SYSID '94

10th IFAC Symposium on System Identification
4 - 6 July 1994

Editors:

Mogens Blanke and Torsten Söderström

Preprints Vol. 3

Published by Danish Automation Society 1994

93-54 C

COMPUTING SOLUTIONS OF THE FRISCH SCHEME USING A POTENTIAL REDUCTION METHOD[1]

Johan DAVID and Bart DE MOOR[2]

ESAT - Katholieke Universiteit Leuven, Kardinaal Mercierlaan 94, 3001 Leuven (Heverlee), Belgium, tel 32/16/22 09 31, fax 32/16/22 18 55, email: bart.demoor@esat.kuleuven.ac.be

Abstract. Despite the fact that the Frisch scheme has been studied for quite some time now, only solutions where the rank of the matrix is one lower than its order are well understood and characterized. Good numerical methods to compute solutions with lower rank do not even exist.
In this paper an algorithm is presented to compute solutions of the Frisch scheme when the resulting rank is lower than the order of the matrix minus one. The algorithm is based on a potential reduction algorithm for convex optimization. The problem is, however, not convex. The resulting solution is not always guaranteed to achieve the minimal possible rank. However, a reduced rank solution is always returned.

Key Words. Identification; optimization; Frisch scheme; linear matrix inequalities.

1. INTRODUCTION

The Frisch scheme deals with the problem of identifying linear relations between measured data. Suppose n variables are measured over m time instants. The measurements are aggregated in an $m \times n$ real matrix M. It is assumed that the number of measurements exceeds the number of variables: $m > n$ so that the matrix M has more rows than columns. This assumption of overdetermination is of course necessary since otherwise the problem would be trivial. An existing linear relation will reveal itself via an n-vector x that belongs to the kernel of the matrix M:

$$Mx = 0.$$

The number of independent linear relations is indicated by the algebraic rank r of M. The *corank* of M is defined as $n - rank(M)$. The corank equals the number of linear independent relations between the variables.

Typically, however, measurements are noisy so that no linear relations between the data exist. A fundamental assumption on the noise is that it is additive and uncorrelated. The measured matrix M can then be written as:

$$M = \hat{M} + N,$$

where $\hat{M}$ denotes the noisefree data and N the noise variables. The noise on the different channels is assumed to be uncorrelated, hence:

$$N^tN = \text{ diagonal}$$

and $ker(N) = 0$. No linear relations between the noise and the exact data exist either:

$$\hat{M}^tN = N^t\hat{M} = 0.$$

The measured, exact and noise Grammians are denoted as:

$$A = M^tM, \quad \hat{A} = \hat{M}^t\hat{M}, \quad D = N^tN.$$

The Frisch scheme can now be formulated as follows:

Given a positive definite matrix A, find a diagonal, positive definite matrix D so that:

1. $\hat{A} = A - D$ is positive semi definite.
2. The corank of $\hat{A}$ is maximal.

[1]The following text presents research results obtained within the framework of the Belgian programme on interuniversity attraction poles (IUAP-17,IUAP-50) initiated by the Belgian State - Prime Minister's Office - Science Policy Programming. The scientific responsibility is assumed by its authors.

[2]Bart De Moor is a senior research associate with the N.F.W.O. (Belgian National Fund for Scientific Research).

The maximization of the corank is necessary since the goal is to find the maximum number of linear relations. Generically the rank of $\hat{A}$, cannot be reduced arbitrarily. There exist a generic lower bound, which is called the Wilson-Lederman bound. See e.g. De Moor (1988) and references therein. The rank of $\hat{A}$ can be reduced under this bound only in special cases.

Although the Frisch scheme has been studied already for a long time, its solutions are only well characterized when the minimal rank of $\hat{A}$ is only 1 lower than the rank of A and necessary and sufficient conditions for this case are well understood now (Kalman,1982). In this paper a method to find solutions of lower rank is presented for the cases where these conditions are not met. It is based on a potential reduction method for convex optimization. In section 2, it is described briefly how the potential reduction method can be used. Section 3 contains some numerical examples. The conclusions are summarized in section 4.

