restrict the attention to r-dimensional subspaces of the ambient space (a recent rationale is developed in [12]; the literature covering such rank-decision tests is enormous, including maximum likelihood eigenvalue ratio tests [2], Akaike's Information Criterion ([1]), Rissanen's Minimum Description Length Criterion [9], Willems' recent results on complexity/misfit approximate modeling [18] etc....) Once the rank r has been fixed, the question of interest is to find the optimal r-dimensional subspace, in which the desired signal sequence $\{a_k\}$ can be optimally distinguished from the corrupting sequence $\{b_k\}$. It will be shown that this is equivalent to determining the maximal minimal signal-to-signal ratio of the two vectorsequences. In order to avoid unnecessary complication, caused by possible rank deficiency of B, it is assumed from now on that B is of full row rank i.e. $\text{rank}(B) = m$. If B is rank deficient with $\text{rank}(B) = r < m$, the vector sequence $\{b_k\}$ has no energy in the orthogonal complement of the r-th left principal subspace of B. For every direction q in this orthogonal complement, the signal-to-signal ratio $R_q[A, B]$ is infinite if $E_q[A] \neq 0$. Such directions can easily be dealt with in advance by some kind of deflation-orthogonalisation procedure applied to the column spaces of the matrices A and B. In a signal processing context however, the full row rank situation is the generic one. For these reasons, the possible rank deficiency of B will not be considered in detail in our further discussion. Two elements will be used : The invariance of the signal-to signal ratio distribution under non-singular transformations (theorem 2) and the generalized singular value decomposition of the pair of $m \times n$ matrices A and B.

Theorem 7 *Given two sequences of m-vectors $\{a_k\}$ and $\{b_k\}, k = 1, \ldots, n$ with associated $m \times n$ matrices A and B $(n > m)$ where $\text{rank}(B) = m$. Consider the GSVD of A and B as in (22):*

$$A = X^{-1} \cdot D_A \cdot U_1^t$$
$$B = X^{-1} \cdot D_B \cdot V_1^t$$

Define the linear transformation:

$$T = D_B^{-1} \cdot X \tag{23}$$

and the transformed vector sequence $\{c_k\}$ via $c_k = T .a_k$ with associated $m \times n$ matrix $C = T \cdot A$. Then:

$$E_{q'}[C] = R_q[A, B] \tag{24}$$

where

$$q' = \frac{(T^{-1})^t \cdot q}{\|(T^{-1})^t \cdot q\|} \tag{25}$$

Proof: This theorem is an immediate consequence of the invariance theorem 2 for the signal-to-signal ratio. Indeed, with q' given by (25), we have:

$$\begin{aligned} R_q[A, B] &= R_{q'}[T \cdot A, T \cdot B] \\ &= R_{q'}[D_B^{-1} \cdot D_A \cdot U_1^t, V_1^t] \\ &= E_{q'}[D_B^{-1} \cdot D_A \cdot U_1^t] \end{aligned}$$

because the oriented energy distribution of the sequence V_1^t is isotropic : $E_{q'}[V_1^t] = 1$. □

Theorem 7 links the signal-to-signal ratio with the oriented energy distribution in the following way. A linear transformation T transforms the vectorsequence $\{b_k\}$ into an isotropic sequence with unit energy distribution. By the invariance theorem 2, it is guaranteed that the signal-to-signal ratio distribution is preserved . Moreover, all information on the signal-to-signal ratio is now available in the oriented energy distribution of the transformed sequence $T \cdot A = D_B^{-1} \cdot D_A \cdot U_1^t$. The sequence $T \cdot B$ has become isotropic. Hence, the linear transformation T could be considered as some kind of 'whitening' operator, an idea which is commonly applied in statistics (in e.g. minimum variance and Markov-type estimators) [7]. But, the most important observation is that, by the very choice of the linear transformation T as

$T = D_B^{-1} \cdot X$, the resulting sequence $D_B^{-1} \cdot D_A \cdot U_1^t$ has precisely the form of a singular value decomposition, in which the left singular matrix equals the identity matrix. This allows to adapt theorem 4 and corollary 1 directly to the properties of the signal-to-signal ratios in a straightforward way:

