Athanasios C. Antoulas (Ed.)

Mathematical System Theory

The Influence of R. E. Kalman

A Festschrift in Honor
of Professor R. E. Kalman
on the Occasion
of his 60th Birthday

With 49 Figures

Springer-Verlag
Berlin Heidelberg NewYork
London Paris Tokyo
Hong Kong Barcelona
Budapest

Identification of Dynamic Systems from Noisy Data: The Case $m^* = 1$*

M. Deistler[1] and B. D. O. Anderson[2]

[1] Institute of Econometrics Operations Research and System Theory, Technical University of Vienna, A-1040 Vienna, Austria
[2] Department of Systems Engineering, Research School of Physical Sciences and Engineering, The Australian National University, GPO Box 4, Canberra, ACT 2601, Australia

Introduction

In identification of linear systems the "main stream" approach to noise modeling is to add all noise to the outputs (assuming orthogonality), or to the equations (which is the same for our analysis). In econometrics these models are named errors-in-equations models. Here we are concerned with the case where in principle all variables may be contaminated by noise. Such models are called *errors-in-variables* (EV) or *latent variables models*, or using a slightly different but equivalent formulation *factor models*. Whereas in the errors-in-equations approach the deterministic system is embedded into its stochastic environment in an asymmetric way, EV modeling is (in principle) more general and corresponds to symmetric noise modeling. The asymmetry of errors-in-equations modeling can be justified in many situations, in particular in prediction, however there are a number of cases where this asymmetry is not appropriate and leads to "prejudiced" results. The symmetric EV modeling is appropriate for instance:

(i) If we are interested in the true system generating the data (rather than in prediction or in encoding the data by system parameters) and we cannot be sure a priori that the observed inputs are not corrupted by noise.
(ii) If we want to approximate a high dimensional data vector by a relatively small number of factors.
(iii) If we have no sufficient a priori information about the number of equations in the system or about the classification of the variables into inputs and outputs; then we have to perform a more symmetric system modeling, which in turn demands a more symmetric noise model.

The statistical analysis of EV models has a long history in econometrics, psychometrics and statistics (see, e.g. Frisch 1934; Koopmans 1937; Bekker and de Leeuw 1987). Recently there has been a resurging interest in such models; these models also attracted attention in system and control theory. Partly this

* Support by the Austrian "Fonds zur Förderung der wissenschaftlichen Forschung" "Schwerpunkt Angewandte Mathematik" (S32/02) is gratefully acknowledged.

development was triggered by a number of seminal papers authored by Kalman (see, e.g. Kalman 1982, 1983). In these papers the theoretical and practical appeal of this approach, as a general approach to the problem of identification of linear systems has been pointed out as well as the technical complications and the main open problems. The price to be paid for the generality of noise modeling is a significantly increased complication in the statistical analysis. The main problem is a basic "nonidentifiability" in the sense that in general the system is not uniquely determined from the (population) second moments of the observations, since the separation between the system and the noise part is not unique.

Our work was definitely inspired by Kalman's ideas. Whereas Kalman analysed the static case, our focus is on the case of dynamic systems; for the dynamic case see also Anderson and Deistler (1984, 1987), Picci and Pinzoni (1986), and Deistler and Anderson (1989).

The system considered is of the form

$$w(z)\hat{z}_t = 0 \tag{1.1}$$

where $\hat{z}_t$ is the n-dimensional vector of latent (i.e. in general unobserved) variables, z is used for the backward-shift on $\mathbb{Z}$ (i.e. $z(\hat{z}_t | t \in \mathbb{Z}) = (\hat{z}_{t-1} | t \in \mathbb{Z})$) as well as for a complex variable and

$$w(z) = \sum_{j=-\infty}^{\infty} W_j z^j; \quad W_j \in \mathbb{R}^{m \times n} \tag{1.2}$$

We will call $w(z)$ the *relation function* of the exact relation (1.1) (compare Willems 1986). Clearly, systems of the form (1.1) are *symmetric* in the sense that we need no a priori classification of the variables $\hat{z}_t$ into inputs and outputs and no a priori information about causality directions; without restriction of generality we will assume that $m \leq n$ holds and that $w(z)$ contains no linearly dependent rows; also in general m is not known a priori.

The observed variables are of the form

$$z_t = \hat{z}_t + u_t \tag{1.3}$$

where u_t is the noise vector.

Throughout the paper we will assume:

(i) (z_t), $(\hat{z}_t)$ and (u_t) respectively are [wide sense] stationary processes [with real valued components] and spectral densities Σ, $\hat{\Sigma}$ and D. [In addition limits of random variables are understood in the sense of mean square convergence.]

(ii) $E\hat{z}_t = Eu_t = 0$

(iii) $E\hat{z}_t u'_s = 0$ and finally

(iv) D is diagonal.

