THE UNCERTAINTY PRINCIPLE OF MATHEMATICAL MODELLING

Bart De Moor ; Joos Vandewalle
ESAT (Electronics, Systems, Automation, Technology)
Katholieke Universiteit Leuven
Kardinaal Mercierlaan 94
3030 Heverlee
Belgium
tel 32/16/220931 telex 25941 elekul fax 32/16/221855

Abstract

In this paper, a survey is presented of some recently obtained results in the problem of identifying linear relations from noisy data. Starting from a geometrically inspired definition of noise and linear relations, the mathematical problem is formulated. Existing identification techniques, such as (total) linear least squares, factor analysis and principal component analysis, are classified within the provided framework. Some existing results describing special cases are summarized. Finally, new geometrical concepts are defined and used to characterize the global solution set : the set of all linear relations that are compatible with the data and the modelling assumptions consists of a collection of convex polyhdral cones. This leads naturally to a fundamental uncertainty principle of mathematical modelling.

Keywords : Factor analysis, (total) linear least squares, maximal corank, Wilson - Ledermann bound, Frisch scheme, communalities.

1 Introduction

Identification of mathematical models from noisy data is one of the enduringly central problems in system theory and statistics and has profound implications in all branches of applied sciences such as electronics,mechanics, time series analysis, econometrics, biometrics, psychometrics... One of the basic questions is : *Can observed values of a finite family of variables be 'explained' by some underlying linear relations between the variables?* Using elementary linear algebra, the answer is trivial when the data are noise-free. The noisy problem however is highly non-trivial both from the conceptual as from the mathematical point of view. *Noise* may mean one or all of many things: inaccuracy of the model, measurement errors, unknown effects, non-linearities (when dealing with linear models), any causal or random factors which cannot be modelled, of which no further information is available, etc.... This indicates that the origin of the *noise* may not be clear: Is it due to our ignorance of infinitely precise data or equivalently, is 'noise' the manifestation of our lack of complete information which is caused by the limited precision of our measurement equipment? Or are the phenomena that we can observe in Nature so complex that they cannot be modelled by something as simple as linear relations, although good approximations might exist? The first question corresponds to the classical view of descriptive modelling: Nature operates consistently according to some universal laws and ours is the task to discover these. However, as is indicated in [23], the cornerstone of the philosophy of science is falsification rather than deduction. *Models and laws are postulated*, based on criteria like simplicity, esthetic appeal or by parallels with other disciplines and applications. It is only later that one finds out that they could also have been deduced from already existing knowledge. Therefore, the following statement is fundamental in that it really defines what is meant by noise (in any application !):

Noise = what is not explained by the model.

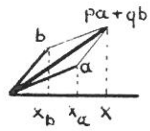
Hence, once the class of models has been fixed, at the same time, the notion of noise is well defined. If it is the user's desire to model the phenomenon under study by linear relations that are to be discovered in the data, one has immediately :

Noise is absence of linear relations.

Remains the question : How to define linear relations ? First, observe that linearity is not really a question of fact nor of evaluation, but a self-imposed limitation on the types of operations or devices that are to be used. One of those tools that are to be chosen is the metric. Many mathematicians believe that they have freedom in the choice of a metric for their mathematical model. This is true for pure mathematics (where the choice of norm can be dictated by pure 'intellectual' motivations such as for instance solvability of the problem) but it may no longer be true for the mathematical modelling of real processes. In a lot of cases, there are physical invariants that imply the use of a certain metric (as an example consider the theory of special relativity with its indefinite metric). Taking into account the necessary invariance principles that a metric should satisfy, there are strong indications that for the purpose studied in this paper - the identification of linear relations from noisy data -, the ordinary Euclidean metric is appropriate [17] as defined in the usual way: If x, y are real vectors with components x_i, y_i then the inner product is the real number $x_1y_1 + \ldots + x_ny_n$. Two vectors are orthogonal if their inner product equals 0 and the geometrical concept of an angle can be defined. Formally, one can now state that :

Orthogonality = absence of linear relations.

When 2 vectors are not orthogonal, at least part of one can be explained to be 'proportional' to the other. An intuitive representation of a linear relation is provided in the following vector scheme :



$$x = x_a + x_b$$
$$a \Longrightarrow pa \quad p \in \mathcal{R}$$
$$b \Longrightarrow qb \quad q \in \mathcal{R}$$
$$x \Longrightarrow px_a + qx_b$$