Corollary 2 *Consider the GSVD of the matrix pair A and B as in (22). Assume that the generalized singular values are ordered in non-increasing order of magnitude:* $(\alpha_i/\beta_i) \geq (\alpha_{i+1}/\beta_{i+1})$. *Denote by* t_i^t *the i-th row of the matrix* $T = D_B^{-1}X$ *and by* x_i^t *the i-th row of the matrix X. (Clearly* $t_i^t = (1/\beta_i)x_i^t$*).* UB *is the unit ball. Under the assumptions and notations of theorem 7:*

1. *For* $q = t_i/\|t_i\|, R_q[A, B] = (\alpha_i/\beta_i)^2$ (26)
2. *If* $q = \sum_{i=1}^m \gamma_i \cdot t_i^t$, *then*
$$R_q[A, B] = \frac{\sum_{i=1}^m(\gamma_i \cdot \alpha_i/\beta_i)^2}{\sum_{i=1}^m \gamma_i^2} \tag{27}$$
3. $\max_{q\in \mathrm{UB}} R_q[A, B] = R_{t_1/\|t_1\|}[A, B] = (\alpha_1/\beta_1)^2$ (28)
4. $\min_{q\in \mathrm{UB}} = R_q[A, B] = R_{t_m/\|t_m\|}[A, B] = (\alpha_m/\beta_m)^2$ (29)

Proof: Properties 1/ and 2/ follow by straightforward substitution while properties 3/ and 4/ are special cases of property 2 and of the extremal relationships between oriented energy and the singular value decomposition (theorem 4 and corollary 1). □

First remark that $t_i/\|t_i\| = x_i/\|x_i\|$. As is expressed in property (26), the GSVD not only provides the extrema of the signal-to-signal ratio but also the directions in which those extrema occur: These are simply the rows of the matrix X . Hence the extreme directions of oriented signal-to-signal ratio need not to be orthogonal. The minimal and maximal signal-to-signal ratios and direction can be found in properties (28) and (29). If the coordinates of a vector q are known with respect to the basis generated by the rows of the matrix T, then the signal-to-signal ratio follows immediately from the generalized singular values (property (27)) Let us now consider some optimal and worst case signal-to-signal ratios.

Definition 5 Maximal minimal and minimal maximal signal-to-signal ratio
The maximal minimal signal-to-signal ratio of two m-vector sequences $\{a_k\}$ *and* $\{b_k\}$ $k = 1, \ldots, n$ *over all r-dimensional subspaces* $(r \leq m < n)$*, denoted by* MmR$[A, B, r]$, *is defined as:*

$$\mathrm{MmR}[A, B, r] = \max_{Q^r \subset \mathcal{R}^m} \min_{q\in Q^r} R_q[A, B] \tag{30}$$

The minimal maximal signal-to-signal ratio, mMR$[A, B, r]$ *over all r-dimensional subspaces, is defined as:*

$$\mathrm{mMR}[A, B, r] = \min_{Q \subset \mathcal{R}^m} \max_{q\in Q^r} R_q[A, B] \tag{31}$$

The idea behind these definitions is the following: For a given subspace Q^r, there is a certain direction q in Q^r for which the signal-to-signal ratio of the two vector sequences $\{a_k\}$ and $\{b_k\}$ is minimal. This direction corresponds to the worst direction q in the subspace Q^r in the sense that the energy of A is difficult to distinguish from the energy of B. This worst case of course depends upon the subspace Q_r. Among all r-dimensional subspaces, there must be at least one subspace where the worst case is better than all other worst cases. Hence the maximal minimal signal-to-signal ratio is in some sense the best of all worst cases : In the corresponding maximizing subspace, it can be guaranteed for all directions that the energy of A is at least MmR$[A, B, r]$ times larger than that of B. A similar explanation can be derived for the minimal maximal signal-to-signal ratio: it is the worst of all best cases considered over the r-dimensional subspaces. Remark that 3 elements are involved in the definition of MmR$[A, B, r]$(*resp.* mMR$[A, B, r]$):

- Some motivation must be available to determine a suitable r.
- In each possible r-dimensional subspace, there is a worst (best) direction q that minimizes (maximizes) $R_q[A, B]$
- That r-dimensional subspace is selected in the ambient space $\mathcal{R}^m$ where the worst (best) case is best (worst).

