Math. Control Signals Systems (1993) 6: 10–29

Mathematics of Control, Signals, and Systems

Identification of Dynamic Systems from Noisy Data: Single Factor Case*

Brian D. O. Anderson† and M. Deistler‡

Abstract. Linear dynamic errors-in-variables (or factor) models in the framework of stationary processes are considered. The noise process is assumed to have a diagonal spectral density. We analyze the relation between the (population) second moments of the observations and the system and noise characteristics; of particular interest are the number of equations (or the number of factors) and a description of the set of all systems compatible with the second moments of the observations. In this paper emphasis is put on the case which can be reduced to a single factor. The problems considered arise in the context of identification and noise modeling.

Key words. Identification, Errors-in-variables, Linear systems.

1. Introduction

In the 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-equation models. Here we are concerned with the case where in principle all variables, i.e., inputs as well as outputs, 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*. They have been analyzed and used in econometrics, psychometrics, statistics, and system engineering (see, e.g., [AD1], [AD2], [BL], [DA1], [EW], [G], and [K]).

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.

* Date received: November 11, 1990. Date revised: March 3, 1992. This research was supported by the Austrian "Fonds zur Förderung der wissenschaftlichen Forschung" "Schwerpunkt Angewandte Mathematik" (S32/02), and the Australian Commonwealth Government under the Cooperative Research Centres Program through the Cooperative Research Centre for Robust and Adaptive Systems.

† Department of Systems Engineering, Research School of Physical Sciences, Australian National University, GPO Box 4, Canberra, ACT 2601, Australia.

‡ University of Technology Vienna, Institute of Econometrics, Operations Research and Systems Theory, Argentinierstrasse 8, A-1040 Vienna, Austria.

(ii) If we want to approximate a high-dimensional data vector by a relatively small number of factors.
(iii) If we have insufficient *a priori* information about the number of independent 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 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

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

We call $w(z)$ the *relation function* of the exact relation (1.1) (compare [W]). 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 loss of generality we assume that $m \leq n$ holds (usually, $m < n$) and that $w(z)$ contains no linearly dependent rows; 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 assume:

(i) (z_t), $(\hat{z}_t)$, and (u_t) respectively are (wide sense) stationary processes (with real-valued components and) with spectral densities Σ, $\hat{\Sigma}$, and D, respectively. (In addition limits of random variables are understood in the sense of mean square convergence.)
(ii) $E\hat{z}_t = Eu_t = 0$ for all t.
(iii) $E\hat{z}_t u_s' = 0$ for all t and s.
(iv) D is diagonal.

For a discussion of assumption (iv), see, e.g., [DA1] and [D]. In our analysis, unless the contrary is stated explicitly, the frequency λ is kept fixed (i.e., we study (1.1) with $z = \exp i\lambda$). (We comment further on this assumption below.) In this sense Σ, $\hat{\Sigma}$, 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 of the form

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

where * denotes the conjugate transpose.

In this paper we analyze 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 problems are (compare [DA2]):

(a) Find the maximum number, say m^*, 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 of Σ explicit.
(b) Give a description of the set of all $(w(z), D)$ compatible with given Σ; in addition describe the subsets corresponding to different numbers of linear relations m.
(c) Describe the set of all Σ corresponding to a given m^*, $n > m^* \geq 1$.

Thus the problems we consider are (a) to find the (maximum) number of equations consistent with a 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 the main emphasis is on the case of general n and $m^* = n - 1$, i.e., on the case which can be modeled by a single factor. (For the "opposite" case, namely the one equation case, i.e., $m^* = 1$, see [H]).

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 where $w(z)$ is constant with real entries, this problem has a long history, beginning with the work of Charles Spearman, see, e.g., [S], [AR], and [BL]. For the dynamic case see [G] and [EW]. The dynamic cases $n = 2$ and $n = 3$ have been treated in detail in [AD1] and [AD2].

Let us now comment on the assumption that the frequency λ is kept fixed. There are practical problems where this is a natural assumption, and others where the assumption must be seen in a wider context. In sonar array processing, see, e.g., [H], an array of n sensors is assumed to provide outputs at a fixed frequency λ. The noise is often modeled as independent noise on each sensor, and if there is a single (noncoherent) transmitter (the detection of which is of interest), then the signal component of the array output has a covariance of rank 1. (Quite often, but not always, the hermitian dyadic expansion of this covariance is structured, i.e., it is not arbitrary, which means that additional information to that assumed here is available.) It may be that the signal processing system in use looks over a number of frequencies, and the received signal may be wideband (though otherwise without a dynamical structure). However, the narrowband case is common.

As a second example, we can consider a physical (mechanical) system subject to a periodic vibration (perhaps due to a motor). A number of sensors are located at different places in the system to measure the vibration response, and the spectrum at each sensor is narrowband, i.e., approximately a single spectral line. When there is a single vibration source we would expect $\hat{\Sigma}$ (associated with the vibration) to be of rank 1, and D (the measurement noise covariance associated with the vibration sensors) to be diagonal.

More generally, a true dynamical system, or collection of systems, giving rise to the data can be considered. Then our methods do no more than identify what is happening at discrete frequencies (which is what happens in normal, nonparametric spectral estimation). There is no dynamical order estimation (indeed no assumption of finite order); there is no attempt to deduce what is happening at frequency λ_0, using measurements at frequencies $\lambda_1, \lambda_2, \lambda_3, \ldots$. If a finite-order model is desired, there would have to be further steps of model building after determination of transfer function values at $\lambda_1, \lambda_2, \ldots$ (and such transfer function values are only obtainable in special cases, as discussed later). Actually, in this paper we are more concerned with identifying a test for the possibility that there is one "causative agent" (apart from measurement noise) for the observations (equivalent to rank $\hat{\Sigma} = 1$) than we are with "identification" in the sense that this term is commonly used by engineers, namely determining the values of parameters in a model whose structure is assumed *a priori*, as opposed to determining that structure.

The paper is organized as follows: In Section 2 we are concerned with problem (b) above, without giving a complete answer. Thereby no restriction on m^* is imposed. In Sections 3 and 4 we give a rather complete analysis of the case $m^* = n - 1$: we characterize the spectral densities Σ corresponding to $m^* = n - 1$ and we give a complete answer to (b) for this case.

Throughout we use the following notation: if A denotes a matrix, $\mathrm{rk}(A)$ and $\ker(A)$ denote its rank and its kernel, respectively; by $\dim_{\mathbb{R}}$ and $\dim_{\mathbb{C}}$ respectively we denote real and complex dimension.

2. The Solution Set—Some Definitions and General Properties

Clearly, relation (1.1) implies

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

If $\hat{\Sigma}$ were known and if we want to explain by the system as much as possible (in the sense that, for given $\hat{\Sigma}$, m is chosen as large as possible) and if we have no additional *a priori* information, then, by (2.1), the rows of w are defined by the requirement that they form an arbitrary basis of the left kernel of $\hat{\Sigma}$; thus w is unique up to basis change.

Clearly, in general only Σ is known and thus (1.4) is the starting point of our analysis. Remember that Σ, $\hat{\Sigma}$, and D are nonnegative definite and that $\hat{\Sigma}$ is singular and D is diagonal. We formalize this as follows.

Definition. *$\hat{\Sigma}$ and D are called feasible for given Σ if*

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

holds, where $\Sigma - D = \hat{\Sigma}$ is singular and D is diagonal.