For a discussion of assumption (iv) see, e.g. Deistler and Anderson (1989) and Deistler (1989). Without imposing any assumption besides (ii) and (iii), in general every system would be compatible with the second moments of the observations; thus some additional assumptions have to be imposed. By assumption (iv) the common effects are attributed to the system and the individual effect to the noise. Clearly (iv) cannot be justified universally; in other words it will be a "prejudice" in a number of applications. However it is a reasonable assumption for a sufficiently large class of cases.

In our analysis the frequency λ will be kept fixed. In this sense $\Sigma, \hat{\Sigma}$ and D are considered as (constant) Hermitian matrices rather than as spectral densities. From (1.3) we have

$$\Sigma = \hat{\Sigma} + D \tag{1.4}$$

Clearly (1.4) may also be interpreted as coming from a (static) relation between $\mathbb{C}^n$-valued random variables $z, \hat{z}$ and u

$$\begin{aligned} &z = \hat{z} + u; \quad W\hat{z} = 0, \quad W \in \mathbb{C}^{m \times n} \\ &\Sigma = Ezz^*; \quad \hat{\Sigma} = E\hat{z}\hat{z}^*; \quad D = Euu^* \end{aligned} \tag{1.5}$$

In this paper we will analyse the relation between the second moments of the observations Σ and the system and noise characteristics $w(z)$ and D. Such an analysis is a necessary first step for an analysis of the properties of estimation and inference procedures. The main question are (compare Deistler and Anderson 1989).

(a) Find the maximum number, m^* say, of (linearly independent) rows of $w(z)$ among the set of all $w(z)$ compatible with given Σ. Sometimes we also use the symbol $mc(\Sigma)$ for m^* if we want to make the dependence on Σ explicit.
(b) Give a description of the set of all $(w(z), D)$ compatible with given Σ; in addition describe the subset corresponding to different numbers of linear relations m.
(c) Describe the set of all Σ corresponding to a given m^*, $n > m^* \geqq 1$.

Thus the problems we consider are (a) to find the (maximum) number of equations for given Σ, (b) to describe the set of all observationally equivalent (based on second moments only) signal and noise characteristics and (c) to describe the set of spectral densities corresponding to a given m^*. There is no general solution available for these problems up to now.

In this paper, we focus on the case of general n and $m^* = 1$ [for n general and $m^* = n - 1$ see Anderson and Deistler (1990)]. For the static case, where $\hat{z}_t$ and u_t are (real) white noise processes (and thus $\Sigma(\lambda)$ and $\hat{\Sigma}(\lambda)$ are constant with real entries) and $w(z)$ is constant with real entries, this problem has a long history, see, e.g. Frisch (1934), Koopmans (1937), and Kalman (1982), (1983). The dynamic case turns out to be significantly more complicated. The cases $n = 2$ and $n = 3$ have been treated in detail in Anderson and Deistler (1984), (1987).

The paper is organized as follows. In Sect. 2 and 3 we are concerned with question (b) and to a small extent also with question (a). The set of all rows of

w (when suitably normalized) compatible with a given Σ is called the solution set. In Sect. 2 some topological properties of the solution set are shown. In Sect. 3 some additional results concerning the form of the solution set are derived. In Sect. 4 we investigate the set of all Σ with $mc(\Sigma)=1$ and the function attaching to every such Σ the corresponding solution set.

2 The Solution Set—Some General Properties

In a first step, let us assume temporarily that $\hat{\Sigma}$ (rather than Σ) is known. Clearly relation (1.1) implies

$$w(e^{i\lambda})\cdot\hat{\Sigma}(\lambda)=0 \tag{2.1}$$

Conversely, if we commence from $\hat{\Sigma}$ and if we want to explain by the system as much as possible and if we have no additional a priori information, then by (2.1), the rows of w are defined as a basis of the left kernel of $\hat{\Sigma}$ and w is unique up to basis change.

Clearly in general only Σ is known and thus equation (1.4) will be the starting point of our analysis. Remember that $\Sigma, \hat{\Sigma}$ and D are nonegative definite and that $\hat{\Sigma}$ is singular and D is diagonal. In view of this for given $\Sigma, \hat{\Sigma}$ and D are called *feasible* if

$$0\leqq\Sigma-D\leqq\Sigma \tag{2.2}$$

holds, where $\Sigma-D=\hat{\Sigma}$ is singular and D is diagonal. As easily can be shown, for every $\Sigma\geqq 0$ a feasible decomposition (1.4) and thus a corresponding EV representation exists. To avoid having to consider a number of special cases we will assume throughout that

(v) $\Sigma>0$
(vi) $\sigma_{ij}\neq 0 \quad i,j=1,\dots,n$ and
(vii) $\tilde{s}_{ij}\neq 0 \quad i,j=1,\dots,n$ hold.

Here $\tilde{S}=\Sigma^{-1}$ and as a general rule if e.g. Σ is a matrix, its i,j-entry is denoted by the corresponding lower case symbol σ_{ij}.