The results of theorem 7 and corollary 2, lead immediately to a computational procedure to compute MmR$[A, B, r]$ and mMR$[A, B, r]$, based upon the GSVD of the matrix pair A, B and the fact that the signal-to-signal ratio of two signals is invariant under linear transformations (invariance theorem 2).

Corollary 3 *Consider two m-vector sequences* $\{a_k\}, \{b_k\}$ $\;k = 1, \ldots, n$, *associated* $m \times n$ *matrices A and B and integer number* $0 < r \leq m < n$ *Consider the GSVD of A and B as in (22):*

$$A = X^{-1} \cdot D_a \cdot U_1^t$$
$$B = X^{-1} \cdot D_b \cdot V_1^t$$

Let (α_i/β_i) *be the generalized singular values of A and B, arranged in non-increasing order. Denote by* x_i^t *the i-th row vector of X. Then :*

MmR$[A, B, r] = \alpha_r/\beta_r$ *The corresponding subspace is generated by the first r row-vectors* $[x_1 \cdots x_r]$ *of X.*

mMR$[A, B, r] = \alpha_{m-r+1}/\beta_{m-r+1}$ *The corresponding subspace is generated by the last r row-vectors* $[x_{m-r+1} \cdots x_m]$ *of X.*

Proof: Define the linear transformation $T = D_B^{-1} \cdot X$. Use the invariance theorem 2 and theorem 7, relating the signal-to-signal ratio to the oriented energy, in order to find that:

$$\text{MmR}[A, B, r] = \max_{(T^{-1})^t Q^r \subset \mathcal{R}^m} \; \min_{q' \in (T^{-1})^t Q^r} E_{q'}[D_B^{-1} \cdot D_A \cdot U_1^t, V_1^t]$$

and

$$\text{MmR}[A, B, r] = \min_{(T^{-1})^t Q^r \subset \mathcal{R}^m} \; \min_{q' \in (T^{-1})^t Q^r} E_{q'}[D_B^{-1} \cdot D_A \cdot U_1^t, V_1^t]$$

From the Courant-Fischer minimax and maximin characterization properties (Corollary 1, 5/ and 6/), the result follows after back transformation with $T^{-1} = X^{-1} \cdot D_B$. □

4.3 Numerical considerations

Given an $m \times n$ matrix A with $n \geq m$ and a $p \times n$ matrix B, it can be proven that there exists a non-singular $n \times n$ matrix X such that both $X \cdot (A \cdot A^t) \cdot X^t$ and $X \cdot (B \cdot B^t) \cdot X^t$ are diagonal [6, p. 314]. The great value of the GSVD is that these diagonalizations can be achieved without forming the Grammians $A \cdot A^t$ and $B \cdot B^t$, hence avoiding the numerically dangerous implicit squaring, which can lead to a loss of accuracy caused by the limited machine precision. As observed in [6], the proof of the GSVD theorem, which makes use of the C-S decomposition, is constructive since it can be shown how to stably compute the C-S decomposition [16]. Another possible implementation is considered in [8].

Moroever, the GSVD provides a structured algorithm to analyse the oriented signal-to-signal distribution of two vector sequences. It computes directly the several extrema (the generalized singular values) but also the corresponding extremal directions (the rows of the matrix X). Of course, any strategy that first orthonormalises the vector sequence $\{b_k\}$ via a linear transformation T and then considers the oriented energy distribution of the matrix $T \cdot A$ will work for well conditioned vectorsequences $\{b_k\}$ As an example, consider first the R-Q factorization of the matrix B, define $T = R^{-1}$ and then study the oriented energy distribution of $R^{-1} \cdot A$ by computing its singular value decomposition . This requires a R-Q factorization, a matrix inversion and a singular value decomposition. Moreover, several additional matrix matrix multiplications are necessary in the backsubstitution. The big advantage of the GSVD is that it replaces these three algorithms and matrix multiplications by one, which is numerically reliable and can more easily handle the near-singularity case, where B is "almost" rankdeficient.