The vector x is 'linearly' related to the vectors a and b via their 'orthogonal' projections x_a and x_b upon the vector x. Let us now consider the problem of identifying linear relations between measured data. Suppose n variables are measured over m time instants. The measurements are aggregated in a $m \times n$ matrix A. It is assumed that the number of measurements exceeds the number of variables, $m > n$, so that the matrix A has more rows than columns. This assumption of overdetermination is of course necessary since otherwise the problem would be trivial. An existing linear relation would reveal itself via an n-vector x that belongs to the kernel of the matrix A: $Ax = 0$. The number of independent linear relations is indicated by the algebraic rank r of A. The *corank* of A is defined as $n - rank(A)$. The corank equals the number of linear relations between the variables. Now when the data are really measurements on some multi-channel phenomenon, generically there will be no linear relations between the data : $rank(A) = n$ or $corank(A) = 0$. A fundamental assumption is that the noise corrupts the data in an additive way (this is not only a matter of taste, but also of simplicity). The measured matrix A can then be written as: $A = \hat{A} + \bar{A}$, where $\hat{}$ denotes the obtained model of the data and $\bar{}$ denotes the obtained model of the noise. The noise variables cannot be linearly related with each other. Because, if they were, they would satisfy linear relations, hence contradicting the very definition of noise as absence of linear relations. Hence : $\bar{A}^t A$ = diagonal. There can also exist no linear relation between the noise and the exact data, hence : $\hat{A}^t\bar{A} = \bar{A}^t\hat{A} = 0$. Define the measured, exact and noise Grammians as:

$$\Sigma = A^t A \qquad \hat{\Sigma} = \hat{A}^t\hat{A} \qquad \bar{\Sigma} = \bar{A}^t\bar{A}$$

Obviously, Σ is positive definite and $\hat{\Sigma}$ is nonnegative definite. $\bar{\Sigma}$ is diagonal with nonnegative diagonal elemenents. The mathematical problem formulation of the identification of linear relations reduces to :

The Frisch scheme :
Given a positive definite $n \times n$ matrix Σ. Find all nonnegative diagonal matrices $\bar{\Sigma}$ and all n-vectors x such that:

1. $\hat{\Sigma} = \Sigma - \bar{\Sigma}$ is nonnegative definite
2. corank($\hat{\Sigma}$) is maximal
3. $\hat{\Sigma}x = 0$

The maximisation of the corank is essential: One is interested in the *maximum number* of linear relations. These linear relations are described by the vectors x. In fact, the maximisation of the corank is crucial, as we will see furtheron, in order to make the identification problem well defined. For several reasons, the above problem is called the *Frisch scheme*in honour of the 1969 Economics Nobel prize winner Frisch [10]. Of course, the problem looks very similar to what happens in statistics, where also concepts such as uncorrelatedness and statistical orthogonality are exploited to facilitate the identification problem. In fact, it was Wold who used the idea of regarding random variables as elements of a metric space with the distance between two elements as the variance of their difference. This geometric interpretation made it natural to interprete least squares estimation as projection onto a subspace. In the framework of stochastic processes as introduced by Kolmogorov [12], it is possible to split a process in a unique deterministic process and one that can be written as linear combinations (moving average) of a white noise process, which itself is a process with uncorrelated components. Both processes are uncorrelated. This is the so-called Wold decomposition and it may be interesting to mention that Wold was influenced by the work of Frisch [12]. There is indeed a big similarity between the characteristics of the Frisch scheme and the Wold decomposition. However, the Frisch scheme provides a conceptual framework, based upon the Euclidean inner product, that takes explicitly into account the double finiteness of the number of data (finite number of sample realizations and finite number of data in one realization). Hence, no so-called statistical 'sampling' problems occur. Moreover, as will be shown, this formalization leads in a straightforward way to a quantification of what could be called the *uncertainty principle of mathematical modelling*.

This paper is organised as follows: First, some classical identification schemes will be discussed in terms of their properties with respect to the Frisch scheme (section 2). Attention will be paid to the characterization of both, all diagonal matrices $\bar{\Sigma}$ and all vectors x that fulfill the requirements of the Frisch scheme (section 3). It will be discussed how the 'volume' of the polyhedral cones that characterize the solutions, are measures of the incompatibility of the data with the imposed linear static model. Moreover, the polyhedral cones serve as a characterization of the fundamental non-uniqueness of the solutions to the problem. In section 4, we discuss a new geometrical framework in which the requirements of the Frisch scheme are discussed one by one. New geometrical notions are introduced such as orthant and null invariance. It is expected that this framework will lead to a general solution of the Frisch scheme. A major drawback of the Frisch scheme will be discussed in section 5. As a function of n, there is an upper bound (the Wilson-Ledermann bound) on the generically achievable maximal corank of the problem, even when some other information as for instance the singular values strongly suggest a higher corank. This solves at the same time some important stability questions concerning the Frisch scheme. Some conclusions can be found in section 6.

2 Remarkable relations with classical identification schemes

In this section, we will analyse some 'classical' linear identification schemes with respect to the requirements of the Frisch scheme. For a unification theorem, in which it is shown that linear least squares and total linear least squares are special cases of one and the same identification approach, the reader is referred to [8].

2.1 Linear least squares

The classical least squares approach applied to the available $m \times n$ data matrix A $(m > n)$ would proceed as follows:

- Choose a column a^k of the matrix A and denote the remaining $m \times (n-1)$ matrix by A_k.
- Regress all remaining variables with respect to the chosen one. This is achieved by the classical expression:

  $$x_{LLSk} = (A_k^t A_k)^{-1} A_k^t a_k$$

  where x_{LLSk} represents the linear relation obtained via this least squares solution.
- Obviously, there are n such least squares solutions.

