86-12C

THE UNIQUENESS VERSUS THE NON-UNIQUENESS PRINCIPLE IN THE IDENTIFICATION OF LINEAR RELATIONS FROM NOISY DATA.

Bart De Moor*;Joos Vandewalle.

* research sponsored by the IWONL.

ESAT-Katholieke Universiteit Leuven,
Kardinaal Mercierlaan 94 / 3030 Heverlee/ Belgium /tel 016 220931
telex 25941 elekul.

1. Introduction.

The objective of this paper is to provide some new insights in the analysis of the basic modelling problem in the simplest case: The identification of linear relations from noisy data. The central idea is the nonuniqueness principle which states that the solution of a noisy identification problem is intrinsically non-unique. Any uncertainty in the data implies that no particular model can be identified with certainty. In a situation of uncertainty (noise), it is unnatural to have mathematical formulations that lead to unique results. We demonstrate how classical treatments of the problem (least squares, total linear least squares (orthogonal regression), 'white noise' approach lead to a unique solution, because additional a priori assumptions are imposed, as unverifiable prejudices, that force a uniqueness.

For the general case, were the number of a priori assumptions is minimal, Kalman(KAL) obtained interesting results for the case that maximally one linear relation is present in the data. In this case, the solution is not unique, and the solution set is the convex simplex generated by all the least squares solutions to the linear problem. This result can be proven with the classical Perron-Frobeniustheorem for positive matrices (GAN).
We believe however, that the geometrical approach, presented in the companion paper [DEM1], will lead to a general solution for any number of linear relations.
In the companion paper, the geometrical concepts of orthant and null invariance are introduced. They lead to the conjecture that the complete solution set in the general case is described geometrically as a collection of convex polyhedral sets in orthants, of which the vertices are least squares solutions and allowed orthant null invariant vectors with as many zeros as possible.
In this paper, the different conjectures about the solution set are analysed. In order to illustrate the novel geometrical approach to the problem, a clarifying numerical example will be presented.
We will use the following notations and abbreviations.

A	mxn matrix (m>n)	PD	positive definite
A^t	transpose	NND	nonnegative definite
r(A)	rank of A	OI	orthant invariant
cor(A)	corank of (A) = n - r(A)	NI	null invariant
		AONI	allowed orthant null invariant
$\hat{A}$	exact matrix		
$\tilde{A}$	pure noise matrix	LLS	linear least squares

2. Problem statement and the non-uniqueness principle.

Let us consider a sequence a_i; i=1,..,n of data vectors in $\mathbb{R}^n$, stored as rows in a mxn matrix A. The problem is to find all possible n-vectors that describe a linear relation between the columns of A :

$$\underset{m\times n}{A} \cdot \underset{n\times 1}{x} = 0 \quad \text{or} \quad \underset{n\times n}{\Sigma} \cdot \underset{n\times 1}{x} = 0$$

where Σ is the sample or population covariance matrix of the vector process. Generically, when uncertainties (noise) are present on the data, the measurement covariance matrix Σ will satisfy corank(Σ)=cor(Σ)=n-rank(Σ)=0 . Hence no trivial solution can be obtained directly from the data. This means that the data must first be modified in order to obtain a solution. How to do this, is implied in the following fundamental noise assumptions.

Assumption 1: Cor(Σ)=0.
Assumption 2: The uncertainties are additive.
Assumption 3: The noise is independent of the data.
Assumption 4: The noise covariance matrix $\tilde{\Sigma}$ is diagonal.
Assumption 5: All data columns of A (all diagonal elements of $\tilde{\Sigma}$) are perturbed by noise

The problem formulation then becomes: Given a nxn symmetrix, PD matrix Σ. Find all NND diagonal matrices $\hat{\Sigma}$ and solutions x so that:

1/ $(\Sigma - \hat{\Sigma}).x=0$
2/ $\Sigma - \hat{\Sigma}$ is NND.
3/ cor($\Sigma - \hat{\Sigma}$) is maximal.