For given Σ, a vector $\tilde{x}\in\mathbb{C}^n$ is called a *solution* if there exists a feasible $\hat{\Sigma}$ satisfying

$$\tilde{x}\hat{\Sigma}=0 \tag{2.3}$$

The set of all solutions corresponding to a given Σ is called the *solution set* $\tilde{L}$ (of Σ); sometimes we also use the notation $\tilde{L}_\Sigma$. Analogously we define D as the set of all feasible matrices D corresponding to Σ. Since $\tilde{L}$ is the union of linear spaces of dimension greater than zero, we may find a normalization useful. In most parts of the paper, the first component of $\tilde{x}, x_1$ is normalized to one.

Let us define the matrix $S=(\tilde{s}_{ij}\cdot\tilde{s}_{i1}^{-1})$, $i,j=1,\ldots,n$ and let s_j denote the j-th row of S. Now it is easily seen from

$$s_j\cdot\Sigma=(0,\ldots,0,\tilde{s}_{j1}^{-1},0,\ldots,0)=s_jD$$

that s_j is the solution (with first component normalized to one) corresponding to the j-th *elementary regression*, i.e. to the case where all components of z_t, except for the j-th, are assumed to be observed free of noise; s_j will be called the j-th *elementary solution*. Since the first elementary solution s_1 always exists, no matrix Σ is excluded by the normalization $\tilde{x}_1=1$. However, the kernel of $\hat{\Sigma}$ may be orthogonal to $(1,0,\ldots,0)$ and in this sense the normalization may be a restriction of generality. However as will be shown in the subsequent Lemma 3, this situation will not occur in the case $m^*=1$. Clearly elementary solutions can also be defined for singular matrices Σ. They correspond to the projection of the j-th component of z in (1.5) on the space spanned by all other components.

Now, let us state some useful lemmas.

Lemma 1. *Let $\Sigma\geqq 0$ (may be singular). If the n-th row of Σ,σ_n say, is linearly independent from the other rows $\sigma_1,\ldots,\sigma_{n-1}$ of Σ, then the n-th elementary regression gives a noise covariance matrix D of the form*

$$D_n=\mathrm{diag}\{0,\ldots,0,d_{nn}^{(n)}\}.$$

where $d_{nn}^{(n)}>0$ and where $rk(\Sigma-D_n)=rk(\Sigma)-1$ holds (here $r(A)$ denotes the rank of A).

Proof. The proof is straightforward and can be seen from projecting the n-th component of z (see 1.5) on the linear space spanned by the other components of z.

Lemma 2. *Let $D=\mathrm{diag}\{d_{ii}\}$ be feasible and let $d_{ii}^{(i)}$ correspond to the i-th elementary regression. Then*

$$0\leqq d_{ii}\leqq d_{ii}^{(i)} \tag{2.5}$$

Proof. Without restriction of generality, take $i=1$; let $D=A+B$; where $A=\mathrm{diag}\{d_{11},0,\ldots,0\}$ and $B=\mathrm{diag}\{0,d_{22},\ldots,d_{nn}\}$.

First note that for $d_{11}>d_{11}^{(1)}$, the matrix $\Sigma-A$ would not be nonnegative definite. To see this consider

$$\det(\Sigma-A)=(\sigma_{11}-d_{11})\cdot f_1(\Sigma)+f_2(\Sigma) \tag{2.6}$$

where f_1 and f_2 depend only on Σ and where

$$f_1(\Sigma)=\det\begin{pmatrix}\sigma_{22}\cdots\sigma_{2n}\\ \cdots\cdots\cdots\\ \sigma_{n2}\cdots\sigma_{nn}\end{pmatrix}>0$$

holds. The expression is zero for $d_{11}=d_{11}^{(1)}$, and hence is negative for $d_{11}>d_{11}^{(1)}$.

Now $\Sigma - D = \Sigma - A - B = C \geqq 0$ would imply $B + C = \Sigma - A \geqq 0$ which is a contradiction for $d_{11} > d_{11}^{(1)}$.

For fixed Σ, the relation between $\tilde{L}$ and D is given by

$$\tilde{x}\Sigma = \tilde{x}D, \quad \tilde{x}\in\tilde{L}, \quad D = \mathbf{D} \tag{2.7}$$

Let us consider the case $mc(\Sigma) = 1$ in more detail now. We investigate the solution set and the set of all feasible matrices D.

Lemma 3. *For $mc(\Sigma) = 1, \tilde{x}\in\tilde{L}$ and $\tilde{x} \neq 0$ imply that every entry $\tilde{x}_j$, $j = 1,\ldots,n$, of $\tilde{x}$ is unequal to zero.*