2. THE POTENTIAL REDUCTION METHOD

In this section we describe how a potential reduction method can be used to solve reduced rank linear matrix inequality (LMI) problems. The potential reduction method is an interior point method for the optimization of a linear objective function over a convex set. The Frisch scheme is however not a convex problem. Nevertheless, an extension of the potential method for convex optimization can be used to find solutions quickly. A more detailed description of the algorithm can be found in David (1994).

The Frisch scheme can be reformulated as minimum rank problem involving linear matrix inequalities:

$$\min_D \text{ Rank } A - D \geq 0,$$

$$\text{subject to } D > 0,$$

where D is a diagonal matrix. The solutions where $A - D$ is rank deficient, are situated on the boundary of the convex set of positive (semi) definite matrices. The constraints $D > 0$ are also convex. The idea is to start inside the feasible set and to move to the boundary. Therefore the following objective function is constructed:

$$\phi(D) = q \log\det(A-D) - \log\det \begin{pmatrix} D & 0 \\ 0 & A-D \end{pmatrix}$$

The first term is $-\infty$ where $A - D$ is rank deficient. The second term becomes $+\infty$ on the boundary of the domain $D > 0$ and $A - D > 0$. It is a barrier function for the feasible domain. The constraint $D > 0$ has to be included otherwise some of the elements of D could become negative. To find reduced rank solutions $\phi(D)$ has to be minimized. It might look strange that $A - D$ is also included in the convex term that becomes $+\infty$ when $A - D$ is singular. This does not matter if q is taken high enough as $\phi(D)$ is equivalent to $(q-1)\log\det(A-D) - \log\det D$. It makes the implementation easier however. It is easily shown that q has to be larger than 2.

$\phi(D)$ is not a convex function, hence the theory of Nesterov and Nemirovsky (1993) cannot be applied directly. However, if the concave term is linearized, the theory can be applied. As the first term of $\phi(D)$ is concave, the value of $\phi(D)$ will always be lower than that of the linearization. Thus if the linearized function is minimized the corresponding value of $\phi(D)$ will be even lower.

The Newton direction of the linearized function is $-H(D)^{-1}g(D)$, where $H(D)$ is the Hessian and $g(D)$ is the gradient. It can then easily be computed as the solution of a least squares problem (Boyd and El Ghaoui, 1993):

$$-H^{-1}g = \underset{v}{\text{argmin}} \; \|I(1-q)$$

$$+\sum_{i=1}^{n} v_i (A-D)^{-1/2} E_i (A-D)^{-1/2}\|_F + \sqrt{\sum_{i=1}^{n}(1-v_i)^2}$$

where E_i is an $n \times n$-matrix with all zeros except a one on position (i,i).

The Newton algorithm is:

$$D_{k+1} = D_k - \alpha \text{diag}(H^{-1}g).$$

The damping factor α still has to be determined. To ensure that D_{k+1} is again inside the feasible set if D_k is feasible, the Nesterov-Nemirovsky damping factor is taken. This damping factor depends on the so-called Newton decrement $\delta(D)$ (Nesterov and Nemirovsky, 1993):

$$\delta(D) = \|H(D)^{-1/2} g(D)\|.$$

The Nesterov-Nemirovsky damping factor is then:

$$\begin{cases} \alpha = \frac{1}{1+\delta(D)} & \text{if } \delta(D) > .25, \\ \alpha = 1 & \text{if } \delta(D) \leq .25. \end{cases}$$

If v is known, $\delta(D)$ can easily be computed as:

$$\delta(D) = \left\| \sum_{i=1}^{n} v_i (A-D)^{-1/2} E_i (A-D)^{-1/2} \right\|_F + \sqrt{\sum_{i=1}^{n} v_i^2}.$$