It is not so difficult to prove the following observation:
Theorem 1 Least squares solutions
The columns of the matrix $S = \Sigma^{-1}$ which is the inverse of the data Gramian Σ, contain the n linear least squares solutions (each up to a normalizing constant). Let s^k be the k-th column of S with k-th element s_k^k and $\bar{\Sigma}_k$ the diagonal matrix with only 1 nonzero element $1/s_k^k$, $\bar{\Sigma}_k = diag(0, 0, \ldots, 1/s_{kk}, 0, ..0)$. Then $(\Sigma - \bar{\Sigma}_k)s^k = 0$ and the matrix $\Sigma - \bar{\Sigma}_k$ is nonnegative definite.

Proof : see [13][14][16]. □

The interpretation of this result is straightforward: In the least squares scheme, the choice of the k-th column as regressor is equivalent with the statement that only the measurements on the k-th variable are noisy while the remaining measurements are noisefree. The noise energy of the k-th channel is then proportional to $1/s_k^k$ where s_k^k is the (k,k)-th element of Σ^{-1}. However, by no means this identification scheme provides a well motivated choice for k (unless it is known a priori that only the k-th column k is noisy). Moreover, if a choice has been specified, the solution is unique.

2.2 The Total Linear Least Squares Solution

Let the real $m \times n$ matrix A contain measurements with $m > n$. The task is to find linear relations among its columns. Generically, the matrix A will be of full column rank : $rank(A) = n$. The total linear least squares then proceeds by finding the matrix $\hat{A}$ that is closest to A in Frobeniusnorm, such that $\hat{A}$ is rank deficient. Remark that in the statistical community this solution is often referred to as 'orthogonal regression'. The answer to this problem is given by a result of Eckart-Young [9] and consists of computing the singular value decomposition of A and removing the dyadic decomposition term corresponding to the smallest singular value (which will be unique generically). The unique linear relation is then obtained as the right singular vector corresponding to this smallest singular value [8] [11]. Observe that the term which is removed from the dyadic decomposition in fact models the noise: it is a rank one matrix, hence the noise model is very structured: There are $n-1$ linear relations among the noise variables ! Hence, the total linear least squares scheme certainly violates the requirements of the Frisch scheme. Moreover, the solution is again (generically) unique.

2.3 The noise covariance matrix is the identity matrix

The Frisch scheme has at least one solution. Simply compute the smallest eigenvalue λ_n of Σ with its corresponding eigenvector v_n. Take the noise covariance matrix $\tilde{\Sigma} = \lambda_n I_n$ where I_n is the $n \times n$ identity matrix . Then, is is easy to prove that:

1. $\Sigma - \tilde{\Sigma}$ is nonnegative definite
2. corank $(\Sigma - \tilde{\Sigma}) = 1$ (generically)
3. $(\Sigma - \tilde{\Sigma})v_n = 0$

Remark that this scheme provides the same linear relation as in the case of total least squares. The model of the noise however is now more realistic : the matrix $\tilde{\Sigma}$ is not rankdeficient and all noise 'energies' (the diagonal elements of $\tilde{\Sigma}$) are equal. The column vectors of $\hat{A}$ and $\tilde{A}$ are undetermined in the sense that there exists infinitely many models for $\hat{A}$ and $\tilde{A}$. Although the linear relation is unique, at least the models for the 'exact' data and the noise are not.

3 Characterizations of the maximal corank

In this section, we will first descibe conditions that are to be satisfied for two extremes of the Frisch scheme: the maximal corank 1 and the maximal corank $n-1$. Also some tools that allow to estimate intermediate cases are described (section 3.2). We also give some insight in the characterization of the allowed diagonal elements of the noise model matrix $\tilde{\Sigma}$ (section 3.3).

3.1 Extreme maximal corank conditions.

3.1.1 Maximal corank $n-1$: The Spearman case

Factor analysis, which is a classical data analytic investigation tool, can be converted into the same problem formulation as the Frisch scheme [1] [15] [22]. The older factor analysis literature concentrated on studying conditions for the maximal corank of $(\Sigma - \tilde{\Sigma})$ to be equal to $n-1$, while $(\Sigma - \tilde{\Sigma})$ remains nonnegative definite. Matrices Σ possessing this property are called Spearman matrices. A long story, starting from the beginning of this century, is summarized in :

Theorem 2 Spearman matrices
A positive definite irreducible matrix Σ is a Spearman matrix if and only if, after sign changes of rows and corresponding columns, all its elements σ_{ij} are positive and satisfy:

$$\sigma_{ik}\sigma_{jl} - \sigma_{il}\sigma_{jk} = 0 \qquad \sigma_{ik}\sigma_{ji} - \sigma_{ii}\sigma_{jk} \leq 0 \tag{1}$$

for all quadruples $(i \neq j, k, l; j \neq k, l; k \neq l)$

Conditions 1 express the vanishing of the so-called *tetrad differences* [1]. Hence, if the data A are such that A^tA is a Spearman matrix, then the maximal corank of the Frisch scheme is $n-1$: There exists a diagonal matrix $\tilde{\Sigma}$ such that $\Sigma - \tilde{\Sigma}$ is nonnegative definite and such that there are $n-1$ linear independent linear relations. Hence the underlying phenomenon is extremely simple since it can be described by $n-1$ simultaneous, independent linear relations.