As can be easily shown, for every $\Sigma \geq 0$ a feasible decomposition (1.4) exists and to every feasible decomposition EV representation exists. To avoid having to consider a number of special cases we assume, unless the contrary is stated explicitly, that

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

$$\tilde{S} = \Sigma^{-1}. \tag{2.3}$$

(As a general rule, if, e.g., Σ is a matrix, its i, j-entry is denoted by the corresponding lowercase symbol σ_{ij}.)

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

$$\tilde{x}\hat{\Sigma} = 0. \tag{2.4}$$

Definition. The set of all solutions corresponding to a given Σ is called the *solution set* $\tilde{L}$ (of Σ) and $\mathscr{D}$ denotes 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 $\tilde{x}$, $\tilde{x}_1$ is normalized to one. This leads to:

Definition. The (*normalized*) *solution set* L is given by

$$L = \{x \in \mathbb{C}^{n-1} | (1, x) \in \tilde{L}\}.$$

Some particular members of L play a major role in what follows.

Definition. Let

$$S = (\tilde{s}_{i1}^{-1}\tilde{s}_{ij}) \tag{2.5a}$$

$$= \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix}. \tag{2.5b}$$

Then s_j is termed the jth elementary solution.

Remark 2.1. To justify this nomenclature, observe that

$$s_j \cdot \Sigma = (0, \ldots, 0, \tilde{s}_{j1}^{-1}, 0, \ldots, 0) = s_j D_j \tag{2.6}$$

for some diagonal D_j and this defines s_j as the solution (with first component normalized to one) corresponding to the jth elementary regression, i.e., to the case where all components of z_t, except for the jth, are assumed to be free of noise. Note that

$$D_j = \operatorname{diag}(0, \ldots, 0, d_{jj}^{(j)}, 0, \ldots, 0), \tag{2.7}$$

where $d_{jj}^{(j)} = \tilde{s}_{jj}^{-1}$.

Remark 2.2. Since the first elementary solution s_1 always exists, no matrix Σ is excluded by the normalization $\tilde{x}_1 = 1$, i.e., every matrix Σ has a nonempty normalized solution set. However, there exist Σ and associated decompositions where 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. This situation cannot occur in the case $m^* = n - 1$ since (vi) implies that every row of the corresponding $\hat{\Sigma}$ (which has rank 1) can be expressed as a linear combination of every other row.

Remark 2.3. Clearly, elementary solutions can also be defined for singular matrices Σ. In general, they correspond to the projection of the jth component of z in (1.5) on the space spanned by all other components. For the following see also [DA2].

Now let us state some useful lemmas.

Lemma 2.1. *Let $\Sigma \geq 0$ (which is a relaxation of assumption (v)). If the nth row of Σ, σ_n say, is linearly independent of the other rows $\sigma_1, \ldots, \sigma_{n-1}$ of Σ, then the nth elementary regression yields a noise covariance matrix D of the form*

$$D_n = \operatorname{diag}\{0, \ldots, 0, d_{nn}^{(n)}\}, \tag{2.8}$$

where $d_{nn}^{(n)} > 0$ and where $\operatorname{rk}(\Sigma - D_n) = \operatorname{rk}(\Sigma) - 1$ holds. If σ_n is linearly dependent on the other rows, then the noise covariance matrix of the nth elementary regression is zero.

Proof. The proof is straightforward and involves projecting the nth component of z, see (1.5), on the linear space spanned by the other components of z. ■

Lemma 2.2. *Let $D = \operatorname{diag}\{d_{ii}\}$ be feasible and let $d_{ii}^{(i)}$ correspond to the ith elementary regression. Then*

$$0 \leq d_{ii} \leq d_{ii}^{(i)}. \tag{2.9}$$

If $d_{jj} > 0$ for some $j \neq i$, then the second inequality in (2.9) is strict.

Proof. Without loss of generality, take $i = 1$; let $D = A + B$ where $A = \operatorname{diag}\{d_{11}, 0, \ldots, 0\}$ and $B = \operatorname{diag}\{0, d_{22}, \ldots, d_{nn}\}$. First note that if it were true that $d_{11} > d_{11}^{(1)}$, the matrix $\Sigma - A$ would not be nonnegative definite. To see this consider

$$\det(\Sigma - A) = (\sigma_{11} - d_{11}) \cdot f_1(\Sigma) + f_2(\Sigma), \tag{2.10}$$

where f_1 and f_2 depend only on Σ and where

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

holds. Thus $\det(\Sigma - A)$ is zero if $d_{11} = d_{11}^{(1)}$; the form of (2.10) then shows that $\det(\Sigma - A)$ is zero only if $d_{11} = d_{11}^{(1)}$, and then it is negative for $d_{11} > d_{11}^{(1)}$.

Now $\Sigma - D = \Sigma - A - B = C \geq 0$ implies $B + C = \Sigma - A \geq 0$ which would be contradicted if $d_{11} > d_{11}^{(1)}$.

To prove the second section of the lemma, perform the jth elementary regression for $(\Sigma - D_i)$, where D_i is the noise matrix corresponding to the ith elementary regression; since all elements of s_i are unequal to zero by (vii) and (2.5), and $s_i(\Sigma - D_i) = 0$, the jth row of $\Sigma - D_i$ is linearly dependent on the other rows of $(\Sigma - D_i)$. Hence, by Lemma 2.1, any diagonal matrix Δ for which $0 \leq \Delta \leq \Sigma - D_i$ has $\Delta = 0$. Equivalently, the last statement of the lemma is proved. ■

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

$$\tilde{x}\Sigma = \tilde{x}D, \qquad \tilde{x} \in \tilde{L}, \qquad D \in \mathscr{D}. \tag{2.11}$$

In order to investigate the solution set further, let us connect two points, x, $y \in \tilde{L}$, by the complex line

$$\alpha x + (1 - \alpha)y, \qquad \alpha \in \mathbb{C}. \tag{2.12}$$

We now wish to consider the question: for what values of α is it true that $\alpha x + (1 - \alpha)y \in \tilde{L}$? The results obtained in Lemma 2.3 below are valid for general $mc(\Sigma)$, of course under assumptions (i)–(vii), and apply when x and y coincide with elementary solutions. In general, we need to start from the equation

$$(\alpha x + (1 - \alpha)y)\Sigma = (\alpha x + (1 - \alpha)y)D = \alpha x D_x + (1 - \alpha)y D_y, \tag{2.13}$$

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 (2.13). Clearly, $\alpha x + (1 - \alpha)y \in \tilde{L}$ if and only if there is a D satisfying (2.13), $D \geq 0$, and $\Sigma - D \geq 0$.

Of course, the set $\alpha x + (1 - \alpha)y$ for $\alpha \in \mathbb{C}$ is a plane in $\mathbb{R}^{2n} = \mathbb{C}^n$. In the following lemma we identify particular line segments or other curves in this plane.

Lemma 2.3. *With notation as above, consider the set $\alpha s_k + (1 - \alpha)s_j$, with s_k and s_j the* kth *and* jth *elementary solution.*

(i) *If $s_{jk}/s_{kk} > 0$, then $\alpha s_k + (1 - \alpha)s_j$ is a solution if and only if $\alpha \in [0, 1]$, i.e., the solutions are precisely the line segment connecting s_k and s_j.*

(ii) *If $s_{jk}/s_{kk} < 0$, then $\alpha s_k + (1 - \alpha)s_j$ is a solution if and only if $\alpha \in (-\infty, 0] \cup [1, \infty)$, i.e., the solutions are precisely the part of the line connecting s_k and s_j which lies outside the segment joining the points.*

(iii) *If s_{jk}/s_{kk} is not real, then $\alpha s_k + (1 - \alpha)s_j$ is a solution if and only if*

$$\arg(\alpha^{-1} - 1) = \arg s_{kk} - \arg s_{jk}, \tag{2.14}$$

which implies that α lies on an arc of a circle.