Proof. We give a proof by contradiction. If e.g. $\tilde{x}_1 = 0$ holds for $\tilde{x}\in\tilde{L}, \tilde{x}\neq 0$, then as seen from (2.7) we may put $d_{11} = 0$ and D remains feasible. Also the last $n-1$ rows of $\Sigma - D$ are clearly linearly dependent. Then the first row of $\hat{\Sigma} = (\Sigma - D)$ is linearly independent from the other rows of $\hat{\Sigma}$ since otherwise $mc(\hat{\Sigma}) > 1$ would hold. Now performing the first elementary regression for $\hat{\Sigma}$, (not Σ) using an evident notation, corresponds to determining $\hat{s}_1$ and $\hat{D}$ so that

$$\hat{s}_1\hat{\Sigma} = \hat{s}_1\hat{D},$$

where

$$\hat{D} = \mathrm{diag}\{\hat{d}_{11}, 0, \ldots, 0\} \quad \text{and} \quad 0 \leqq \hat{D} \leqq \hat{\Sigma}$$

By Lemma 1, $\hat{d}_{11} > 0$ holds. From

$$\Sigma \geqq \hat{\Sigma} \geqq \hat{\Sigma} - \hat{D} = \Sigma - (D + \hat{D}) \geqq 0$$

we see that $D + \hat{D}$ is feasible; Lemma 1 then implies $rk(\Sigma - (D + \hat{D})) < rk\,\hat{\Sigma}$ and thus $mc(\Sigma) > 1$.

Therefore in the case $mc(\Sigma) = 1$, the normalization $\tilde{x}_1 = 1$ is no restriction of generality and we can consider the (normalized) solution set

$$L = \{x \mid (1, x)\in\tilde{L}\}$$

The relation between L and $\mathbf{D}$ (remember that Σ is kept fixed) then is of the form

$$x(\Sigma_{22} - D_{22}) = -\Sigma_{12} \tag{2.8}$$

where

$$\Sigma = \begin{pmatrix} \sigma_{11}, \Sigma_{12} \\ \Sigma_{12}^*, \Sigma_{22} \end{pmatrix}, \quad D = \begin{pmatrix} d_{11}, 0 \\ 0, D_{22} \end{pmatrix}$$

$\Sigma_{12}\in\mathbb{C}^{n-1}$; $\Sigma_{22}, D_{22}\in\mathbb{C}^{(n-1)\times(n-1)}$.

Note that for all feasible D, $\hat{\Sigma}_{22} = \Sigma_{22} - D_{22}$ must have full rank $n-1$, since otherwise $mc(\Sigma) > 1$ would hold. Thus x is uniquely determined for given $D\in\mathbf{D}$ by:

$$x = -\Sigma_{12}(\Sigma_{22} - D_{22})^{-1} \tag{2.9}$$

Conversely, for given x, D is uniquely determined by

$$x\Sigma_{22} = xD_{22} \tag{2.10}$$

and

$$\sigma_{11} + x\Sigma_{12}^* = d_{11} \tag{2.11}$$

Thus we have defined a bijection, i say between L and $\mathbf{D}$.

Theorem 2.1

(a) *$mc(\Sigma) = 1$ if and only if no $\tilde{x} \in \tilde{L}\tilde{x} \neq 0$ has a zero entry*
(b) *for $mc(\Sigma) = 1$, the relation between L and $\mathbf{D}$ defined by (2.9)–(2.11) is a homeomorphism*
(c) *for $mc(\Sigma) = 1$, L and $\mathbf{D}$ are compact. If we consider L as a subset of $\mathbb{R}^{2(n-1)} \triangleq \mathbb{C}^{n-1}$, then L is of real dimension $n-1$.*

Proof

(a) One part is just Lemma 3. Conversely if $mc(\Sigma) > 1$ holds, then $\tilde{L}$ contains at least one linear subspace of dimension greater than one and in this subset clearly there is an element $\tilde{x} \neq 0$, with one zero component.
(b) As has been shown already, i is a bijection. The continuity in both directions is easily seen from (2.9)–(2.11).
(c) Clearly, $\mathbf{D}$ is bounded. Let $\hat{\Sigma}_n = \Sigma - D_n, D_n \in \mathbf{D}$, be a convergent sequence with limit $\hat{\Sigma}$. We then have $\Sigma \geqq \hat{\Sigma} \geqq 0$ and $\hat{\Sigma}$ is singular since $\det \hat{\Sigma}_n = 0$ holds and the determinant is a continuous function of its entries and therefore $\hat{\Sigma}$ is feasible. Thus every convergent sequence $D_n \in \mathbf{D}$ takes its limit D in $\mathbf{D}$, in other words $\mathbf{D}$ is closed. Since the image of a compact set by a continuous mapping is compact, L is compact by (b). As is seen from (2.9) and (2.11), for given $d_{22}, \ldots, d_{nn}$ (and given Σ of course) x and thus d_{11} are uniquely determined. Note that $D = \lambda_{\min}(\Sigma) \cdot I$, where $\lambda_{\min}(A)$ is used to denote the smallest eigenvalue of $A \geqq 0$, is always feasible. This follows from

$$\Sigma = U\Lambda U^* = U(\Lambda - \lambda_{\min} I)U^* + \lambda_{\min} I \tag{2.12}$$