3.1.2 The maximal corank=1 case

Having described one extreme, we now turn our attention to the other extreme which was proved by Koopmans and Reiersol [16] [20] [21].

Theorem 3 The maximal corank=1
The maximal corank of the Frisch scheme is 1 if and only if the matrix Σ^{-1} is elementwise (sign-similar to) a strictly positive matrix. If Σ^{-1} is elementwise positive, the solution set of vectors that satisfy the requirements of the Frisch scheme, is the polyhedral cone generated by the n least squares solutions.

Proof : For a proof using the celebrated Perron Frobeniustheorem for nonnegative matrices see [13] [14] [15]. For a proof that does not invoke this theorem , see [1] including also references to other proofs. □

A symmetric matrix is sign-similar to an elementwise positive one, if by appropriate sign changes of rows and corresponding columns all elements can be made positive. It is an easy exercise to prove that a symmetric matrix is sign-similar to an elementwise positive matrix, if and only if all its column vectors belong to at most two orthants [5]. This theorem covers both a quantitative and qualitative aspect of the uncertainty principle. Qualitatively, it states that, even if there is maximally only one linear relation 'hidden' in the data, the solution is intrinsically non-unique. All vectors that are nonnegative linear combinations of the least squares vectors are candidates to describe the linear relations. The maximal corank is 1 if and only if by appropriate sign changes, all least squares solutions can be brought to the first orthant. However, also quantitatively this result is very attractive: Experiments have shown that the cone spanned by the least squares vectors shrinks to a single point when the data tend towards the noise free case. On the other hand, when more noise is artificially added (worse signal to noise ratio), the polyhedral cone generated by the least squares solution enlarges. Hence there is a direct relation between the amount of noise on the data (which could be both model mismatch and measurement inaccuracy) and the uncertainty in the solution set, characterized by the 'volume' of the cone. If the noise energy is increased,the least squares solutions reach orthant planes and the situation changes. This is now described.

3.2 In between the extremes

The following theorem, proved by Kalman in 1982 [13] [14] , allows to check whether the maximal corank is larger than one .

Theorem 4 *When Σ^{-1} is not sign-similar to an elementwise strictly positive matrix, then the maximal corank is larger than 1.*

Proof : The proof, which is constructive, can be found in [13] [14] . □

Hence, it suffices to check if the columns of Σ^{-1} belong to more than two orthants. If so, then the maximal corank of the problem is certainly larger than or equal to 2. The direct characterization of the maximal corank of a matrix from its properties, is at this moment still *terra incognita* . However, there exist some results, as for Metzler matrices:

Definition 1 Metzler matrix
A matrix M with elements m_{ij} is a Metzler matrix if $m_{ij} \geq 0$ for all $i \neq j$.

There is a close connection between Metzler matrices and the Perron-Frobeniustheorem for nonnegative matrices [19] :

Lemma 1 *Let M be a Metzler matrix . Then $-M^{-1}$ exists and is a positive matrix if and only if M has all of its eigenvalues strictly within the left half of the complex plane.*

Proof : [19] □

Hence , it is straightforward to prove the following corollary:

Corollary 1 *If the matrix Σ is such that $-\Sigma$ is (sign similar to) a Metzler matrix, then maximal corank(Σ) = 1 .*

This is the case if the columns of Σ belong to n different orthants. The following theorem is due to Reiersol [20] and is a kind of generalization of the inverse positiveness condition of the maximal corank=1, as descirbed in theorem 3.

Theorem 5 The Reiersol tree procedure
Let $\mathcal{H}_i$ denote all subsets of the set of n variables and let $\Sigma_{\mathcal{H}_i}$ denote the submatrix of the matrix Σ corresponding to the variables in $\mathcal{H}_i$. If k is the maximum order of any submatrix $\Sigma_{\mathcal{H}_i}$ which is inverse-positive elementwise (or sign-similar to it), then maximal corank($\Sigma - \tilde{\Sigma}$) $\leq n - (k-1)$

Proof : [1] [21] □

As an example, consider the matrix :

$$\Sigma = \begin{pmatrix} 4 & 3 & 0 & -1 & 1 \\ 3 & 8 & -1 & -3 & 1 \\ 0 & -1 & 4 & 3 & 1 \\ -1 & -3 & 3 & 6 & 1 \\ 1 & 1 & 1 & 1 & 3 \end{pmatrix}$$

Now verify that all 2×2 submatrices are (sign-similar to) an inverse elementwise positive matrix. However, the matrix formed from columns and rows 2, 4, 5 is inverse positive as well. None of all possible 4×4 matrices is inverse positive (or sign-similar to it). Hence maximal corank ≤ 3!