Starting from an initial feasible point, the algorithm will move towards a minimum of $\phi(D)$. Each of the updates will be inside the feasible set. To stop the algorithm the following criterion is used. It can be shown that when D is near the boundary of $A - D > 0$,

$$\left(\frac{\left\| \sum_{i=1}^{n} v_i (A-D)^{-1/2} E_i (A-D)^{-1/2} \right\|_F}{1-q} \right)^2 \tag{1}$$

is a good estimate of the number of decreasing eigenvalues. When it is rounded towards the nearest integer it indicates the number of decreasing eigenvalues of $A-D$. Let n_d be that number. The stopping criterion is then: Sort the eigenvalues in ascending order and check if the eigenvalue indicated by n_d is smaller than a certain tolerance. If so, stop, else continue. The resulting rank of $A - D$ is $n - n_d$.

To start the algorithm needs an initial point inside the feasible set. We suggest to start from the analytic center of the constraints $A - D > 0$ and $D > 0$ (Boyd and El Ghaoui, 1993). The minimum rank problem is, however, not convex. There is not a single minimum. The intuitive idea behind the algorithm is that if the rank of $A - D$ is lower on a point on the boundary than in other boundary points, $\phi(D)$ will be steeper towards this point. Thus it is more likely that the method will converge to this point. Numerical experiments show that this is often the case. However, the optimization method can get stuck in other minima than the optimal one. Still these minima are reduced rank solutions. This is illustrated in the next section.

3. NUMERICAL EXAMPLES

In this section some numerical examples are shown. First 2×2 examples are solved, to illustrate some facts graphically.

Consider the following example:

$$A = \begin{pmatrix} 8 & 2 \\ 2 & 7 \end{pmatrix}.$$

The lowest possible rank for $A-D$ is 1. The rank minimization algorithm starts at the analytic center of $A - D > 0$ and $D > 0$. The evolution of the eigenvalues is shown in figure 1. If D is represented by:

$$D = \begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix}.$$

The steps of the algorithm can be represented in an xy-plane. This is shown in figure 2. Also the boundaries of the feasible region are plotted. The

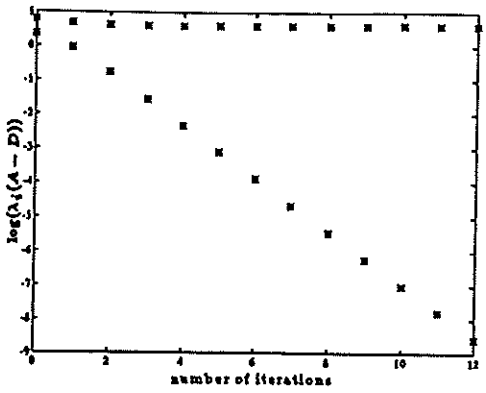

Fig. 1. The logarithm of the eigenvalues for a simple 2×2 example as a function of the number of iterations. One eigenvalue decreases to 0. The resulting rank of $A - D$ is thus 1. Note how the eigenvalue reduces linearly on this logarithmic plot.

boundary of the feasible region consists of the axes $x = 0$ and $y = 0$. The curved line represents the boundary where $\det(A - D) = 0$. All the points on that boundary represent matrices of rank 1. The algorithm starts in the analytic center of the feasible region: $(3.49, 3.06)$. In a few steps the algorithm moves to a singular point on the boundary: $(5.86, 5.13)$.

In figure 2 also the paths followed by the optimization algorithm are shown starting from 40 points around the analytic center. All the paths tend to the boundary of the feasible region where the rank of $A - D$ equals 1.

In the previous example, all the rank deficient matrices had rank 1. Consider the following 2×2 matrices where the minimal possible rank is 0:

$$A = \begin{pmatrix} 10 & 0 \\ 0 & 5 \end{pmatrix}.$$

Again the algorithm is started from the analytic center of the feasible region. In this case two eigenvalues tend to 0. The different steps converge to the intersection of the two boundaries $x = 10$ and $y = 5$ where the rank of $A - D$ is 0. This is shown in figures 3 and 4.