where Λ is the diagonal matrix of eigenvalues of Σ and where U is a unitary matrix. For $mc(\Sigma) = 1$, all principal minors of $(\Sigma_{22} - \lambda_{\min} I_{n-1})$ are strictly positive and since the determinant is a continuous function of the matrix entries, all principal minors of $\Sigma_{22} - D_{22}$ are also positive in a suitably chosen neighborhood of $\lambda_{\min} I_{n-1}$, and moreover this neighborhood can be chosen such that $0 < d_{11} < d_{11}^{(1)}$ holds. Thus this neighborhood is homeomorphic to the corresponding neighborhood of $\lambda_{\min} I$ in $\mathbf{D}$. Replacing $\lambda_{\min} I$ by a general $D \in \mathbf{D}$ and using an analogous argument, we can show that $\mathbf{D}$ is locally homeomorphic to $\mathbb{R}_+^{n-1}$ and thus topological manifold of dimension $n-1$.

Remark 1. Note that for the static case a significantly more far-reaching result is available, see Frisch (1934), Koopmans (1937) and Kalman (1982). In this case $m^* = 1$ if and only if Σ^{-1} is "like" a positive matrix (i.e. Σ^{-1} can be made to a matrix with positive entries by eventually multiplying some of the rows and the corresponding columns by -1) and the solution set L is the convex hull generated by the elementary solutions. The proof of this result relies on a Perron–Frobenius type argument which cannot be carried over to matrices with complex entries. Note in particular that whereas it is easy to check whether Σ^{-1} is like a positive matrix, condition (a) is not of much value for that purpose. Also note that, as can be seen from Anderson and Deistler (1987) for the case $n=3$, in the dynamic situation we may have that Σ^{-1} is not like a positive matrix (with properly complex entries) and still $m^* = 1$ holds. As we will see in Sect. 3, in general for the complex case, the solution set L will not be a polytope since it has curved bounding hyper-surfaces.

Remark 2. Note that L is compact if and only if $mc(\Sigma)=1$ holds: Assume $mc(\Sigma)>1$; let $\hat{\Sigma}$ be feasible such that the dimension of the kernel of $\Sigma-\hat{\Sigma}$, $\ker(\Sigma-\hat{\Sigma})$, is equal to $mc(\Sigma)$; now it is straight-forward to show that the dimension of $\ker(\Sigma-\hat{\Sigma})\cap\{x\in\mathbb{C}|x_1=1\}$ is equal to $mc(\Sigma)-1$ and thus is greater than zero. Also note that for $mc(\Sigma)>1$, the relation between L and $\mathbf{D}$ is not a function in either direction.

3 Solutions on Complex Lines

In this section we further investigate the solution set. Again Σ is kept fixed. The main idea here is to connect two points, x and y say, from the solution set by the complex line

$$\alpha x+(1-\alpha)y, \quad \alpha\in\mathbb{C} \tag{3.1}$$

and to investigate for which α, $\alpha x+(1-\alpha)y\in\tilde{L}$ holds. Note that $\alpha x+(1-\alpha)y$, $\alpha\in\mathbb{C}$ is a plane in $\mathbb{R}^{2n}$. The results obtained in this section are valid for general $mc(\Sigma)$. We start from the equation

$$(\alpha x+(1-\alpha)y)\Sigma=(\alpha x+(1-\alpha)y)D=\alpha x D_x+(1-\alpha)yD_y \tag{3.2}$$

where $x,y\in\tilde{L}$, $x_1=y_1=1$ and D_x and D_y correspond to x and y respectively; D is diagonal and the unknown variable in (3.1). Clearly $\alpha x+(1-\alpha)y\in\tilde{L}$ if and only if there is a D satisfying (3.2) and $D\geqq 0$ and $\Sigma-D\geqq 0$ hold.

First consider the case $x=s_1, y=s_j$, $j>1$, $D_x=D_1$ and $D_y=D_j$, i.e. we investigate the real plane given by the first and the j-th elementary solutions. Then the first equation in (3.2) is of the form

$$\alpha d_{11}^{(1)}=d_{11} \tag{3.3}$$

and the j-th equation is of the form

$$(1-\alpha)s_{jj}d_{jj}^{(j)}=(\alpha s_{1j}+(1-\alpha)s_{jj})d_{jj}$$

which gives

$$\frac{1}{1+\dfrac{\alpha}{1-\alpha}\dfrac{s_{1j}}{s_{jj}}}\cdot d_{jj}^{(j)} = d_{jj} \tag{3.4}$$