5 Applications and examples

In this section, several examples are presented in order to illustrate the practical significance of the above derived framework of oriented energy and oriented signal-to-signal ratio. In section 5.1., the oriented energy distribution and the technique of prewithening are considered. In section 5.2., the concept of total linear least squares is formalized using the complexity/misfit approximative modeling framework of [18] and the oriented energy concept. In section 5.3., it is shown how a lot of factor analysis like modeling problems lend themselves very naturally to a formulation in terms of oriented signal-to-signal ratios.

5.1 The oriented energy distribution of stochastic vector sequences.

Consider a stochastic process, consisting of a m-vector sequence $\{b_k\}, k = 1, \ldots, n$. The process is assumed to be ergodic and the elements b_{ij} of the associated matrix B are independenlty distributed. Under these assumptions, the Grammian $B \cdot B^t/n$ is an estimate for the second-order joint moment matrix of the vector stochastic process and for increasing n , it will tend to become a symmetric Toeplitz matrix: By applying theorem 5, one can replace the actual time realization contained in the matrix B, by a sequence of m m-vectors having the same oriented energy distribution. As a special case, assume that the components of the vector process are independently and identically distributed with first and second order moments m_1 and m_2^2. Then the covariance matrix will have the following specific symmetric

Toeplitz structure :

$$\begin{matrix} B \cdot B^t \\ m \times m \end{matrix} \approx n \begin{bmatrix} m_2^2 & m_1^2 & m_1^2 & \dots & m_1^2 \\ m_1^2 & m_2^2 & m_1^2 & \dots & m_1^2 \\ \dots & \dots & \dots & \dots & \dots \\ m_1^2 & m_1^2 & m_1^2 & \dots & m_2^2 \end{bmatrix}$$

The nice fact about this matrix is that its eigenstructure is straightforward to compute: There are $m-1$ eigenvalues $n \cdot (m_2^2 - m_1^2)$. There is one largest eigenvalue $\gamma_1 = n \cdot [(m-1) \cdot m_1^2 + m_2^2]$ with corresponding eigenvector $u_1 = 1/\sqrt{m}(1 \quad 1 \cdots 1)^t$. Via the equivalence theorem 5, a compact presentation of a stochastic sequence with the above mentioned characteristics, is the vector-sequence:

$$[u_1 \cdot \sigma_1 \quad u_2 \cdot \sigma_2 \cdots u_m \cdot \sigma_m]$$

where $u_1 = 1/\sqrt{m}(1\ 1 \cdots 1)^t$ and the vectors $u_j, j = 2, \dots, m$ are an arbitrary orthonormal set of vectors orthogonal to u_1. Moreover, $\sigma_1 = \sqrt{n[(m-1)m_1^2 + m_2^2]}$ and $\sigma_j = \sqrt{n(m_2^2 - m_1^2)}$, $j = 2, \dots, m$. In words, the oriented energy distribution of a stochastic sequence which is ergodic and which has identically independently distributed elemenents, is isotropic except for one principal direction along the direction $(1\ 1\ \dots\ 1)$ in the first orthant in which the energy is larger. Clearly, it is in this direction that this stochastic disturbance sequence will have the largest corrupting influence on any 'exact' signal sequence. A special case is of course obtained if the first moment $m_1 = 0$. In this case, the second order joint moment matrix reduces to a diagonal matrix and the oriented energy distribution is isotropic. This is the case in a lot of engineering applications, where it is assumed that the noisy vector sequence consists of independent identically normally distributed zero mean random variables. This assumption is quite natural if the central limit theorem is invoked to argue that the macroscopic effect of noise is due to the superposition of a lot of independent microscopic causes, and if all offsets are eliminated a priori.