In case $k = 1$, $s_{jk}/s_{kk} > 0$ and case (i) *applies.*

Proof. (i) $\alpha \in [0, 1]$. To help fix ideas, let us 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 jth elementary solution. Then the first element in the second equation in (2.13) is of the form

$$d_{11} = \alpha d_{11}^{(1)} \tag{2.15}$$

and the jth element gives

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

which gives

$$\frac{1}{1 + (\alpha/(1 - \alpha))(s_{1j}/s_{jj})} \cdot d_{jj}^{(j)} = d_{jj}. \tag{2.16}$$

By Lemma 2.2, $d_{jj}^{(j)} \geq d_{jj} \geq 0$ must hold for every feasible D. Also note that $s_{1j} \cdot s_{jj}^{-1} > 0$; thus (2.15) and (2.16) imply $\alpha \in [0, 1]$.

Now we show that for every $\alpha \in [0, 1]$ we can find a diagonal $D \geq 0$ with $\Sigma - D \geq 0$ satisfying (2.13). Put

$$d_{ii} = 0 \qquad \text{for} \quad i \neq 1, \quad i \neq j. \tag{2.17}$$

Then such a prescription for D satisfies (2.13) and $D \geq 0$ for every $\alpha \in [0, 1]$. In order to show that a D given by (2.15)–(2.17) is feasible for every $\alpha \in [0, 1]$, it remains to show that $\Sigma - D \geq 0$ holds. Note that, for the jth elementary regression, $d_{jj}^{(j)}$ is the *unique* solution of the equation

$$\det(\Sigma - \operatorname{diag}(0, \ldots, d_{jj}, 0, \ldots, 0)) = 0 \tag{2.18}$$

in the variable $d_{jj} \in \mathbb{R}$. This is a direct consequence of the fact that (2.18) is a linear equation with a positive coefficient for $(\sigma_{jj} - d_{jj})$ (compare (2.10)). Now performing the jth elementary regression for $\Sigma - \alpha D_1$, $\alpha \in (0, 1)$, we see that the corresponding noise covariance matrix is $\operatorname{diag}\{0, \ldots, d_{jj}, 0, \ldots, 0\}$ with d_{jj} given by (2.16) and thus $\Sigma - D = (\Sigma - \alpha D_1) - \operatorname{diag}\{0, \ldots, d_{jj}, 0, \ldots, 0\} \geq 0$.

(ii) and (iii) We first exhibit the condition which α must satisfy. We proceed in an analogous way as before in the case $x = s_k$, $y = s_j$, $k, j \neq 1$; $k \neq j$. Then from (2.13) we obtain

$$\alpha s_{kk} d_{kk}^{(k)} = (\alpha s_{kk} + (1 - \alpha) s_{jk}) d_{kk} \tag{2.19}$$

for the kth equation and an analogous term for the jth equation. Since $d_{kk}^{(k)} \geq d_{kk}$, (2.19) implies

$$\left(\frac{1}{\alpha} - 1\right) \cdot \frac{s_{jk}}{s_{kk}} \geq 0,$$

which, for negative s_{jk}/s_{kk}, implies $\alpha \in (-\infty, 0] \cup [1, \infty)$ and, for not real s_{jk}/s_{kk}, is equivalent to (2.14).

We now argue the necessity of the conditions in (ii) and (iii). We follow the argument as for case (i), identifying d_{kk} via (2.19) and d_{jj} similarly, with $d_{ii} = 0$ for $i \neq j$, $i \neq k$. With minor adjustment to the argument for case (i), we can show that $\Sigma - D \geq 0$ holds. ■

Remark 2.4. It is easily verified that the three conditions in the lemma are all consistent with (2.14); i.e., if $s_{jk}/s_{kk} > 0$ or $s_{jk}/s_{kk} < 0$, then (2.14) implies $\alpha \in [0, 1]$ or $\alpha \in (-\infty, 0] \cup [1, \infty)$, respectively.

Remark 2.5. We can actually also obtain a small result for the case when $x = s_1$ and $y \in F \cap \tilde{L}$ where

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

For every $y \in F \cap \tilde{L}$, there exists a corresponding feasible D_y with $d_{11} = 0$. The first element in the second equation of (2.13) yields $d_{11} = \alpha d_{11}^{(1)}$, and so we see it is necessary (but now not sufficient) that $\alpha \in [0, 1]$.

3. The Spectral Densities of the Observations for the Single Factor Case

In this and in the next section we give a complete analysis of the case $m^* = n - 1$, i.e., the case which can be described by one single factor: We give a characterization of Σ in this case, a description of the set of all Σ corresponding to $m^* = n - 1$, and finally a description of the solution set (for given Σ, also in terms of different coranks and not just $m = n - 1$). Part of the results have been available for the static case for a long time (see, e.g., [BL]). For the dynamic $n = 3$ case see [AD2]. The next theorem gives a characterization of the case $m^* = n - 1$.

Theorem 3.1. *For $n > 3$ the following statements are equivalent:*

(a) *$mc(\Sigma) = n - 1$, $\hat{\Sigma}$, D feasible.*

(b) *There exists a diagonal unitary matrix U such that $U^*\Sigma U = (\tau_{ij})$, $i, j = 1, \ldots, n$, is real with all entries τ_{ij} positive and satisfying*

$$\tau_{ik} \cdot \tau_{jl} - \tau_{il} \cdot \tau_{jk} = 0, \qquad i, j, k, l \text{ all different}, \tag{3.1}$$

$$\tau_{ii} \cdot \tau_{jk} - \tau_{ik} \cdot \tau_{ji} \geq 0, \qquad i, j, k \text{ all different}. \tag{3.2}$$

(c) *There exists a diagonal unitary matrix U such that $U^*\Sigma^{-1}U = (t_{ij})$, $i, j = 1, \ldots, n$, is real with all off-diagonal elements negative and satisfying*

$$t_{ik} \cdot t_{jl} - t_{il} \cdot t_{jk} = 0, \qquad i, j, k, l \text{ all different}. \tag{3.3}$$

Proof. (a) implies (b). Since there is a feasible $\hat{\Sigma}$ with rank equal to one, we can write

$$\hat{\Sigma} = \lambda\lambda^*, \qquad \lambda \in \mathbb{C}^n.$$

In addition, for $U = \operatorname{diag}\{e^{i \arg \lambda_j}\}$, where $\arg \lambda_j$ denotes the phase of the jth entry of λ, $\hat{\lambda} = U^*\lambda$ has all elements positive (Note: since Σ has all entries nonzero, so has λ.) Thus all entries in

$$U^*\Sigma U = \hat{\lambda}\hat{\lambda}' + D \tag{3.4}$$

are positive. Equations (3.1) and (3.2) are straightforward.