By Lemma 2, $d_{ll}^{(l)} \geqq d_{ll} \geqq 0$ must hold for every feasible D. Also note that $s_{1j} \cdot s_{jj}^{-1} > 0$; thus (3.3) and (3.4) imply $\alpha \in [0, 1]$. Put

$$d_{ii} = 0 \quad \text{for } i \neq 1, i \neq j \tag{3.5}$$

then such a prescription for D satisfies (3.2) and $D \geqq 0$ for every $\alpha \in [0, 1]$. In order to show that a D given by (3.3)–(3.5) is feasible for every $\alpha \in [0, 1]$, it remains to show that $\Sigma - D \geqq 0$ holds. Note that for the j-th elementary regression $d_{jj}^{(j)}$ is the unique solution of the equation

$$\det(\Sigma - \operatorname{diag}(0, \ldots, d_{jj}, 0, \ldots, 0)) = 0 \tag{3.6}$$

in the variable $d_{jj} \in \mathbb{R}$. This is a direct consequence of the fact that (3.6) is a linear equation with a positive coefficient for $(\sigma_{jj} - d_{jj})$ (compare 2.6). Now performing the j-th elementary regression for $\Sigma - \alpha D_1$, $\alpha \in (0, 1)$ we see that the corresponding noise covariance matrix is diag $\{0, \ldots, d_{jj}, 0, \ldots, 0\}$ with d_{jj} given by (3.4) and thus $\Sigma - D = (\Sigma - \alpha D_1) - \operatorname{diag}\{0, \ldots, d_{jj}, 0, \ldots, 0\} \geqq 0$.

Let us consider

$$F = \left\{ \sum_{j=2}^{n} \beta_j s_j \,\middle|\, \sum_{2}^{n} \beta_j = 1, \beta_j \in \mathbb{C} \right\} \tag{3.7}$$

For every $y \in F \cap \tilde{L}$ we can choose a corresponding feasible D with $d_{11} = 0$. Thus, for $x = s_1$ and $y \in F \cap \tilde{L}$ we have (3.3) and thus $\alpha \in [0, 1]$. [Clearly, here, in general not every $\alpha \in [0, 1]$ gives a solution.]

In an analogous way as before we proceed in the case $x = s_l$, $y = s_j$, $l, j \neq 1$; $l \neq j$. Then from (3.2) we obtain

$$\alpha s_{ll} d_{ll}^{(l)} = (\alpha s_{ll} + (1-\alpha) s_{jl}) d_{ll} \tag{3.8}$$

for the l-th equation and

$$(1-\alpha) s_{jj} d_{jj}^{(j)} = (\alpha s_{lj} + (1-\alpha) s_{jj}) d_{jj} \tag{3.9}$$

for the j-th equation. Again we put $d_{ii} = 0$, $i \neq l$, $i \neq j$. Equation (3.2) then are equivalent to

$$\left(\frac{1}{\alpha} - 1\right) \cdot \frac{s_{jl}}{s_{ll}} \geqq 0$$

which in turn is equivalent to

$$\arg\left(\frac{1}{\alpha} - 1\right) = \arg s_{ll} - \arg s_{jl} = \arg \tilde{s}_{jl} - \arg \tilde{s}_{ll} \tag{3.10}$$

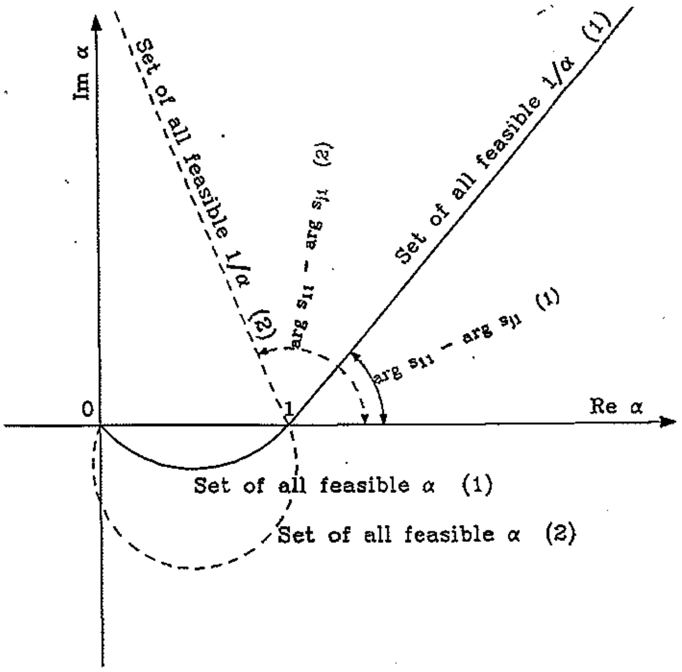

Fig. 1

Using the same argument as above, we can show that $\Sigma - D \geqq 0$ holds. Now let us discuss condition (3.10). We may distinguish between three cases.

(i) If $s_{jl}/s_{1l} > 0$ holds, then (3.10) is equivalent to $\alpha \in (0, 1]$.
(ii) If $s_{jl}/s_{1l} < 0$ holds, then (3.10) is equivalent to $\alpha \in [1, \infty) \cup (0, -\infty)$.
(iii) If s_{jl}/s_{1l} is not real, then the set of all feasible α is an arc of a circle as shown in Fig. 1.

For a more detailed discussion for the cases $n = 3$ and $n = 4$ see Scherrer et al. (1990).