3.3 Some results on the allowed diagonal elements of $\tilde{\Sigma}$.

In [3] [6], one finds the following result on the maximal allowable amount of noise that can be present in one variable:

Theorem 6 The maximal noise theorem
Let the i-th diagonal element of $\tilde{\Sigma}$ be $\tilde{\sigma}_i$. Then:

$$0 \leq \sigma_i \leq det(\Sigma)/det(\Sigma_i)$$

where Σ_i is the matrix obtained from Σ by deleting its i-th row and column.

Proof : Follows directly from the requirments of the Frisch scheme and a well known theorem on the determinant of a partitioned matrix [3]. □

Remark that the maximum for $\tilde{\sigma}_i$ is reached for the i-th linear least squares solution, as in theorem 1. If s_i^i is the i-th element of Σ^{-1} then it is easy to prove that:

$$1/s_i^i = det(\Sigma)/det(\Sigma_i)$$

Also, a geometrical characterization is provided of the matrices $\tilde{\Sigma}$ that are such that $corank(\Sigma - \tilde{\Sigma}) = 1$ and $\Sigma - \tilde{\Sigma}$ is nonnegative definite.

Theorem 7 *The family of all the noise matrices $\tilde{\Sigma}$ leading to a corank 1 solution of the Frisch scheme and which is parametrized by the n diagonal elements $\tilde{\sigma}_i$ of $\tilde{\Sigma}$ is a continuous and convex hypersurface. Its intersection with any plane parallel to a coordinate plane is an hyperbola concave toward the origin.*

Proof : see [3]. □

4 Towards a general solution of the Frisch scheme

In the previous section, a summary was given of some special cases of the Frisch scheme. In this section, we summarize a new geometrically inspired framework, of which the authors conjecture that it will lead to the general solution of the Frisch scheme, since already a lot of promising results and insights have been obtained. The geometrical tools that are developed are *orthant invariance* and *null invariance*. It will be demonstrated that these are necessary properties that have to be satisfied by any vector that is a solution of the Frisch scheme. A nice thing is that the algorithms needed to compute the vectors satisfying these properties, have been developed recently [8].

4.1 Orthant invariance

An orthant of the n dimensional vector space $\mathcal{R}^n$ will be characterized by a diagonal matrix E with +1 and -1 along the diagonal : $E = \text{diag}(\pm 1)$. The nonnegative (first) orthant is denoted by the $n \times n$ unit matrix I_n. A vector x is said to belong to orthant $E : x \in E$ if the vector Ex belongs to the first orthant : $Ex \in I_n$. A vector $x \in E$ is orthant invariant for the matrix Σ if $\Sigma x \in E$. It is really verified by inspection that:

Theorem 8 *All solutions x of the Frisch scheme are orthant invariant.*

Define the vector $y = \Sigma x$. If $x \in E$ and x is orthant invariant, also $y \in E$ and equivalently : $Ey = (E\Sigma E)(Ex)$ where now both (Ey) and (Ex) are nonnegative vectors. This observation permits to compute explicitly and characterize geometrically all orthant invariant vectors of the matrix Σ :

Theorem 9 *In orthant E, the orthant invariant vectors can be obtained from the nonnegative solution of the set of linear equations:*

$$\begin{pmatrix} E\Sigma E & -I_n \end{pmatrix} \begin{bmatrix} Ex \\ Ey \end{bmatrix} = 0$$

Hence. the solution set is a convex polyhedral cone .

For algorithms and proofs, the reader is referred to [5] [6] [7] [8]. Geometrically, the solution of the above problem is the intersection of the kernel of the matrix $[(E\Sigma E)\ (-I_n)]$ with the first orthant. Hence one can expect that the 'boundary' vectors (the vertices of the polyhedral cones), will contain zero components. The behavior of these zero components is now investigated.

4.2 Null invariance

A second straightforward observation will appear to be crucial in the computation of the maximal corank. Suppose that the i-th component x_i of a solution vector x is zero : $x_i = 0$. This implies that also the i-th component of the product Σx will be zero. The vector x with zero component $x_i = 0$ will be said to be *null invariant* for component i with respect to the matrix Σ if from $x_i = 0$ it follows that $(\Sigma x)_i = 0$. Then we have:

Theorem 10 Null Invariance
All solution vectors of the Frisch scheme with a zero component are null invariant for that component with respect to the matrix Σ.

So far, two necessary properties have been derived: any potential solution vector has to be *orthant* and *null* invariant. The important role of zeros in the solution vectors is also highlighted by the next observation.

4.3 Recognition of the maximal corank

Temporarily, we return to the problem in terms of the matrix A instead of the 'covariance-Grammian' formulation. If the exact matrix $\hat{A}$ is of rank r, then there exist $n - r$ linear independent solution vectors which will be denoted by $x_1 \ldots x_{n-r}$. Define the $n \times (n - r)$ matrix $X = [\, x_1 \ldots x_{n-r} \,]$, then obviously, $\hat{A}X = 0$.