In figure 4 also the paths followed by the optimization algorithm are shown starting from 40 points around the analytic center of the feasible region. All the paths tend to the corner point of the boundary of the feasible region where the rank of $A - D$ equals 0.

The method is not restricted to small examples. Consider a 10×10 matrix A_{10}. The numerical

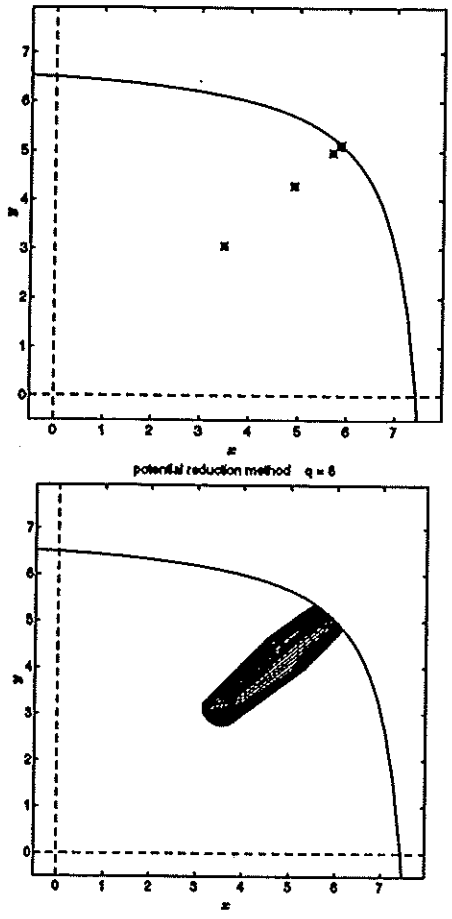

Fig. 2. The upper figure shows the steps taken by the algorithm starting from the analytic center of the feasible region. The curved line is the boundary where $A - D$ is rank deficient. The algorithm quickly moves from the analytic center of the feasible region towards the boundary. The lower figure shows the paths followed starting from 40 points around the analytic center. All the paths move to the boundary where the rank of $A - D$ equals 1.

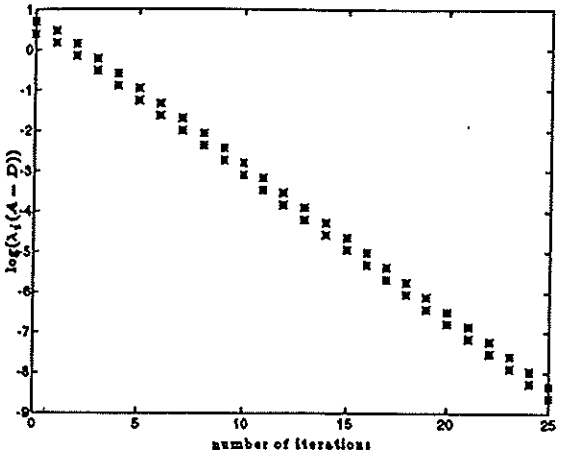

Fig. 3. The logarithm of the eigenvalues for a special 2×2 example as a function of the number of iterations. Two eigenvalue decrease to 0. The resulting rank of $A - D$ is thus 0. Remark how the eigenvalues reduces linearly on this logarithmic plot.

values are shown in the appendix. This matrix is a positive definite matrix of full rank. The matrix is constructed such that the minimal rank for the Frisch scheme is known. A is constructed as $A = BB^t + D_1$, with $B \in \mathbb{R}^{10\times 3}$, a random matrix. D_1 is a diagonal matrix with strictly positive elements. The minimal obtainable rank of $A - D$ is thus 3. The eigenvalues as a function of the iterations are shown in figure 5. 7 eigenvalues decrease. The rank of the resulting matrix is thus 3. The optimal D is exactly D_1. Note that this is not a *typical* order 10 example, since the Wilson-Lederman bound for $n = 10$ equals 6.