4 The Relation Between Σ and L

Let $\sum_m$ denote the set of all $\Sigma > 0$, satisfying (vi) and (vii), such that $mc(\Sigma) = m$. In a next step we now consider some properties of $\sum_1$ and of solutions for varying (rather than for fixed) $\Sigma \in \sum_1$.

We will consider the set $\sum$ of all Σ satisfying (v)–(vii) as an open subset of $\mathbb{R}^{n^2-n}$. (Note that the set of all positive definite matrices is open in $\mathbb{R}^{n^2-n}$.)

Theorem 4.1. *$\sum_1$ is an open subset of $\sum$.*

Proof. We have to show that for every $\Sigma \in \sum_1$, there is a neighborhood contained in $\sum_1$. If this were not true, there would be a sequence Σ_n converging to Σ and $mc(\Sigma_n) > 1$ for all n. Thus, there would exist $\hat{\Sigma}_n$, $\hat{\Sigma}_n$ feasible for Σ_n and $rk\hat{\Sigma}_n < n-1$. By $\hat{\Sigma}_n \leqq \Sigma_n$ and $\Sigma_n \to \Sigma$, $\hat{\Sigma}_n$ is a bounded sequence, and thus has at least one limit point, $\hat{\Sigma}$ say. Since the determinant is a continuous function of the entries of a matrix, we then have $rk\,\hat{\Sigma} < n-1$ and $0 \leqq \hat{\Sigma} \leqq \Sigma$ and thus $mc(\Sigma) > 1$ in contradiction to our assumption.

From the idea of the proof above, we immediately obtain:

Corollary. $\sum_1 \cup \sum_2 \cdots \cup \sum_m$, $1 \leqq m \leqq n-1$, *is open in* $\sum$.

Consider the function 1, defined on $\sum_1$, which attaches to every Σ the corresponding (normalized) solution set $L = L_\Sigma$. For any two compact subsets, L, M say, of $\mathbb{C}^{(n-1)\times(n-1)}$ a metric can be defined by

$$d(L,M) = \sup(\rho(L,M), \rho(M,L)) \tag{4.1}$$

where

$$\rho(L,M) = \sup_{y \in M} \inf_{x \in L} \| y - x \| \tag{4.2}$$

d is called the Hausdorff distance (see, e.g. Dieudonne (1969), p. 61). By C we denote the set of all compact subsets of $\mathbb{C}^{n-1}$ endowed with the Hausdorff distance.

Theorem 4.2. *The function* $1: \sum_1 \to C: 1(\Sigma) = L_\Sigma$ *is continuous.*

Proof. Consider a sequence $\Sigma_n \to \Sigma \in \sum_1$. By Theorem 4.1 we may assume $\Sigma_n \in \sum_1$. We will show that

$$\sup_{x_n \in L_{\Sigma_n}} \inf_{x \in L_\Sigma} \| x_n - x \| \to 0 \tag{4.3}$$

As the norm is a continuous function and the (normalized) solution sets are compact we may replace the inf by min and sup by max.

As we are in the case $m^* = 1$ every principal minor up to size $n-1$ of $(\Sigma - D)$, $D \in \mathbf{D}$ is strictly positive. Since these minors are continuous functions of D and since $\mathbf{D}$ is compact by Theorem 2.1 all these principal minors are bounded away from zero by a positive constant. Furthermore, by the same type of argument we see that there is a compact neighborhood, $U_\delta := \{A \geqq 0 \mid \exists D \in \mathbf{D}$ s.t. $\|A - (\Sigma - D)\| \leqq \delta\}$, of the set $\{\Sigma - D \mid D \in \mathbf{D}\}$, where the same statement remains valid, i.e. there is a $\delta > 0$ and a corresponding $c > 0$ such that all principal minors in the set U_δ are uniformly bounded away from zero by c. In the following we will assume that $\varepsilon_1 > 0$ chosen sufficiently small so that we will stay in U_δ by performing the subsequent steps.

Since $\Sigma_n \to \Sigma$, for every $\varepsilon_1 \in (0,1)$ there is a n_0 such that

$$(1-\varepsilon_1)\Sigma \leqq \Sigma_n \leqq (1+\varepsilon_1)\Sigma, \qquad n \geqq n_0 \tag{4.4}$$

Let $\hat{\Sigma}$ and D be feasible for Σ and define

$$\check{D}_n=(1-\varepsilon_1)D, \qquad \check{\Sigma}_n=\Sigma_n-\check{D}_n$$

Then $\check{D}_n \geqq 0$ and

$$2\varepsilon_1\Sigma+(1-\varepsilon_1)\hat{\Sigma}=(1+\varepsilon_1)\Sigma-(1-\varepsilon_1)D \geqq \check{\Sigma}_n \geqq (1-\varepsilon_1)\check{\Sigma} \geqq 0 \tag{4.5}$$

Now let us perform the first elementary regression for $\check{\Sigma}_n$ which gives a feasible decomposition of the type (1.4)

$$\check{\Sigma}_n=\hat{\Sigma}_n+D_n^1, \qquad D_n^1=\operatorname{diag}\{\hat{d}_{11}^1,0,\ldots,0\}$$

Let $D_n=\check{D}_n+D_n^1$, then $\hat{\Sigma}_n$ and D_n are feasible for Σ_n.