Lemma 2 *Let $\hat{A}$ be a $m \times n$ matrix $(m > n)$, $rank(\hat{A})=r$. Let X be a $n \times (n-r)$ matrix, $rank(X) = n - r$ such that $\hat{A}X = 0$. It is always possible to find a non-singular $(n - r) \times (n - r)$ matrix P such that each column of the matrix XP contains $n - r - 1$ zeros.*

This property implies that, if the maximal corank of the identification scheme is $n - r$, then there will always exist $n - r$ linearly independent solution vectors each with $n - r - 1$ zeros. These vectors are of course to be found among the orthant null invariant vectors of Σ.

4.4 Allowed vectors.

The properties of orthant and null invariance are necessary for a solution vector but not yet sufficient. In other words, there exist orthant null invariant vectors that are no solution to the problem, simply because the corresponding diagonal matrix $\tilde{\Sigma}$ is such that the difference $\Sigma - \tilde{\Sigma}$ is not nonnegative definite. The diagonal matrix corresponding to an orthant null invariant vector x can be computed by the following scheme: If x is orthant null invariant with components x_i, then:

- if $x_i = 0$, set $\tilde{\sigma}_i = 0$ (i-th diagonal element of $\tilde{\Sigma}$).
- if $x_i \neq 0$, compute $y = \Sigma x$, set $\tilde{\sigma}_i = y_i/x_i$.

An orthant null invariant vector x will be called *allowed* if the difference matrix $\Sigma - \tilde{\Sigma}$ is nonnegative definite. While this last requirement of allowedness has not been solved yet in full generality, it has been solved for some 'special' cases (corank 1 for instance). The geometrical framework however permits to find some more allowed vectors. Hereto, we confine now the attention to some remarkable properties satisfied by the least squares vectors.

4.5 About vectors that are convex combinations of least squares vectors

In theorem 1 it was shown that the columns of the inverse matrix Σ^{-1} are the linear least squares solutions to the problem. Theorem 3 states that whenever the inverse matrix Σ^{-1} is (sign-similar to) an elementwise positive matrix, all linear least squares vectors can be brought by appropriate sign changes into one orthant. The solution set of vectors that satisfy the Frisch scheme conditions is then the convex polyhedral cone generated by the least squares vectors. If Σ^{-1} is not (sign-similar to) an inverse-positive elementwise matrix however, the least squares solutions can never be 'moved' into one orthant. If now two least squares vectors belong to two different orthants, the line segment that connects these two vectors must have an intersection with at least one 'orthant' plane. The corresponding intersection vector hence will have at least one zero. The following theorem states the conditions for such a vector to be allowed [5].

Theorem 11 *Let E be a diagonal sign matrix. Let s^i and s^j be the i-th and j-th column of the matrix $E\Sigma^{-1}E$. Let x be a convex combination of s^i and s^j, such that $x_k = 0$ where :*

$$x = \alpha s^i + (1-\alpha)s^j \quad with \quad \alpha = s^j_k/(s^j_k - s^i_k)$$

If x is orthant null invariant, then x is allowed if and only if $s^i_j = s^j_i > 0$.

Proof : see [5] □

This theorem gives at least heuristically the reason of the importance of the inverse positiveness of the matrix Σ. If Σ^{-1} is elementwise positive (or sign-similar with an inverse positive matrix), then it is impossible to find a convex combination x of two least square vectors that is allowed. Hence, there does not exist an allowed vector with at least one zero. Lemma 1 then suggests that the maximal corank is 1, which as theorem 3 demonstrates is indeed the case. A similar result has been derived in [5].

4.6 The global solution set of the identification

The solution set consists of those vectors x that are orthant null invariant and allowed with respect to the matrix Σ. As a special case, the linear least squares solutions as derived in section 2.1., belong to this class of vectors. They play a prominent role in the maximal corank = 1 case. However, they play also a fundamental role in the description of the corank higher than 1 solution set. It is conjectured by the authors that in this case, the general soution set is a collection of polyhedral convex cones. Their vertices are the least squares solutions and allowed orthant null invariant vectors with zero components. The maximal corank is $n-r$ if there exist maximally $n-r$ linear independent allowed orthant null invariant vectors each with $n-r-1$ zeros. In order to prove completely and rigorously this statement, important partial results have been obtained [5] [6] [7]. Recently, the authors have proved that the set of allowed orthant null invariant vectors is closed and bounded. This supports the strategy of first computing orthant invariant vectors, restricting this set to orthant null invariant vectors and then in a third stage, throwing away the non-allowed vectors.