(b) implies (a). Consider the matrix $T = \{\tau_{ij}\} = U^*\Sigma U$; then we can find a decomposition $T = \hat{T} + \tilde{T}$ defined by $\hat{t}_{ij} = \tau_{ij}$, for $i \neq j$, and by

$$\hat{t}_{ii} = \tau_{ik} \cdot \tau_{ji} \cdot \tau_{jk}^{-1}, \qquad i, j, k, \text{ all different}.$$

Then as can be easily checked from (3.1), $\hat{t}_{ii}$ is independent of j, k and all minors of $\hat{T}$ of order 2 are equal to zero and thus $\hat{T}$ has rank 1. Clearly, $\hat{t}_{ii} > 0$ since $\tau_{ij} > 0$, $i, j = 1, \ldots, n$, and thus $T \geq 0$. Furthermore, $\tilde{T} \geq 0$ holds by (3.2) and thus $\hat{T}$ and $\tilde{T}$ are feasible for T, which immediately implies that $U\hat{T}U^*$ and $U\tilde{T}U^*$ are feasible for Σ.

(b) implies (c). Assume the decomposition (3.4). Suppose temporarily that $D > 0$ holds. It is easy to check that (3.4) implies

$$(U^*\Sigma U)^{-1} = D^{-1} - \frac{1}{1 + \hat{\lambda}'D^{-1}\hat{\lambda}} \cdot D^{-1}\hat{\lambda}\hat{\lambda}'D^{-1} = D^{-1} - \hat{\mu}\hat{\mu}', \tag{3.5}$$

where $\hat{\mu}$ is a vector with all positive entries. It is immediate that all off-diagonal entries of $(U^*\Sigma U)^{-1}$ are negative.

Now suppose $D \geq 0$. Since Σ is nonsingular, the entires of $(U^*(\Sigma + \varepsilon I)U)^{-1}$, $\varepsilon \geq 0$, depend continuously on ε. For all $\varepsilon > 0$ the off-diagonal elements are negative; hence in the limit for $\varepsilon \to 0$ they are nonpositive and thus by assumption (vii) negative also.

Equation (3.3) directly follows from (3.4).

(c) implies (b). Let $T = (t_{ij}) = U^*\Sigma^{-1}U$ and $T = \hat{T} + \tilde{T}$, where $\hat{T}$ is defined by $\hat{t}_{ij} = t_{ij}$, $i \neq j$, and

$$\hat{t}_{ii} = \frac{\hat{t}_{ik} \cdot \hat{t}_{ji}}{\hat{t}_{jk}}, \qquad i, j, k, \text{ all different.}$$

Again it is easy to see, using (3.3), that $\hat{t}_{ii}$ is independent of j, k and that $\hat{T}$ has rank 1. Furthermore, since all t_{ij}, $i \neq j$, are negative, $\hat{T}$ consists of negative elements only and thus can be written as

$$\hat{T} = -\beta\beta', \qquad \beta \in \mathbb{R}^n_+ \quad (\text{i.e., } \beta_i > 0),$$

$\tilde{T}$ then is diagonal with strictly positive diagonal elements. From (3.5) we then obtain

$$T^{-1} = U^*\Sigma U = \tilde{T}^{-1} + \frac{1}{1 + \beta'\tilde{T}^{-1}\beta} \cdot \tilde{T}^{-1}\beta\beta'\tilde{T}^{-1},$$

which implies (b). ■

Remark 3.1. Note that, for $m^* = n - 1$,

$$x(\Sigma - D) = 0 \qquad \text{if and only if} \quad xUU^*(\Sigma - D)U = 0$$

(with U defined as above). Hence, for every complex problem, x, U gives a one-to-one relation to a solution with a real Σ. Note that in general the entries of x, U will be complex.

Remark 3.2. If we drop assumptions (vi) and (vii), then it can be argued that Theorem 3.1 remains true if we replace (b) by:

(b*) There exists a diagonal unitary matrix U such that $U^*\Sigma U = (\tau_{ij})$, $i, j = 1, \ldots, n$, is real with all entries τ_{ij} nonnegative and satisfying (3.1), (3.2) and

$$\tau_{ij} > 0, \qquad i \neq j, \quad \text{for at least one pair } (i, j).$$

and (c) by:

(c*) Either (c) holds, or there exists a diagonal unitary matrix U such that $U^*\Sigma^{-1}U = (t_{ij})$, $i, j = 1, \ldots, n$, is real with all off-diagonal entries in one row and column negative, and all other off-diagonal entries zero.

Remark 3.3. Several observations can easily be made concerning the set Σ_{n-1} of all Σ (satisfying our assumptions) with $mc(\Sigma) = n - 1$. Consider the mapping i defined by $i(\lambda, D) = \lambda\lambda^* + D$, where $\lambda \in \mathbb{C}^n$ and $D \geq 0$, diagonal. If we impose the additional normalization assumption $\lambda_1 > 0$, it is easily seen from the condition that all 2×2 minors of $\hat{\Sigma}$ are zero, that, for $n \geq 3$ (and imposing (vi)), λ and D are uniquely determined from Σ, and that i and i^{-1} are continuous. Thus Σ_{n-1} is a subset of real dimension $3n - 1$ of the set of all $\Sigma > 0$, which is of real dimension n^2 and i^{-1} provides a useful parametrization for Σ_{n-1}. In particular, we see that, for $n \geq 3$, the single-factor case is highly nongeneric for the spectral densities of the observations.

4. The Solution Set for the Single-Factor Case

In this section we give a complete description of the set of all solutions corresponding to a given Σ with $mc(\Sigma) = n - 1$. We also provide a characterization of solutions corresponding to different m.

We need some further notation.

Definition. L_m denotes the set of $x \in L$ for which there exists a feasible $\hat{\Sigma}$ with $(1, x)\hat{\Sigma} = 0$ and $\dim_{\mathbb{C}} \ker \hat{\Sigma} = m$.

Definition. For $z = (z_1, z_2, \ldots, z_n) \in \mathbb{C}^n$, the projection π is defined by $\pi(z) = (z_2, \ldots, z_n) \in \mathbb{C}^{n-1}$ and $\pi(M) = \{\pi(x) | x \in M\}$ for a set $M \subset \mathbb{C}^n$.

In the following theorem we are concerned with characterizing, for the case $mc(\Sigma) = n - 1$, not just those solutions corresponding to feasible $\hat{\Sigma}$ of corank $n - 1$ but also those corresponding to feasible $\hat{\Sigma}$ of corank $n - 2, n - 3, \ldots, 1$. Among the conclusions of the theorem, one that stands out is that the solution set L_{n-1} associated with feasible $\hat{\Sigma}$ of rank 1 includes the solution set associated with feasible $\hat{\Sigma}$ of all ranks between 2 and $n - 2$, and has a nonempty intersection with the solution set L_1 associated with feasible $\hat{\Sigma}$ of rank $n - 1$, which is in general a strict subset of both L_1 and L_{n-1}.

Theorem 4.1. *Let $n \geq 3$ and $mc(\Sigma) = n - 1$. Then:*

(a) *There is a unique feasible D, $D^* = \operatorname{diag}\{d_{ii}^*\}$ say, given by*

$$d_{ii}^* = \sigma_{ii} - \frac{\sigma_{ik} \cdot \sigma_{ji}}{\sigma_{jk}}, \qquad i, j, k \text{ all different,} \tag{4.1}$$

such that $\Sigma - D^*$ has rank equal to one; in this case

$$L_{n-1} = \pi(\ker(\Sigma - D^*) \cap \{x \in \mathbb{C}^n | x_1 = 1\}) \tag{4.2}$$

and

$$\dim_{\mathbb{C}}(L_{n-1}) = n - 2 \tag{4.3}$$

hold.