Now, let us write, using an obvious notation [compare (2.8)]:

$$\begin{aligned}\|x_n-x\| &= \|\Sigma_{12}(\Sigma_{22}-D_{22})^{-1}-\Sigma_{12,n}\cdot(\Sigma_{22,n}-D_{22,n})^{-1}\| \\ &= [\det(\Sigma_{22}-D_{22})\cdot\det(\Sigma_{22,n}-D_{22,n})]^{-1}\cdot\|\Sigma_{12}\cdot\operatorname{adj}(\Sigma_{22}-D_{22}) \\ &\quad \cdot\det(\Sigma_{22,n}-D_{22,n})-\Sigma_{12,n}\cdot\operatorname{adj}(\Sigma_{22,n}-D_{22,n})\cdot\det(\Sigma_{22}-D_{22})\| \\ &\leqq c^{-2}\cdot\varepsilon_2\end{aligned} \tag{4.6}$$

where $\varepsilon_2>0$ is determined from ε_1 by (4.4) and (4.5); ε_2 can be chosen arbitrarily small by a suitable choice of ε_1 and can be chosen independently of the choice of $D\in\mathbf{D}$. Thus we have shown (4.3). By the same argument we can show

$$\sup_{x\in L_\Sigma}\ \inf_{x\in L_{\Sigma_n}} \|x_n-x\| \to 0 \tag{4.7}$$

which proves the theorem

Theorems 4.1 and 4.2 are important for a statistical analysis. In particular we see that a consistent estimator of Σ can directly be used to obtain a consistent estimator for L_Σ.

Next we consider the situation where the noise variance is converging to zero; in this case the solution sets are converging to a singleton:

Theorem 4.3. *Let* $\Sigma_n\to\Sigma$, *where* $\Sigma_n\in\sum_1$ *and* $rk\,\Sigma=n-1$, $L_\Sigma=\ker\Sigma\cap\{x\in\mathbb{C}^n|x_1=1\}\neq\emptyset$. *Then* $L_{\Sigma_n}\to L_\Sigma$.

Proof. Since Σ_n converges to a singular matrix, also the variances $d_{ii,n}^{(i)}$ of the corresponding noise terms for the elementary regressions converge to zero. Thus, by Lemma 2, we have $\hat{\Sigma}_n\to\Sigma$ for any $\hat{\Sigma}_n$ feasible for Σ_n. The rest is straightforward from

$$\|x_n-x\|=\|\Sigma_{12}\cdot\Sigma_{22}^{-1}-\Sigma_{12,n}\cdot(\Sigma_{22,n}-D_{22,n})\|^{-1}$$

References

Anderson, B.D.O. and M. Deistler, 1984, Identifiability in dynamic errors-in-variables models, *Journal of Time Series Analysis* 5, 1–13

Anderson, B.D.O. and M. Deistler, 1987, Dynamic errors-in-variables systems with three variables, Automatica 23, 611–616

Anderson, B.D.O. and M. Deistler, 1990, Identification of dynamic systems from noisy data: The case $m^* = n - 1$, Mimeo

Bekker, P. and J. de Leeuw, 1987, The rank of reduced dispersion matrices, Psychometrica 52, 125–135

Deistler, M., 1989, Symmetric modeling in system identification, in: H. Nijmeijer and J.M. Schumacher, eds., Three Decades of Mathematical System Theory. Springer Lecture Notes in Control and Information Sciences, no. 135, Springer-Verlag, Berlin

Deistler, M. and B.D.O. Anderson, 1989, Linear dynamic errors-in-variables models, some structure theory, Journal of Econometrics 41, 39–63

Dieudonne, J., 1969, Foundations of modern analysis. Academic Press, New York

Frisch, R., 1934, Statistical confluence analysis by means of complete regression systems, Publication no. 5 (Economic Institute, University of Oslo, Oslo)

Kalman, R.E., 1982, *System identification from noisy data*, in: A. Bednarek and L. Cesari, eds., Dynamic systems II, a University of Florida international symposium (Academic Press, New York, NY)

Kalman, R.E., 1983, Identifiability and modeling in econometrics, in: P.R. Krishnaiah, ed., Developments in statistics, Vol. 4 (Academic Press, New York, NY)

Koopmans, T.C., 1937, Linear regression analysis of economic time series, Netherlands Economic Institute, Haarlem

Picci, G. and S. Pinzoni, 1986, Dynamic factor-analysis models for stationary processes, IMA Journal of Mathematical Control and Information 3, 185–210

Scherrer, W., M. Deistler, M. Kopel and W. Reitgruber, 1990, Solution sets for linear dynamic errors-in-variables models, Mimeo

Willems, J.C., 1986, From time series to linear systems, Part I, Automatica 22, 561–580