Now consider the situation of an 'exact' m-vector signal contained in the $m \times n$ matrix A, which is of rank $r < m < n$ and assume that only the $m \times n$ matrix $C = A + B$ is observable, where B is some stochastic noisy sequence, with a priori known (for instance from experiments) second order statistics, that are summarized in an equivalent m-vector sequence of m vectors, that are the columns of the matrix $U_b.S_b$, where U_b is $m \times m$ orthonormal and S_b is $m \times m$ positive definite diagonal. Moreover, assume that the row spaces of A and B are orthogonal (which under mild conditions [5] is the case for large overdetermination n/m). Then, it is not difficult to see that the generalized singular value decomposition of the matrix pair $[A, U_b.S_b]$ or equivalently the singular value decomposition of the matrix $S_b^{-1}.U_b^t.A$ will provide the subspaces of maximal minimal signal-to-signal ratio, which are the subspaces in which the vector sequence A can best be distinguished (in the average) from the perturbing influence of the fuzzy sequence B. The generalized singular values are appropriate tools to make meaningful decisions for the correct dimension, because the transformation $S_b^{-1}.U_b^t$ has (under mild conditions) caused a noise threshold in the singular values equal to 1. It can be shown in a straightforward way that this technique is nothing else than the 'classical' prewhitening technique, in which the data are transformed via a so-called Mahalanobis transformation [10] in such a way that the noise covariance matrix equals the identity matrix. Hereto, assume that Σ_c is the measurement sample covariance matrix and that Σ_b is the noise covariance matrix. Then, the problem to be solved is the generalized symmetric eigenvalue problem

$$\det(\Sigma_c - \gamma.\Sigma_b) = 0$$

The whitening transformation then consists in converting this expression into the eigenvalue problem, under the assumption that Σ_b is non-singular :

$$\det(\Sigma_b^{-\frac{1}{2}}.\Sigma_c.\Sigma_b^{-\frac{1}{2}} - \gamma I) = 0$$

where $\Sigma_b^{\frac{1}{2}}$ is any symmetric square root of Σ_b. Via lemma 5 this establishes of course the link with the oriented signal-to-signal ratio framework and the generalized singular value decomposition. In a certain sense, these are even more general, since even (almost) rank deficient Σ_b can be allowed without numerical complications.

Another example is the use of GSVD in prewhitening the data for the estimation of parameters in a general Gauss-Markov linear model [7]

5.2 Total Linear Least Squares

In [18] a conceptual framework is developed in which the modeling problem is translated into an approximation context based upon the paradigm of low complexity and high accuracy models. The key concepts in this approach are the complexity of a model and the misfit between a model and the observations. Approximate modeling then consists of implementing the principle that either the desired optimal model is the least complex one in a given model class which approximates the observed data up to a preassigned tolerated misfit, or that it is the most accurate model within a preassigned tolerated complexity level. A particularly simple example is the total linear least squares approach [6,15] which consists of fitting a linear subspace to a finite number of points . Consider an $m \times n$ matrix A ($n \gg m$) containing n measurements on a m-vector signal. Denote by a_i its i-th column. Let Q^r be a r-dimensional subspace of $\mathcal{R}^m$ then, the complexity is defined as :

$$c(r) : Q^r \rightarrow C = [0,1] : c(r) = \dim(Q^r)/m = r/m$$

Suppose that we are looking for linear relations among the m measurement channels of the form $x^t.A = 0$ Define the error between the data and the law $x^t.A = 0$ as:

$$d(A,x) = \frac{\sqrt{[\frac{1}{n}\sum_{i=1}^{n}(x^t.a_i)^2]}}{\|x\|}$$

and the misfit associated with the r-dimensional subspace Q^r as:

$$\epsilon(A,Q^r) = \max_{x \perp Q^r} d(A,x)$$

Then, we have the following theorem [18] :

Theorem 8 *Let $\frac{1}{\sqrt{n}}A = U.\Sigma.V^t$ be the SVD of the $m \times n$ matrix A of rank s ($s \leq m < n$) with singular values $\sigma_1 \geq \cdots \geq \sigma_s > 0$ and left singular vectors $u_i, i = 1,\ldots,m$. The unique optimal approximate model Q^r with complexity $c(Q^r) = \frac{r}{m}$ and misfit $\epsilon(A,Q^r) = \sigma_{r+1}$ is an r-dimensional subspace where :*