(b) *Let*

$$x_j = \alpha^* s_1 + (1 - \alpha^*) s_j, \qquad j = 2, \ldots, n, \tag{4.4}$$

where

$$\alpha^* = \left[1 - \frac{s_{1i}}{s_{ji}}\right]^{-1}, \qquad 1, i, j \text{ all different,} \tag{4.5}$$

and α^* is independent of i and j. Then all entries of x_j not in positions 1 and j are equal to zero and

$$L_{n-1} = \pi\left[\left\{\sum_{j=2}^{n} \beta_j x_j \middle| \sum_{2}^{n} \beta_j = 1\right\}\right]. \tag{4.6}$$

(c) *For every m such that $1 < m < n-1$, L_m is a nonempty subset of L_{n-1}, additionally satisfying $L_m \subset L_{m+1}$. In particular,*

$$L_m = \pi\left[\sum_{j=2}^{n} \beta_j x_j \middle| \sum_{2}^{n} \beta_j = 1, \text{at most } m\ \beta_j \text{ are nonzero}\right]. \tag{4.7}$$

The set of diagonal noise matrices D in decompositions $\Sigma = \hat{\Sigma} + D$, where $\dim_{\mathbb{C}} \ker(\Sigma - D) = m$, is given by $D = D^* - \tilde{D}$, where $0 \le \tilde{D} \le D^*$ and $\tilde{D}$ has precisely $(m+1)$ zero entries, including the first. Moreover, $\ker(\Sigma - D)$ is fully determined by these $(m+1)$ entries, and is otherwise independent of $\tilde{D}$.

(d) *L_1 can be written as $L_1 = L_{1,U} \cup L_{1,0} \cup L_{1,L}$, where*

$$L_{1,U} = \pi\left(\left\{\alpha s_1 + (1-\alpha) \sum_{j=2}^{n} \beta_j x_j \middle| \alpha \in (0,1], \beta \in [0,1], \sum_{2}^{n} \beta_j = 1\right\}\right), \tag{4.8}$$

$$L_{1,0} = \pi\{x_2, x_3, \ldots, x_n\}, \tag{4.9}$$

and

$$L_{1,L} = \pi\left(\left\{\alpha s_1 + (1-\alpha) \sum_{j=2}^{n} \beta_j x_j | \alpha \in [\alpha_{\min}, 0), \alpha_{\min} = -\alpha^*(1-\alpha^*)^{-1}, \right.\right.$$
$$\left.\left.\sum_{2}^{n} \beta_j = 1, \beta_k \ge 1 \text{ for some } k \ge 2 \text{ and } \beta_j \le 0 \text{ for } j \ne 1, j \ne k\right\}\right) \tag{4.10}$$

which can be reexpressed as

$$L_{1,L} = \pi\left(\left\{\alpha s_1 + (1-\alpha) \sum_{j=2}^{n} \beta_j s_j | \alpha \in [0, \alpha^*), \right.\right.$$
$$\left.\left.\sum_{2}^{m} \beta_j = 1, \beta_k \ge 1 \text{ for some } k \ge 2 \text{ and } \beta_j \le 0 \text{ for } j \ne 1, j \ne k\right\}\right)$$

Moreover, $L_{1,U} \cap L_{n-1} = \varphi$, $L_{1,L} \cap L_{n-1} = \varphi$, and $L_{1,0} \cap L_{n-1} = L_{1,0}$.

Proof. (a) The first part of (a) directly follows from the fact that all 2×2 minors of $(\Sigma - D^*)$ have to be zero and from assumption (vi) (set out below (2.2)), which together imply that D^* is uniquely determined. In order to show (4.3) note that the intersection between $\ker(\Sigma - D^*)$ and $\{x \in \mathbb{C}^n | x_1 = 1\}$ is always nonempty: Since $\Sigma - D^*$ has rank equal to one and (vi) holds, it is always possible to express the first row of $\Sigma - D^*$ as a linear combination of the others. There are $(n-1)$ linearly independent vectors y_i say with first entry 1 in $\ker(\Sigma - D^*)$. The set of linear combinations $\sum_{i=1}^{n-1} \beta_i y_i$ forms an $(n-1)$-dimensional space, but the constraint $\sum_{i=1}^{n-1} \beta_i = 1$ reduces $\dim_{\mathbb{C}}(L_{n-1})$ to $n-2$.

(b) Consider the equation system

$$\alpha s_1 D_1 + (1-\alpha)s_j D_j = (\alpha s_1 + (1-\alpha)s_j)D \tag{4.11}$$

(with $s_1\Sigma = s_1 D_1$, $s_j\Sigma = s_j D_j$). From Section 2 we know that, for every $\alpha \in [0, 1]$, $x = \alpha s_1 + (1-\alpha)s_j$ is a solution.

Let $U = \operatorname{diag}(u_j)$ in statement (c) of Theorem 3.1. Then, for 1, i, and j all different,

$$\begin{aligned}\frac{s_{1i}}{s_{ji}} &= \frac{\tilde{s}_{1i}\tilde{s}_{11}^{-1}}{\tilde{s}_{ji}\tilde{s}_{j1}^{-1}}\\ &= \frac{(u_1 t_{1i} u_i^*)(u_j t_{j1} u_1^*)}{u_j t_{ji} u_i^* t_{11}}\\ &= \frac{t_{1i}t_{j1}}{t_{ji}t_{11}}\\ &< 0.\end{aligned}$$

Using (3.3), it is trivial to show that the value of s_{1i}/s_{ji} for 1, i, and j all different is independent of i and j.

Hence if we put $\alpha^* = (1 - s_{1i}\cdot s_{ji}^{-1})^{-1}$, then $\alpha^* \in (0, 1)$. Furthermore, if $k \neq 1, k \neq j$,

$$\begin{aligned}(x_j)_k &= \alpha^* s_{1k} + (1-\alpha^*)s_{jk}\\ &= \frac{1}{1 - t_{1i}t_{j1}/t_{ji}t_{11}}\cdot\frac{u_1 t_{1k} u_k^*}{t_{11}} - \frac{t_{1i}t_{j1}/t_{ji}t_{11}}{1 - t_{1i}t_{j1}/t_{ji}t_{11}}\cdot\frac{u_j t_{jk} u_k^*}{u_j t_{j1} u_1^*}\\ &= 0\end{aligned} \tag{4.12}$$

on using (3.3). Now define $D = \hat{D}$ in (4.11) by

$$\hat{d}_{11} = \alpha^* d_{11}^{(1)}, \qquad \hat{d}_{jj} = (1 + \alpha(1-\alpha)^{-1}s_{1j}\cdot s_{jj}^{-1})^{-1} d_{jj}^{(j)} \tag{4.13}$$

(see (2.15) and (2.16)) and, for $i \neq 1, j$, define

$$\hat{d}_{ii} = \sigma_{ii} - \frac{\sigma_{ik}\sigma_{1i}}{\sigma_{1k}}, \qquad k \neq i, \quad 1 \neq i, \quad 1 \neq k. \tag{4.14}$$

We assert that $\hat{D} = D^*$, as defined in (4.1). To see this, note that, because all entries of x_j other than the first and jth are zero, the 2×2 minors of $\Sigma - \hat{D}$ formed from rows 1 and j are all zero. The only way this can hold is if $\hat{d}_{11} = d_{11}^*$ and $\hat{d}_{jj} = d_{jj}^*$, as in (4.1). Also, for $i \neq 1, j$, the definition of (4.14) coincides with (4.1).