The maximality requirement for the corank is crucial. It is the maximal number of linear relations that is implicitely present in the data under the given assumptions. The maximal corank as such is the only invariant of the problem that can be uniquely identified from the data.
As demonstrated in the companion paper [DEM1], the problem can be attacked by two different approaches. In the noise space approach, one reformulates the problem as an eigenvalue problem for the matrix $\Sigma^{-1}.\tilde{\Sigma}$ (KAL). In the solution space characterization, the concepts of orthant and null invariance are introduced to characterize the canonical solutions, wherefrom the maximal corank can be determined. Both approaches lead to the same fundamental conclusion: Even in the simplest case where maximally one linear relation is present in the data, the solution for the vectors x, describing the linear relations, is non-unique. It is then a convex simplex in the n-dimensional space. In [DEM1], some mathematical tools are provided that seem to justify the conjecture that in the general case the solution is a collection of convex polyhedral sets in several orthants of the n-dimensional space. Before clarifying this general conjecture, we first illustrate where some frequently used noise schemes of identification techniques fail to recognize the non-uniqueness of the solution.

3. The uniqueness principle in 'classical' identification schemes.

In this section three classical identification schemes are considered. They all show a very clear disadvantage: By some unverifiable 'prejudice'(KAL), the solution is forced to be unique.

3.1. The linear least squares solutions (LLS).

A non conventional way to present the LLS solutions, is the following theorem (KAL),[DEM1].

> Theorem: Consider a nxn PD matrix Σ. Then the i-th LLS solution of $\Sigma.x=0$ is the i-th column s_i of Σ^{-1} with diagonal noise covariance matrix $\tilde{\Sigma}_i$=diag(0,..,0,1/s_{ii},0,..,0) (1/s_{ii} in the i-th position)
>
> where s_{ii} is the i-th diagonal element of Σ^{-1}.
> Hence: $(\Sigma-\tilde{\Sigma}_i).s_i=0$; $(\Sigma-\tilde{\Sigma}_i)$ is NND ; cor($\Sigma-\tilde{\Sigma}_i$)=1.

Proof: see [KAL],[DEM1].

86-12

De Moor B., Vandewalle J., ``The uniqueness versus the non-uniqueness principle in the identification of linear relations from noisy data", in *Proc. of the 25th IEEE Conference on Decision and Control*, Athens, Greece, vol.3, Dec. 1986, pp. 1663-1665., Lirias number: 181688.

Observe that each LLS solution is compatible with the first 4 fundamental noise assumptions. This theorem is important because it shows how the LLS solutions are determined by the data. The noise is modelled via a rank one covariance matrix $\tilde{\Sigma}_i$. This implies that only the i-th column of the data matrix A is considered to be noisy, while all others are considered to be noisefree. This i-th column is projected into the column space formed by the 'noisefree' columns. The noise $\sigma_i = 1/s_{ii}$ is the 'energy' needed to make the matrix Σ singular. So, unfortunately, in many cases, the LLS solutions do not provide a satisfactory answer. Always, a unique solutionn is provided to the noisy identification problem.

3.2 The total linear least squares approach - Orthogonal Regression.

In an attempt to perturbe all the data columns in order to obtain a solution, instead of modifying only one column, Golub [GOL] formulated the total linear least squares approach and solved it via the singular value decomposition, based on the classical result of Eckart and Young [ECK]. The problem formulation has a long story with an early contribution already in 1878 (see [AND] for a list of references). It was reinvented many times and is sometimes known as orthogonal regression. Using assumptions 1,2,3,of the fundamental noise assumptions, the problem formulation in the covariance formalism becomes:

Given a nxn PD matrix Σ, find $\tilde{\Sigma}$ such that $\tilde{\Sigma} = \Sigma - \tilde{\Sigma}$ is NND and the 'perturbation energy', defined as :

$$\|\tilde{\Sigma}\|_F = \|\Sigma - \tilde{\Sigma}\|_F$$

is minimal (F denotes the Frobeniusnorm of a matrix)

The solution is straightforward, using the classical result of Eckart and Young [ECK]. Under the generic assumption that the smallest eigenvalue d_n of Σ is simple, take $\tilde{\Sigma} = d_n.v_n.v_n^t$ where v_n is the eigenvector corresponding to d_n which is also the unique solution vector:

$$(\Sigma - \tilde{\Sigma}).v_n = 0$$

Clearly, v_n is the direction that minimizes the scatter [AND].