4.7 Example

As an example of the newly defined geometrical concepts, consider the data matrix Σ:

$$\Sigma = \begin{pmatrix} 9 & 4 & 2 \\ 4 & 17 & 2 \\ 2 & 2 & 8 \end{pmatrix}$$

It can be checked that the inverse matrix Σ^{-1} is not sign similar to an elementwise positive matrix, hence the maximal corank is 2. One can find a geometrical impression of the allowed solution set in figure 1. There are 6 polyhedral convex cones that are generated by linear least squares vectors and allowed orthant null invariant vectors with one zero, which are each a convex combination of two least squares vectors (only 3 of them are depicted in figure 1, the other three lie in the opposite orthants). The representation of the solution in terms of the diagonal elements of $\hat{\Sigma}$ can be found in figure 2.

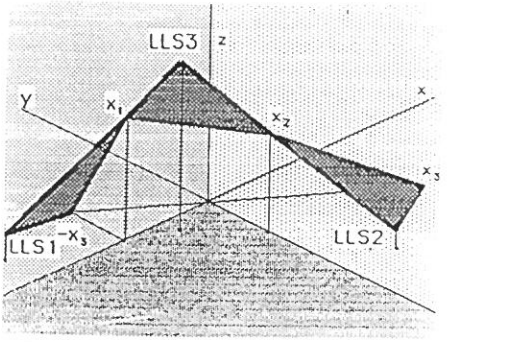


Figure 1: Polyhedral solution cones of the Frisch Scheme for 3 variables and corank=2

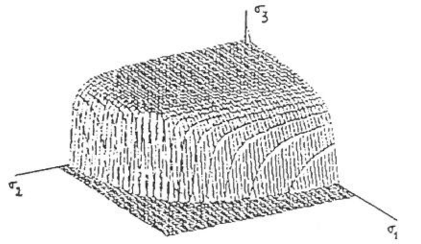


Figure 2: The set of allowed diagonal elements of the noise covariance matrix.

5 The genericity of the solution and its stability

As must be obvious by now, the solution of the Frisch scheme is nothing else than trying to reduce the rank of Σ as much as possible by only changing its diagonal entries while maintaining nonnegative definiteness. Recently [22] , Shapiro has proved that:

Theorem 12 *With probability one, the rank of $\Sigma - \hat{\Sigma}$ cannot be reduced below the so-called Wilson-Ledermann bound :*

$$\phi(n) = [(2n+1) - \sqrt{8n+1}]/2$$

Proof : A complete and general proof can be found in [22] . □

A proof by Baratchart and Kalman, using Thom's topological transversality theorem has been announced [15]. For a derivation of the Wilson-Ledermann formula, see [15] [18] [21] [24]. Hence, $\phi(n)$ can be considered as almost surely a lower bound on the reduced rank of Σ. The set of symmetric $n \times n$ matrices for which the rank can be reduced below the Wilson-Ledermann bound $\phi(n)$ is *thin*, or to be more specific, it has *Lebesgue measure zero*. Now, with respect to the modelling purposes of the Frisch scheme, this observation has important consequences. Suppose that one performs the following simulation experiment. Start from an exact $m \times n$, $(m > n)$ matrix $\hat{A}$ of rank r and add some random noise to it in the form of a noise matrix $\tilde{A}$ so that the elementwise signal-to-noise ratio is really very high. Then generically, one will not be able to recover the exact rank r from the measurement matrix $\Sigma = A^tA$, where $A = \hat{A} + \tilde{A}$, if r is smaller than the Ledermann bound $\phi(n)$.

n	2	3	4	5	6	7	8	9	10	11	12
$\phi(n)$	1	1	2	3	3	4	5	6	6	7	8

Hence, for instance for the case $n = 10$, the maximal corank that is generically obtainable, is 4. However, especially when the signal to noise ratio is high, some other measures such as for instance the singular values of A (or the eigenvalues of Σ), may indicate that the matrix is very close (in Frobeniusnorm) to one of rank r. This genericity result obviously represents a severe requirement for the Frisch scheme to be useful in realistic identification problems. One has to have some a priori knowledge of the expected (co-)rank, in order to chose an appropriate number of measurement channels. Because, if one chooses the number of measurement channels n too high, also the Ledermann bound may become too high so that a low rank can not be achieved generically. It also immediately follows that a reduced rank less than the Ledermann bound, cannot be stable in the sense that, when the elements of Σ are slightly changed, the matrix Σ can generically not be adjusted so that the reduced rank is preserved. Stability is usually expected for a rank $r \geq \phi(n)$ because, if Σ is a matrix for which there exists a matrix $\tilde{\Sigma}$ such that $\Sigma - \tilde{\Sigma}$ satisfies the Frisch scheme requirements and is of rank $r \geq \phi(n)$, then the whole neighbourhood of Σ is reducible to rank r [22]. The Wilson-Ledermann bound arose in the works of Wilson [24] and Ledermann [18]. It was derived by simply counting the number of unknown parameters and the number of equations for a fixed corank, while stating that 'any scientific theory must be *overdetermined* by the data' [18]. In order to obtain real solutions for the Frisch scheme, the corank q should satisfy the inequality :

$$d(n,q) = n - q(q+1)/2 \geq 0$$

When $d(n,q) = 0$, there is a zero dimension solution (i.e. only distinct points) for the so-called communalities (the diagonal elements $\tilde{\sigma}_i$ of $\tilde{\Sigma}$). For example, $d(3,2) = 0$, hence there is a unique solution, which in this case, consists of one point in the parameterspace of the $\tilde{\sigma}_i$ (as is the case in the example of section 4.7). Wilson [24] gives an example $d(6,3) = 0$, where there are 2 distinct numerical solutions.