Hence we see that $x_j \in \ker(\Sigma - D^*)$ for $j = 2, \ldots, n$ and, by (4.2) and (4.3), the x_j generate L_{n-1}.

(c) First we show that if D is a feasible matrix with $\dim_{\mathbb{C}} \ker(\Sigma - D) = m$, $1 < m < n - 1$, then D must be of the form

$$D = D^* - \tilde{D}, \tag{4.15}$$

where $\tilde{D}$ is a nonnegative diagonal matrix.

This conclusion follows from the fact that, for every feasible D, all 2×2 principal minors of $\Sigma - D$ are nonnegative. On the other hand, all 2×2 principal minors of $\Sigma - D^*$ are zero. Hence $(\sigma_{ii} - d_{ii})(\sigma_{jj} - d_{jj}) \geq (\sigma_{ii} - d_{ii}^*)(\sigma_{jj} - d_{jj}^*)$ for all $i \neq j$. Then it is impossible to have $d_{ii}^* < d_{ii}$ and $d_{jj}^* < d_{jj}$, i.e., D^* could have at most one element smaller than the corresponding element in D. Moreover, if there was such a strictly smaller element in D^*, then all other elements in D^* would be strictly larger than the corresponding elements in D. Suppose, without loss of generality, $d_{ii}^* > d_{ii}$ for all $i > 1$. Now the last $(n-1)$ rows and columns of $\Sigma - D^*$ necessarily form a nonnegative definite matrix. Hence the last $(n-1)$ rows and columns of $\Sigma - D$ will form a positive definite matrix, i.e., $\dim_{\mathbb{C}} \ker(\Sigma - D) = 1$. Hence D with $\dim_{\mathbb{C}} \ker(\Sigma - D) = m$, $1 < m < n - 1$, must be of the form (4.15) with $\tilde{D} \geq 0$.

Now we show that the $\tilde{D}$ associated with D when $\dim_{\mathbb{C}} \ker(\Sigma - D) = m$ has precisely $(m+1)$ zero diagonal entries. Recall also that $\ker(\Sigma - D)$ must contain a vector with 1 as the first entry. Now observe that for two matrices $A \geq 0$ and $B \geq 0$ we have

$$\ker(A + B) = \ker A \cap \ker B$$

and so

$$\ker(\Sigma - D) = \ker[(\Sigma - D^*) + \tilde{D}] = \ker(\Sigma - D^*) \cap \ker \tilde{D}.$$

We have shown that $\ker(\Sigma - D^*)$ is spanned by the $n-1$ vectors x_j defined in (4.4), and x_j has zero entries in positions other than 1 and j.

In order that there be any vector in $\ker(\Sigma - D)$ with 1 as the first entry, the first diagonal entry of $\tilde{D}$ must evidently be zero. Suppose also that the diagonal entries $j_1, j_2, \ldots, j_\alpha$ for $1 < j_1 < j_2 \cdots < j_\alpha \leq n$ are zero; then $\ker \tilde{D}$ is evidently spanned by the unit vector with 1 in the first entry and $x_{j1}, x_{j2}, \ldots, x_{j\alpha}$. Hence $\ker(\Sigma - D)$ will be spanned by $x_{j1}, x_{j2}, \ldots, x_{j\alpha}$. In order that the kernel of $\Sigma - D$ have dimension m, it is necessary that $\alpha = m$, i.e., that $\tilde{D}$ have precisely $(m+1)$ zero diagonal entries, including the first.

Now we show that $\ker(\Sigma - D)$ is fully determined by the zero diagonal entries of any $\tilde{D}$ with $0 \leq \tilde{D} \leq D^*$. Thus suppose that $\tilde{D}$ satisfies $0 \leq \tilde{D} \leq D^*$, and $\dim_{\mathbb{C}} \ker(\tilde{D}) = m + 1$, with $\tilde{d}_{11} = 0$. Then we certainly have a decomposition $\Sigma = \hat{\Sigma} + D$, with $0 \leq D = D^* - \tilde{D}$, $0 \leq \hat{\Sigma} \leq \Sigma$, and $\ker \hat{\Sigma} = m$. Thus we have a complete characterization of all feasible D yielding $\dim_{\mathbb{C}} \ker(\Sigma - D) = m$.

As noted above, if the diagonal entries $j_1, \ldots, j_m$ of $\tilde{D}$ are zero (as well as $\tilde{d}_{11}$), then $\ker(\Sigma - D)$ is spanned by $x_{j1}, x_{j2}, \ldots, x_{jm}$. Conclusion (c) and the nesting property are now immediate.

(d) Consider the equation system

$$\begin{aligned}\left(\alpha s_1 + (1-\alpha)\cdot \sum_{j=2}^{n} \beta_j x_j\right)\Sigma &= \alpha s_1 D_1 + (1-\alpha)\left[\sum_{2}^{n} \beta_j x_j\right] D^* \\ &= \left(\alpha s_1 + (1-\alpha)\sum_{2}^{n} \beta_j x_j\right) D, \\ \sum_{2}^{n} \beta_j = 1. & \end{aligned} \tag{4.16}$$

Notice that any normalized solution x will be expressible as $\alpha s_1 + (1-\alpha)\sum_{j=2}^{n} \beta_j x_j$ with $\sum_2^n \beta_j = 1$, since $s_1, x_2, \ldots, x_n$ are linearly independent, and the first component of the weighted sum is 1. (On the other hand, not every x of this form has to be a solution, of course.)

The first entry of the second equation in (4.16) gives

$$\alpha(d_{11}^{(1)} - d_{11}^{*}) = (d_{11} - d_{11}^{*}) \tag{4.17}$$

and the remaining equations give

$$d_{jj} = d_{jj}^{*}\left(1 + \frac{\alpha s_{1j}}{(1-\alpha)\beta_j(x_j)_j}\right)^{-1}, \qquad j \geq 2, \tag{4.18}$$

where $(x_j)_j$ denotes the jth component of x_j. From Lemma 2, we know that $d_{11}^{(1)} > d_{11}^{*}$ and $d_{11}^{(1)} \geq d_{11}$. It follows that $\alpha \leq 1$.

We next consider separately the case $\alpha \in (0, 1]$, $\alpha = 0$, and $\alpha < 0$. For the case $\alpha > 0$, from (4.17) we have $d_{11} > d_{11}^{*}$. Thus by the same argument as in the first paragraph of the proof of (c) we see that if $\alpha > 0$ gives a solution, then it can only correspond to $m = 1$. Furthermore, $d_{11} > d_{11}^{*}$ implies $d_{jj} < d_{jj}^{*}$ for all $j > 1$ and as can be easily checked $s_{1j} \cdot (x_j)_j^{-1} > 0$ holds. Thus (4.18) implies $\beta_j \geq 0$ for $\alpha \in (0, 1)$. Since $\sum_2^n \beta_j = 1$, evidently, $\beta_j \in [0, 1]$.