Observe that the noise covariance matrix $\tilde{\Sigma}$ is not diagonal and that $cor(\tilde{\Sigma})=n-1$. Hence, there exist n-1 independent linear relations between the noise columns, which are all proportional to v_n. Hence, the solution may look somewhat more sophisticated than the LLS scheme but the noise scheme is a very structured one ($r(\tilde{\Sigma})=1$!!) and highly conflicts with ones intuitive appreciation of noise. However, extensive simulations on numerical examples [VANH] showed that the total least squares solution is always closer to the exact solution than is any LLS. This is not surprising if one considers the following classical noise scheme.

3.3. The 'white noise' approach.

When all data columns are perturbed by uncertainties it is very natural to try to modify all data columns to get a solution. The easiest non trivial noise scheme (in the sense that $r(\tilde{\Sigma})=n$, no linear relation exists between the noise columns) is to put $\tilde{\Sigma} = d_n.I_n$, where d_n is the smallest eigenvalue of Σ, with eigenvalue decomposition $V.D.V^t$. Then obviously :

$$\Sigma - \tilde{\Sigma} = V.D.V^t - d_n.I_n = V.(D - I_n).V^t$$

and the unique solution is $x=v_n$, the eigenvector corresponding to d_n. Hence all columns are modified with an equal amount of energy d_n and the noise variables are uncorrelated. The solution v_n is exactly the same as the total least squares solution. This of course explains the better results of total least squares compared to ordinary least squares. The total least squares approach does not provide a very realistic noise covariance matrix $\tilde{\Sigma}$, while the 'white noise' covariance matrix $\tilde{\Sigma} = d_n.I_n$ is more correct. However, despite of the uncertainties present in the data, still a unique solution is provided. This solution however, is an element of the global solution set we are now going to describe.

5. Non-unique solutions for the noisy identification problem.

Kalman's main result obtained in [KAL], using the celebrated Perron - Frobeniustheorem for positive matrices can be stated as follows:

Theorem: If Σ is a given nxn PD covariance matrix and Σ^{-1} is sign similar to a strictly positive matrix, then the solution set of the noisy identification problem $(\Sigma - \tilde{\Sigma}).x=0$ is the convex simplex of which the vertices are the n least squares solutions. In this case, for each allowed noise matrix $\tilde{\Sigma}$, there is maximally one solution (max.cor=1).

Σ^{-1} is sign similar to a positive matrix if there exists a sign matrix $E=diag(\pm 1)$ so that $E.\Sigma^{-1}.E$ is strictly positive. The columns of $E.\Sigma^{-1}.E$ are then the vertices of the convex solution simplex. This important result illustrates very clearly the 'non-uniqueness' principle. If Σ^{-1} is sign similar to a positive matrix, then there exist infinitely many diagonal NND covariance matrices $\tilde{\Sigma}$ so that $\Sigma-\tilde{\Sigma}$ is singular and NND. For each $\tilde{\Sigma}$ there is a solution x satisfying $(\Sigma-\tilde{\Sigma}).x=0$. It is a most unexpected fact from the point of view of classical statistics, that a quasi linear object (a simplex) measures the uncertainty inherent in the data., relative to the question of identifying a single linear relation (max cor = 1).

The difference in the solution characterization between the algebraic approach in [KAL] and that proposed in [DEMI], is that Kalman fixes the conditions for a diagonal NND covariance matrix Σ to be allowed. This goes via an eigenvalue reformulation. In [DEMI] however, not the matrices $\tilde{\Sigma}$ are investigated, but on the contrary, the necessary conditions for a vector x to be a solution vector are derived. This leads quite straightforward to the novel concepts of orthant and null invariance.

For every orthant null invariant (ONI) vector x, there exists a diagonal matrix $\tilde{\Sigma}_x$ such that $\Sigma.x=\tilde{\Sigma}_x.x$. An ONI vector x is allowed is $\Sigma-\tilde{\Sigma}_x$ is NND. It is expected that this geometrical framework will lead to a complete global solution of the problem even for the max cor > 1 case. In fact, there are strong indications, confirmed by numerous experiments, that the following conjectures are correct:

Conjecture 1: On the convex solution sets in orthants.
The solution set (for any max.cor case), is a collection of convex polyhedral sets in orthants. The vertices of these sets are the least squares solutions and the allowed orthant null invariant vectors with as many zeros as possible.