- *If c_{adm} is the maximal admissible complexity, then :*
  - *if $\text{int}[m.c_{\text{adm}}] = 0$, $r = 0$ and $Q^r = 0$.*
  - *if $\text{int}[m.c_{\text{adm}}] \geq s$, $r = s$, $Q^r = \text{span}_{\text{col}}[A]$*

- *if* $\sigma_k > \sigma_{\text{int}[m.c_{\text{adm}}]+1}$, $r = k, Q^r = S_U^k$

- *If* ϵ_{tol} *is the maximal tolerated misfit, then :*
  - *if* $\epsilon_{\text{tol}} \geq \sigma_1$, $r = 0$ *and* $Q^r = 0$
  - *if* $\epsilon_{\text{tol}} < \sigma_s$, $r = s, Q^r = \text{span}_{\text{col}}[A]$
  - *if* $\sigma_k > \epsilon_{\text{tol}} \geq \sigma_{k+1}$, $r = k$ *and* $Q^r = S_U^k$

Proof : see [18] □

In this framework of approximate modeling, the appropriate rank r is determined from either an a priori fixed admissible complexity or a maximal tolerable misfit. Observe that these concepts readily reduce to the framework of oriented energy in that :

$$\begin{aligned} [d(A,x)]^2 &= E_x[A] \\ [\epsilon(A,Q^r)]^2 &= \min_{Q^r} \max_{q \in Q^r} E_q[A] \end{aligned}$$

where $p = m - r$. Hence the misfit is nothing else than a subspace of minimal maximal oriented energy . The authors conjecture that also the dynamical case for the identification of state space models developed in [18] can be translated in the oriented signal-to-signal ratio framework.

5.3 Factor-analysis like subspace methods

A lot of identification and modeling problems can be formulated in a factor-analysis like framework:

Given noisy measurements of an m-vector process which can be modeled as :

$$\underset{m \times 1}{x(t)} = \underset{m \times r}{Q(\theta)} \cdot \underset{r \times 1}{s(t)} + \underset{m \times 1}{n(t)} \qquad r < m$$

where $Q(\theta)$ contains the r linear independent so-called factor loadings, $s(t)$ are the source signals and $n(t)$ are the corrupting noise signals. The subspace generated by the columns of $Q(\theta)$ is called the factor loading subspace which is in one sense or another parametrized by unknown parameters θ. The task is then to estimate the parameters θ, given a priori knowledge of the second order statistics of the measurement noise and varying degrees of knowledge concerning the sensor response function.

Factor analysis is not that well reputed in the statistical community. The reason is that the loadings are undetermined and that only the subspace that they generate is well determined under appropriate conditions on the noise. Classically, attempts were undertaken to resolve this problem by fixing so-called structural zeroes, which corresponds to fixing a coordinate system in the loading subspace. The determination of the remaining non-zero components however frequently leads to an ill-conditioned parameter estimation problem, which explains (at least heuristically) the bad reputation of factor analysis. However, in modern applications (so-called subspace methods), the indeterminancy is immaterial because the properties (the parameters θ) are in a one-to-one correspondence with the generated subspace : Any set of vectors that generates a basis for this subspace, will allow to determine the parameters θ : a precise choice of a basis in that space is not important.

We will now present 3 applications : high resolution spectral analysis and sensor array processing techniques, the separation of fetal ECG and maternal ECG and realization of dynamical systems.

5.3.1 High resolution sensor array processing.

As is derived in [10,11] a factor analysis model can be used to model the arrival of narrowband signals impinging on an array consisting of sensor pairs that are separated by a fixed distance δ. With m sensor pairs and when the signals from sensor pair i are $x_i(t)$ and $y_i(t)$, the following model is appropriate when $d < m$ narrowband sources $s_j(t)$ are present [10,11] :

$$\underset{m \times 1}{x(t)} = \underset{m \times d}{A} \; . \; \underset{d \times 1}{s(t)} + \underset{m \times 1}{n_x(t)}$$

$$\underset{m \times 1}{y(t)} = \underset{m \times d}{A} \; . \; \underset{d \times d}{\Omega} \; . \; \underset{d \times 1}{s(t)} + \underset{m \times 1}{n_y(t)}$$