Next we have to show that, for every $\alpha \in (0, 1]$, $\beta_j \in [0, 1]$, $\sum_2^n \beta_j = 1$, the element $\alpha s_1 + (1-\alpha)[\sum_2^n \beta_j x_j]$ is actually a solution. Observe from (4.16) that, with d_{jj} as defined in (4.17) and (4.18),

$$\left(\alpha s_1 + (1-\alpha)\sum_{j=2}^{n}\beta_j x_j\right)[\Sigma - \operatorname{diag}(0, d_{22}, \ldots, d_{nn})]$$
$$= \left(\alpha s_1 + (1-\alpha)\sum_{j=2}^{n}\beta_j x_j\right)\operatorname{diag}(d_{11}, 0, \ldots, 0).$$

Also $0 \leq \Sigma - \operatorname{diag}(d_{11}^{*}, d_{22}^{*}, \ldots, d_{nn}^{*}) < \Sigma - \operatorname{diag}(0, d_{22}, \ldots, d_{nn})$ since (4.18) yields $d_{jj} < d_{jj}^{*}$. So $\alpha s_1 + (1-\alpha)\sum_{j=2}^{n}\beta_j x_j$ is the first elementary regression for $\Sigma - \operatorname{diag}(0, d_{22}, \ldots, d_{nn})$ and $\Sigma - D$ is nonnegative. This shows the solution property for $L_{1,U}$.

Now, we consider the case $\alpha = 0$. Then $d_{11} = d_{11}^{*}$ from (4.17); furthermore, $d_{jj} = d_{jj}^{*}$ from (4.18), provided $\beta_j \neq 0$. To understand the possibilities when one or more $\beta_j = 0$, return to (4.16), which says (with $\alpha = 0$)

$$\sum_2^n \beta_j x_j D^{*} = \sum_2^n \beta_j x_j D.$$

If $\beta_j = 0$, then this equation constitutes no constraint on d_{jj}. The argument in the proof of (c) just after (4.15) shows, identifying $i = 1$, that $d_{jj} \leq d_{jj}^{*}$, independent of whether $\beta_j = 0$. This means that $D = D^{*} - \tilde{D}$ for some nonnegative $\tilde{D}$. In order that this D does *not* result in $\dim_{\mathbb{C}} \ker(\Sigma - D) = m > 1$, it follows, as argued in the proof of (c), that $\tilde{D}$ can have only two zero entries, including the first, i.e., $d_{11} = d_{11}^{*}$ and $d_{jj} = d_{jj}^{*}$ for only one $j \neq 1$, i.e., only one β_j can be nonzero. Hence if $\alpha = 0$ and $L_{1,0}$ is nonempty, it could only consist of the points $x_2, x_3, \ldots, x_n$, obtained by setting $\alpha = 0$ and all but one $\beta_j = 0$ in the description $\alpha s_1 + (1-\alpha)\sum_{j=2}^{n}\beta_j x_j$, with $\sum_2^n \beta_j = 1$. We still need to verify though that the associated D ensures $\dim \ker(\Sigma - D) = 1$. This, however, follows by the following variant of the argument of (c). Recall that if $D = D^{*} - \tilde{D}$, with $\tilde{D} \geq 0$, $D \geq 0$, then

$$\ker(\Sigma - D) = \ker(\Sigma - D^{*}) \cap \ker \tilde{D}.$$

We have shown that $\ker(\Sigma - D^*)$ is spanned by $x_j, j = 2, \ldots, n$, and $\ker \tilde{D}$ is spanned by the unit vector e_1 and x_j, where j is such that $\tilde{d}_{jj} = 0$. Hence recalling that the x_i all have 1 in the first entry, and only one other nonzero entry (the ith), we see that $\ker(\Sigma - D)$ is spanned by x_j, and thus has dimension 1. This establishes the structure of $L_{1,0}$.

Finally, we consider the case $\alpha < 0$ and $L_{1,L}$. Here the proof is similar to that establishing the form of $L_{1,U}$. With $\alpha < 0$, it folows that (4.17) implies $d_{11} < d_{11}^*$ and (since $d_{11} \geq 0$ holds) $\alpha \geq \alpha_{\min} = -d_{11}^* \cdot (d_{11}^1 - d_{11}^*)^{-1}$. Using (4.13) and the fact established below (4.13) that $\tilde{D} = D^*$, it is evident that $\alpha_{\min} = -\alpha^*(1 - \alpha^*)^{-1}$. Since $d_{jj} < d_{jj}^*$ must hold with the possible exception of one value of j, call it $k \geq 2$, this implies, by (4.18), that $\beta_j \leq 0$ for $j \neq k$. However, then from $\sum_2^n \beta_j = 1$ we must have $\beta_k \geq 1$ for one $k \geq 2$; in this case $d_{kk} > d_{kk}^*$; thus these solutions correspond to $m = 1$. This argument shows that any vector in $L_{1,L}$ necessarily has the form indicated in (4.10). For the converse, let us a vector be given. Define $d_{11}, d_{22}, \ldots, d_{k-1,k-1}, d_{k+1,k+1}, \ldots, d_{nn}$ via (4.18), and notice that $d_{jj} < d_{jj}^*$ for all j, save $j = k$. It is not hard to check that $\Sigma - D$ is nonnegative by considering the kth elementary regression for $\Sigma - \operatorname{diag}(d_{11}, d_{22}, \ldots, d_{k-1,k-1}, 0, d_{k+1,k+1}, \ldots, d_{nn})$.

To obtain the alternative expression for $L_{1,L}$, observe using (4.4) and the properties of β_j that

$$\alpha s_1 + (1 - \alpha) \sum_{j=2}^{n} \beta_j x_j = [\alpha + (1 - \alpha)\alpha^*] s_1 + \{1 - [\alpha + (1 - \alpha)\alpha^*]\} \sum_{j=2}^{n} \beta_j s_j$$

$$= \alpha' s_1 + (1 - \alpha') \sum_{j=2}^{n} \beta_j s_j$$

with $\alpha' = \alpha + (1 - \alpha)\alpha^*$. When $\alpha \in [\alpha_{\min}, 0)$, then $\alpha' \in [0, \alpha^*)$ holds.

It remains to verify the set intersection properties just after (4.10). Suppose that $L_{1,U}$ and L_{n-1} have a nonempty intersection, in order to obtain a contradiction. Then

$$\alpha s_1 + (1 - \alpha) \sum_{j=2}^{n} \beta_j x_j = \sum_{j=2}^{n} \bar{\beta}_j x_j$$

for some $\alpha \in (0, 1]$, $\sum_2^n \beta_j = 1$, $\sum_2^n \bar{\beta}_j = 1$. By (4.4),

$$\left[\alpha + (1 - \alpha) \sum_{j=2}^{m} \beta_j \alpha^*\right] s_1 + (1 - \alpha) \sum_{j=2}^{n} \beta_j (1 - \alpha^*) s_j = \sum_{j=2}^{n} \bar{\beta}_j \alpha^* s_1 + \sum_{j=2}^{n} \bar{\beta}_j (1 - \alpha^*) s_j.$$

Since the s_i are linearly independent, this means that the coefficients of s_1 on each side are equal, i.e.,

$$\alpha + (1 - \alpha)\alpha^* = \alpha^*$$

or

$$\alpha(1 - \alpha^*) = 0.$$

Since $\alpha^* \neq 1$, it follows that $\alpha = 0$. This is a contradiction.

The intersection property for $L_{1,0}$ and L_{n-1} is trivial.