6 Conclusions

In this paper, a conceptually new approach for the identification of linear relations from noisy data was proposed. A rigorous and consistent definition of noise and linear relations resulted in a fundamental *uncertainty principle of mathematical modelling*. The uncertainty inherent in the initial data reveals itself in a geometrical way in the convex polyhedral cones that characterize the uncertainty of the solution set. Future work will concentrate on a further investigation of the conditions of allowedness within the geometrical framework of section 4. Although the diagonality of the noise matrix $\tilde{\Sigma}$ is required for the Frisch scheme, there exist applications in the domain of identification under feedback, in which one can allow noise matrices $\tilde{\Sigma}$ with non-zero off-diagonal elements. This will be the subject of future research.

References

[1] Becker P.A., De Leeuw J. *The rank of reduced dispersion matrices.* Psychometrika, Vol.52., no.1, pp.125-135, March 1987.

[2] Berman A., Plemmons R.J. *Nonnegative matrices in the mathematical sciences.* Academic Press, New York, 1979.

[3] Beghelli S., Guidorzi R. *Problemi di stima da dati affeti da rumore.* Atti dell'incontro nazionale dei ricercatori del progretto nazionale M.P.I., Como, Villa Olmo, June 1985.

[4] Bode H.W. and C.E. Shannon. *A simplified derivation of linear least square smoothing and prediction theory.* Proc. of the I.R.E., 38, pp.417 - 425, 1950.

[5] De Moor B., Vandewalle J. *A geometrical approach to the maximal corank problem in the analysis of linear relations.* Proc. of the 25th IEEE CDC Conference, Athens, Dec. 1986,pp.1990-1995.

[6] De Moor B., Vandewalle J. *The uniqueness versus the non-uniqueness principle in the identification of linear relations from noisy data.* Proc. of the 25th IEEE CDC Conference, Athens, Dec.1986, pp.1663-1665.

[7] De Moor B., Vandenberghe L., Vandewalle J.*A new approach to the analysis of piecewise linear resistive circuits.* Submitted , February 1988.

[8] De Moor B., Vandewalle J.*A unifying theorem for linear and total linear least squares identification.*, June 1987, Accepted for IEEE Transactions on Automatic Control, to appear in 1988.

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

[10] Frisch R. *Statistical confluence analysis by means of complete regression systems.* Publication no.5., University of Oslo Economic Institute, 192 pages, 1934.

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

[12] Kailath T. (Editor)*Linear Least Squares estimation.* Benchmark papers in Electrical Engineering and Computer Science/17, Dowden, Hutchinson and Ross, Inc., 1977.

[13] Kalman R.E. *Identification from real data.* In 'Current developments in the Interface : Economics, Econometrics, Mathematics'(edited by M.Hazewinkel and A.H.G. Rinrooy Kan), D.Reidel Publishing Co., Dordrecht, pp.161-196, 1982.

[14] R.E. Kalman. *System Identification from noisy data.* Proc. Int. Symp.on Dynamical Systems (Gainesville, Florida 1981), edited by A.Bednarek, Academic Press, 1982.

[15] Kalman R.E., Los C.A. *The Prejudices of Least Squares, Principal Components and Common Factor Schemes.* 6th International Conference on Mathematical Modelling, St. Louis, Missouri, USA, August 1987.

[16] Koopmans T.C. *Linear regression analysis of economic time series.* De Erven F.Bohn NV, Haarlem, 1937.

[17] P. Kovanic. *A new theoretical and algorithmic basis for estimation, identification and control.* Automatica, Vol.22, no.6, 1986, pp.657-674.

[18] Ledermann W. *On the rank of the reduced correlation matrix in multiple factor analysis.* Psychometrika, 1937, 2, 85-93.

[19] Luenberger D.G. *Introduction to dynamic systems. Theory, models and applications.* John Wiley and Sons, New York, 1979.

[20] Reiersol O. *Confluence analysis by means of lag moments and other methods of confluence analysis.* Econometrica 9, 1-24, 1941.

[21] Reiersol O. *On the identifiability of parameters in Thurstone's multiple factor analysis.* Psychometrika, Vol.15, no.2, june 1950.

[22] Shapiro A. *Rank reducibility of a symmetric matrix and sampling theory of minimum trace factor analysis.* Psychometrika, 47-2, pp.187-199, 1982.

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

[24] Wilson E.B., Worcester J. *The resolution of six tests into three general factors.* Proc. of the National Academy of Sciences, 25:73-77, 1929.