Conjecture 2: On the allowed convex combinations of noise matrices.
Let a,b be vertices of the solution set in orthant E and let $\Sigma.a=\tilde{\Sigma}_a.a$, $\Sigma.b=\tilde{\Sigma}_b.b$ where $\tilde{\Sigma}_a=diag(\alpha_i)$, $\tilde{\Sigma}_b=diag(\beta_i)$ and $(\Sigma-\tilde{\Sigma}_a)$, $(\Sigma-\tilde{\Sigma}_b)$ are NND. Then define a convex combination $0 \leq \tau \leq 1$, $x = \tau.a+(1-\tau).b$ with corresponding noise matrix $\tilde{\Sigma}=diag(\sigma_i)$ determined by:

$$\Sigma.x = \tilde{\Sigma}.x = \tau.(\Sigma.a)+(1-\tau).(\Sigma.b)$$
$$= \tau.\tilde{\Sigma}_a.a+(1-\tau).\tilde{\Sigma}_b.b$$

so that $\sigma_i = d_i.\alpha_i + (1-d_i).\beta_i$
where $d_i = (\tau.a_i)/(\tau.a_i+(1-\tau).b_i)$
and $d_i \geq 0$ because a and b are in the same orthant.
Hence: $\tilde{\Sigma} = D.\tilde{\Sigma}_a + (I-D).\tilde{\Sigma}_b$ with $D=diag(\sigma_i)$.
The conjecture now is that $\Sigma - \tilde{\Sigma}$ is NND.

This conjecture shows how noise matrices should be combined to produce new noise matrices. This cannot happen in an arbitrary way but depends on corresponding solution vectors a and b.

Conjecture 3: On the maximal corank.
If there exist z linear independent AONI (allowed orthant null invariant) vectors, each with minimum z-1 zero's , then the maximal corank is at least z.

To illustrate these conjectures, let us now consider the following simple 3x3 example.

5. Example

$$\Sigma = \begin{bmatrix} 9 & 4 & 2 \\ 4 & 17 & 2 \\ 2 & 2 & 8 \end{bmatrix} \qquad \Sigma^{-1} = 1/1024 \begin{bmatrix} 132 & -28 & -26 \\ -28 & 68 & -10 \\ -26 & -10 & 137 \end{bmatrix}$$

The following table gives the results for the OI,ONI and AONI vectors with one zero in the $2^{n-1}=4$ first orthants and for the least squares solutions.

Orthant	OI-vectors	ONI ?	AONI ?	Noise=diag(σ_i)
(1 1 1)	All positive vectors	--	--	--
(1 1 -1)	(0;4;-1)	no	--	--
	(0;1;-2)	yes	yes	diag(0 13 7)
	(4;0;-1)	no	--	--
	(1;0;-2)	yes	yes	diag(5 0 7)
	LLS 3	yes	yes	diag(0 0 1024/137)
(1 -1 1)	(0;-2;17)	no	--	--
	(0;-1;2)	yes	yes	diag(0 13 7)
	(1;-1;0)	yes	yes	diag(5 13 0)
	(17;-4;0)	no	--	--
	LLS2	yes	yes	diag(0 1024/68 0)
(1 -1 -1)	(1;-1;0)	yes	yes	diag(5 13 0)
	(4;-9;0)	no	--	--
	(2;0;-9)	no	--	--
	(1;0;-2)	yes	yes	diag(5 0 7)
	LLS1	yes	yes	diag(1024/132 0 0)

The vertices of the convex polyhedral sets in the orthants can be chosen as in fig.1. where x1=(-1;0;2), x2=(0;-1;2), x3=(1;-1;0). It is easily checked that for $\tilde{\Sigma}$=diag(5 13 7), cor($\Sigma-\tilde{\Sigma}$)=2. The line segments x1-x2, x2-x3, x1-x3 contain solutions for cor=2. It can also be verified that any vector which is a convex combination of the vertices in one orthant is an AONI vector.