As a last example the evolution of the eigenvalues for a random 7×7 matrix A_7 are shown in figure 6. The numerical values of the entries are shown in the appendix. Two eigenvalues decrease. The resulting rank of $A - D$ is 5. This is higher than the Wilson-Lederman bound, which is 4 for a 7×7 matrix. This is correct as the Wilson-Lederman bound is only a generic *lower* bound. The resulting elements of D are:

$$D = \text{diag}\ (\ 0.008258 \quad 2.475 \quad 0.0109 \quad 0.01552 \quad 0.2606 \quad 0.8981 \quad 0.01883\ )$$

4. CONCLUSIONS

In this paper an algorithm is presented to find reduced rank solutions for the Frisch scheme. The algorithm is based on an algorithm for convex optimization. This problem is however not convex.

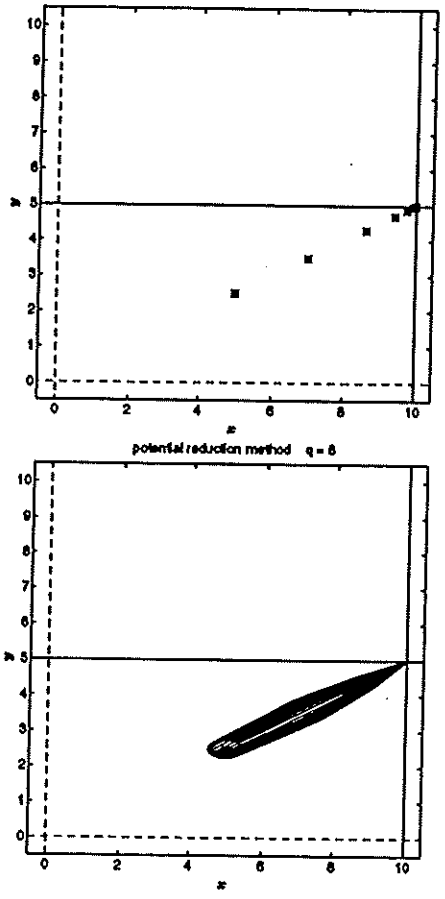

Fig. 4. The upper figure shows the steps taken by the algorithm starting from the analytic center of $A - D > 0$ and $D > 0$. The point with the lowest rank of $A - D$ is $(10, 5)$. The algorithm moves quickly towards that point. The lower figure shows the paths followed by the optimization algorithm starting from 40 points around the analytic center. All the paths move to the corner point of the boundary where the rank of $A - D$ is 0.

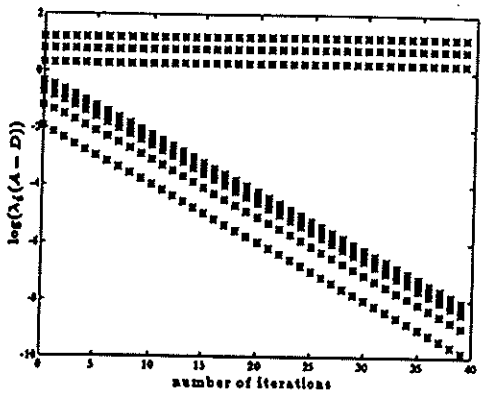

Fig. 5. The logarithm of the eigenvalues for an artificially constructed 10×10 example as a function of the number of iterations. Seven eigenvalues decrease to 0. The resulting rank of $A - D$ is thus 3.

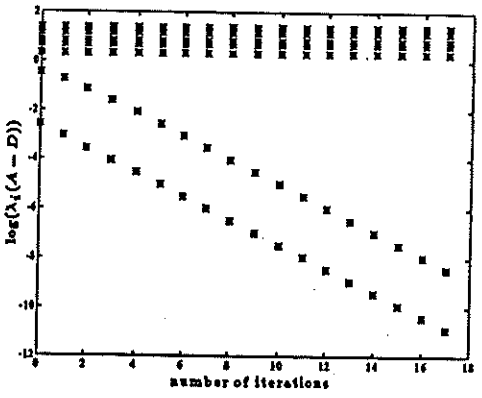

Fig. 6. The logarithm of the eigenvalues for a random 7×7 example as a function of the number of iterations. Two eigenvalues decrease to 0. The resulting rank of $A - D$ is thus 5.