The d-dimensional columnspace of A is generated by the so-called steering vectors while the matrix Ω is a complex diagonal shift matrix that contains information on the phase shifts between the pairs of sensors from which the direction of arrival can be estimated. Schematically, this can be achieved as follows : Define $z(t)$ as

$$z(t) = \begin{bmatrix} A \\ A.\Omega \end{bmatrix} .s(t) + n_z(t)$$

Now store n consecutive samples $z(i), i = 1, \ldots, n$ in a $2m \times n$ matrix Z, n consecutive samples $s(i)$ in a $d \times n$ matrix S and the noise in a $2m \times n$ matrix N

$$Z = \begin{bmatrix} A \\ A.\Omega \end{bmatrix} .S + N$$

If the second order noise statistics of $n_z(t)$ are known, one can obtain an equivalent m-vector sequence of m vectors $U_m.\Sigma_m$ from the SVD of the sample covariance $N.N^t$.

Using the insights of [10,11], one can then prove that the best approximation for the loading subspace follows from the maximal minimal signal-to-signal ratio and corresponding subspace that can be estimated from the GSVD of the matrix pair $[Z, U_m\Sigma_m]$ The number of sources can be estimated from the generalized singular values via rank determination tests [1,2,9,10,12] while the imposed shift-structure of the corresponding subspace of maximal minimal signal-to-signal ratio can be exploited to determine the angles of arrival of the signals, hence the location of the sources.

5.3.2 Realization of dynamical systems.

Realization theory of dynamical systems reduces to the determination of the matrices of a state space model for a linear finite dimensional systems starting from (possibly noisy) measurements of its Markov parameters. As it is well known, the relation between the state space model with states x_k, inputs u_k and outputs y_k

$$\underset{n \times 1}{x_{k+1}} = \underset{n \times n}{A} \; . \; \underset{n \times 1}{x_k} + \underset{n \times m}{B} \; . \; \underset{\times 1}{u_k}$$
$$\underset{l \times 1}{y_k} = \underset{l \times n}{C} \; . \; \underset{n \times 1}{x_k}$$

and its Markov - parameters, is $H_k = C.A^{k-1}.B$. Classically, the realization of the model from the H_k proceeds by the following algorithm (there are several variants [4] and references therein) :

- Construct a sufficiently large block Hankel matrix with the H_k.
- Determine its rank via SVD. The rank decision results in an estimate of the minimal state dimension n
- The matrices A, B , C are then realized up to a similarity transformation:
  - B and C can be read off from certain block- and column rows in the SVD
  - The matrix A follows from the shift invariant structure of the column space of the block Hankel matrix.

When the measurements are noisy, the following novel realization framework, based upon the oriented signal-to-signal ratio concept, is appropriate to determine an estimate of the matrix A up to a similarity transformation.

Construct a sufficiently large block Hankel matrix H and partition it in two blocks H_1 and H_2

$$H = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}$$

The reader may now wish to verify that the best approximation for the appropriate shift invariant subspace follows from the GSVD of the matrix pair $[H_1, H_2]$. The generalized singular values allow to estimate the minimal order n. The corresponding subspace of maximal minimal signal to signal ratio contains information on the minimal poles of the system via an imposed shift structure [4]

5.3.3 The separation of fetal ECG from maternal ECG.

In this biomedical application, the cutaneous measurements (typically 6 to 9 channels) are contained in a vector $m(t)$ which is modeled as :

$$m(t) = T.s(t) + n(t)$$

where the signal $s(t)$ corresponds to the sources (electrical activity of the heart of mother and fetus), $n(t)$ is the noise, and the columns of T are the so-called lead vectors (typically 2 for the fetus, 3 for the mother)[14,19]. Under certain conditions (orthogonality of source signals, certain statistical conditions on the noise, placement of electrodes), it can be verified that one singular value decomposition suffices to determine the factor loading subspace generated by the lead vectors of the fetus. This allows to project the measurements into this subspace, hence eliminating almost completely the maternal ECG. The conceptual framework is provided by the oriented energy distribution of the vector signal $m(t)$. In [17], the same problem is solved using an approach that can be interpreted in the oriented signal-to-signal framework. By visual inspection, two matrices A and B are constructed from the measurements $m(t)$. A window in time is selected visually so that A contains only fetal ECG complexes. Another window is chosen so that B contains only maternal ECG complexes. The loading subspace generated by the fetal lead vectors is then nothing else than the subspace of maximal minimal oriented signal-to-signal ratio. This subspace and its dimension can be computed from the GSVD of the matrix pair $[A, B]$.