6.Conclusions.

In this paper, the non-uniqueness principle was illustrated for the simplest case of the modelling problem: The identification of linear relations from noisy data. Any uncertainty in the data implies that no particular model can be identified with certainty. It was shown how some classical treatments of the problem such as least squares, orthogonal regression and the 'white noise' approach, lead to a unique solution. Starting from some fundamental noise assumptions and the novel geometrical concepts of orthant and null invariance, three conjectures are stated that describe the general solution set as a collection of convex polyhedral sets in orthants. These conjectures are illustrated with an example.

7. References.

[AND] Anderson T.W. The 1982 Wald Memorial Lectures: Estimating Linear Statistical Relationships. The Annals of Statistics, 1984, Vol.12,no.1,1-45.

[DEM1] De Moor B., Vandewalle J. A geometrical approach to the maximal corank problem in the analysis of linear relations. to be presented at this 25th IEEE CDC conference , Athens, Dec. 1986.

[DEM2] De Moor B., Vandewalle J. On the use of the inverse of the covariance matrix in multivariate data analysis. Submitted for publication.

[ECK] Eckart C., Young G. The approximation of one matrix by another of lower rank. Psychometrika,Vol.1.,211-218,1936.

[GOL] Golub G.,Van Loan C.F. Matrix Computations. North Oxford Academic.J. Hopkins Univ.Press,1983.

[KAL] Kalman R.E. System identification from noisy data. Dynamical systems II. A.R. Bednarek and L. Cesari, editors, Academic Press 1982.

[VANH] Van Huffel S.,Vandewalle J. The use of total linear least squares techniques for identification and parameter estimation. Proc. 7th IFAC/IFORS Symp. on Ident.&System parameter estimation, july 3-7,1985, York,pp1167-1172.

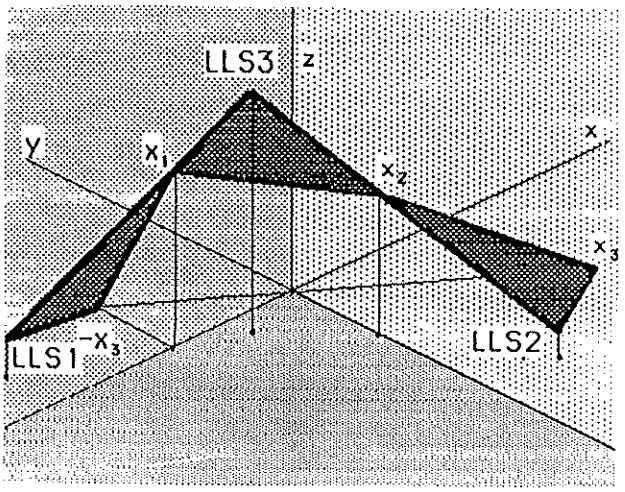


Fig.1. Geometrical representation of the convex polyhedral sets in orthants of the example of section 5. The vertices are the linear least squares solutions and the AONI vectors (x1,x2,x3) with one zero. The line segments x1-x2, x1-(-x3), x2-x3 are corank 2 solutions.

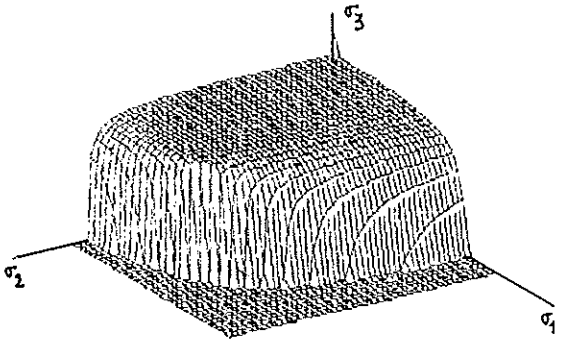


Fig.2. Geometrical representation of the solution in the 'noise' space. All possible values for the diagonal elements of the noise covariance matrix Σ such that $\Sigma - \tilde{\Sigma}$ is NND. The corank 2 solution is here one point: $\tilde{\Sigma}$ = diag(5 13 7).