The algorithm can evolve to points that are not optimal. However, the solutions are always of reduced rank.

The same algorithm can also be used to compute solutions for extensions of the Frisch scheme. E.g. if it is known that certain cross-correlations exist between channels. This can be done by introducing non-zero parameters in off-diagonal elements of D. The only restriction is that D has to remain symmetric (and positive definite).

5. REFERENCES

Boyd, S. and L. El Ghaoui (1993). Method of centers for minimizing generalized eigenvalues. *Linear Algebra and Applications*, 188-189, 63-113.

David, J. (1994). *Algorithms for Analysis and Design of Robust Controllers*, Ph.D. thesis, Dept. Elec. Eng., K.U.Leuven, 1994.

De Moor, B. (1988). *Mathematical concepts and techniques for modelling of static and dynamic systems*, Ph.D. thesis, Dept. Elec. Eng., K.U.Leuven, Leuven.

Kalman, R.E. (1982). System identification from noisy data. In: *Proc. Int. Symp. on Dynamical Systems* (Gainesville, Florida, 1981) (A. Bednarek, Ed.), Academic Press, New York.

Nesterov Y. and A. Nemirovsky (1993). *Interior point polynomial methods in convex programming: Theory and applications*, SIAM.

APPENDIX

$$A_{10} = \begin{pmatrix} 3.67 & 1.78 & -1.68 & 1.295 & -1.644 & 2.474 & 0.247 & 4.011 & 2.127 & -0.348 \\ 1.78 & 1.57 & -0.215 & -0.020 & -0.797 & 0.905 & -0.185 & 1.468 & 0.736 & 0.928 \\ -1.68 & -0.215 & 2.822 & -1.429 & 0.710 & -0.714 & -0.468 & -2.125 & -1.87 & 2.319 \\ 1.295 & -0.020 & -1.429 & 1.954 & -0.662 & 1.269 & 0.449 & 2.146 & 1.25 & -1.479 \\ -1.644 & -0.797 & 0.710 & -0.662 & 1.756 & -1.418 & -0.142 & -2.073 & -0.940 & 0.037 \\ 2.474 & 0.905 & -0.714 & 1.269 & -1.418 & 3.654 & 0.353 & 3.922 & 1.111 & 0.435 \\ 0.247 & -0.185 & -0.468 & 0.449 & -0.142 & 0.353 & 0.699 & 0.589 & 0.347 & -0.614 \\ 4.011 & 1.468 & -2.125 & 2.146 & -2.073 & 3.922 & 0.589 & 6.453 & 2.492 & -0.776 \\ 2.127 & 0.736 & -1.87 & 1.25 & -0.940 & 1.111 & 0.347 & 2.492 & 1.879 & -1.558 \\ -0.348 & 0.928 & 2.319 & -1.479 & 0.037 & 0.435 & -0.614 & -0.776 & -1.558 & 3.582 \end{pmatrix}$$

$$A_7 = \begin{pmatrix} 6.157 & -5.434 & 4.433 & -0.442 & -1.17 & 1.227 & -3.578 \\ -5.434 & 13.45 & -4.839 & 1.944 & 4.976 & -1.737 & 2.56 \\ 4.433 & -4.839 & 9.128 & -5.036 & -2.579 & -2.126 & 0.975 \\ -0.442 & 1.944 & -5.036 & 6.401 & 3.822 & 4.124 & -0.608 \\ -1.17 & 4.976 & -2.579 & 3.822 & 6.669 & 3.881 & 0.278 \\ 1.227 & -1.737 & -2.126 & 4.124 & 3.88 & 7.352 & -1.911 \\ -3.578 & 2.56 & 0.975 & -0.608 & 0.278 & -1.911 & 6.122 \end{pmatrix}$$