6 Conclusions

Two important concepts have been defined : The oriented energy distribution of a vector sequence and the oriented signal-to-signal ratio of two vector sequences. For the former, the singular value decomposition is the appropriate quantification tool while for the latter the general singular value decomposition applies. Both allow a numerically robust implementation. Conceptually important properties have been analysed. With some clarifying examples, the practical significance of the framework in the formalization of so-called subspace methods has been demonstrated. Future work will concentrate around a further theoretical development of the concepts and research for fast algorithms for the required factorisations including up - and down - dating.

Acknowledgement : The authors wish to thank Dr.Ir.Tan Shaohua and Ir.Lieven Vandenberghe for their help in typesetting this manuscript.

References

[1] Akaike H. Information theory and an extension of the maximum likelihood principle, In Proc. 2nd Int.Symp.Inform.Theory, pp.267-281, 1973.

[2] Anderson T.W. Asymptotic theory for principal component analysis, Ann. Math. Statist., 34:122-148, 1963.

[3] Chatfield C. , Collins A. Introduction to multivariate analysis. Chapman and Hall Ltd., London 1980.

[4] 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)

[5] Vandewalle J, De Moor B. A wide variety of applications of singular value decomposition in identification and signal processing. This volume .

[6] Golub G., Van Loan C. Matrix Computations. Johns Hopkins University Press, 1983.

[7] Hammarling S. The Numerical Solution of the General Gauss-Markov Linear Model, NAG-Technical report TR2/85, October 1985.

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

[9] Rissanen J. Modeling by shortest data description, Automatica, 14:465-471, 1978.

[10] Roy R. ESPRIT : Estimation of Signal Parameters via Rotational Invariant Techniques. Ph.D. Dissertation, Stanford University, August 1987.

[11] Roy R., Paulraj A., Kailath T. Estimation of Signal Parameters via Rotational Invariance Techniques - ESPRIT. In Proc. IEEE ICASSP pp.2495-2498, Tokyo, Japan, 1986.

[12] Scharf L.L., Tufts D.W. Rank reduction for modeling stationary signals. IEEE Trans. on Acoustics, Speech and Signal Processing, Vol.ASSP-35,no.3, March 1987.

[13] Staar Jan. Concepts for reliable modeling of linear systems with application to on-line identification of multivariable state space descriptions. PhD thesis, Esat - Katholieke Universiteit Leuven, 1982.

[14] Vanderschoot J., Vandewalle J., Janssen J., Sansen W., Vantrappen G. Extraction of weak bioelectrical signals by means of SVD. Proc. of 6th Int. Conf. on Analysis & Optimization of Systems, Nice, June 1984, Springer Verlag, Berlin, pp.434-448.

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

[16] Van Loan C. Computing the C-S and the Generalized Singular Value Decompositions, Numerische Mathematik, 46:479-491,1985.

[17] Van Oosterom A. Alsters J. Removing the maternel component in the Fetal ECG using singular value decomposition. Electrocardiography '83, I. Rattkay-Nedecky and P.MacFarlane, Eds. Amsterdam, The Netherlands. Excerpta Medica, 1984, pp. 171-176.

[18] Willems J.C. From time series to linear systems. Part I - II - III . Automatica, 1987.

[19] Vanderschoot J., Callaerts D., Sansen W., Vandewalle J., Vantrappen G., Janssens J., "Two methods for optimal MECG elimination and FECG detection from skin electrode signals", IEEE Trans. on Biomedical Engineering, Vol. BME-34, No. 3, pp. 233-243, March 1987.