The intersection property for $L_{1,L}$ and L_{n-1} is proved just like that for $L_{1,U}$ and L_{n-1}. This completes the proof. ∎

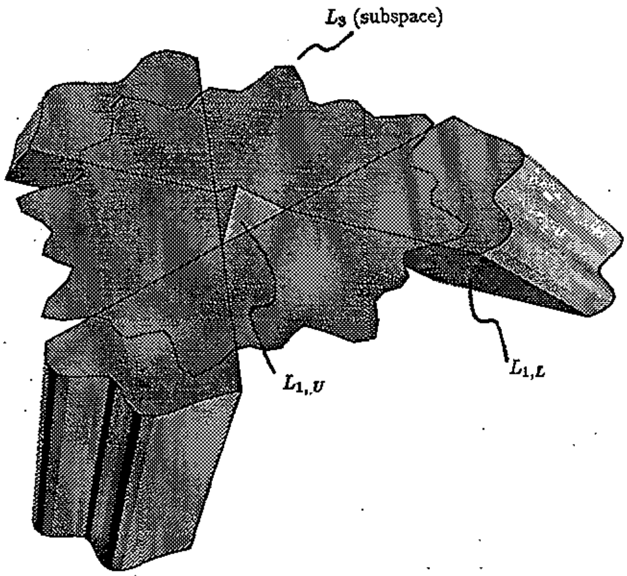

Fig. 1. A "typical" solution-set for $m^* = n - 1$: $n = 4$, $m^* = 3$, static (real) case.

In Fig. 1 we show a solution set L for $n = 4$, $m^* = 3$, in the static (real) case. Note that for the real static case $\dim_{\mathbb{R}} L_m$ is equal to $\dim_{\mathbb{C}} L_m$ for the complex case.

To help the reader understand the figure, let us note that:

(a) The four-cornered convex polytope on the top is $L_{1,U}$, the corners are $\pi(s_1)$, $\pi(x_j)$ with $j = 2, 3$, and 4, and the base of the polytope (corresponding to $\alpha = 0$) is excluded.

(b) The corners of the base of the $L_{1,U}$ polytope, namely $\pi(x_j)$, $j = 2$, 3, and 4, define $L_{1,0}$, which of course lies in L_3.

(c) L_3 is the plane through the base of this polytope, i.e., through the points $\pi(x_j)$, $j = 2$, 3, and 4.

(d) L_2 consists of the three (doubly infinite) lines, in the plane L_3, passing through each pair selected from $\pi(x_j)$, $j = 2$, 3, and 4.

(e) $L_{1,L}$ consists of the three three-dimensional disjoint pieces, all underneath the L_3 plane. These three pieces correspond to $\beta_2 \geq 1$, $\beta_3 \geq 1$, and $\beta_4 \geq 1$ (with otherwise $\beta_j \leq 0$). The top boundary of $L_{1,L}$ lies in the plane of L_3, and, as far as (4.10) is concerned, it obtained by setting $\alpha = 0$ in (4.10); this boundary is not part of $L_{1,L}$ itself. The bottom boundary of $L_{1,L}$, which is in $L_{1,L}$, is obtained by setting $\alpha = \alpha_{\min}$ in (4.10). (The restrictions on the signs of the β_j mean that not all the plane is in $L_{1,L}$.) Each of the three parts of $L_{1,L}$ is of infinite extent in an "outward" direction. The "sides" of each of the three $L_{1,L}$ parts are defined by planes. The part of $L_{1,L}$ in the lower right of the figure has side boundary planes defined by $\beta_3 = 0$ and $\beta_4 = 0$. Within this part of $L_{1,L}$, $\beta_2 \geq 1$, $\beta_3 \leq 0$, and $\beta_4 \leq 0$, and the closure of this part of $L_{1,L}$ interests $L_{1,U}$ at $\pi(x_2)$.

Remark 4.1. For $n = 2$, in general D^* is not unique (see [AD1]). Also for general n, if (vi) is violated, D^* may not be unique, as can be easily seen from the case $\sigma_{ij} = 0$, for all $i \neq j$ except for $(i, j) = (1, 2)$. On the other hand, for n sufficiently large in relation to m^*, D^* may be unique also for $m^* < n - 1$ (under additional conditions which are generically satisfied). For example, for $m^* = n - 2$ and $n \geq 5$ we can generically calculate the corresponding D^* from the condition that all 3×3 minors of $\hat{\Sigma}$ are zero.

Remark 4.2. As is easy to see, for $m^* = n - 1$, the solution sets L_m (in a certain sense) continuously depend on Σ.

5. Conclusions

Let us now review some of the results obtained above with respect to their use for identifying the underlying system. One of the main problems in EV model identification is to determine m^*. The results of Section 3 provide some help in this respect. If the conditions of Theorem 3.1 hold for every frequency λ, then we have $m^* = n - 1$ for the underlying dynamical system. It should be noted, however, that even if $m^* = n - 1$ holds, the usual *estimates* of Σ due to sampling variation are unlikely to satisfy the conditions of the theorem. Therefore in a real data situation, a likelihood-ratio or similar test must be used to determine m^*. In this respect the structure of the parameter space corresponding to the null hypothesis as described in Remark 3.3 is of interest. For given m^* then a restricted estimator for Σ (i.e., an estimator contained in Σ_{m^*}) can be developed.

The results of the previous sections are given for a fixed frequency. Of course they may be interpreted "pointwise" in frequency for a variable frequency situation. In this sense they relate in particular to nonparametric spectral estimation. There are a number of (still open) problems in connection with rational parametrization and parametric estimation in the EV case which have not been addressed in this paper. In this section we abandon the assumption of fixed frequency. For simplicity, we could think of rational spectra, which are defined pointwise in frequency.

For given $\hat{\Sigma}$, an underlying system can be determined as follows: First note that for given $\hat{\Sigma}$ the corresponding relation functions are unique up to left multiplication by nonsingular matrices. This is an immediate consequence of the fact that the rows of w form a basis for the left kernel of $\hat{\Sigma}$. Now select any m linearly independent columns of w. After a reordering putting the linearly independent columns in the first positions we write w as $(w_1 | w_2)$ where w_1 is nonsingular. Then $w_1^{-1}(w_1 | w_2) = (I, w_1^{-1} w_2)$ is an equivalent relation function, which gives a partition of $\hat{z}_t$ into inputs $\hat{x}_t$ and outputs $\hat{y}_t$, so that

$$\hat{y}_t = -w_1(z)^{-1} \cdot w_2(z) \hat{x}_t \tag{5.1}$$

holds. Clearly, in general, the partitioning of the elements of $\hat{z}_t$ into inputs and outputs will not be unique since the selection of m linearly independent columns from w is not unique in general.

For most cases the main interest is in the solutions corresponding to m^*; this corresponds to the maximum degree of explanation by the system versus the "outside world." In the case $m^* = n - 1$, where $\hat{\Sigma}$ and thus its kernel are uniquely determined from Σ, the relation function corresponding to $m^* = n - 1$ is unique up to left multiplication by nonsingular matrices.

One such relation function is

$$w = \left(\frac{x_2}{x_n}\right) \tag{5.2}$$

with x_j defined in (4.4). The systems corresponding to $m = n - 1$ have one input and $n - 1$ outputs. Moreover, by (vi) any one entry of $\hat{z}_t$ may be chosen as an input (which of course in general will not be causal for the outputs). Once the decision of which entry of $\hat{z}_t$ is chosen as the input has been made, for the case $m^* = n - 1$ and $m = n - 1$ we thus have a unique system for given Σ. Let, for instance, the first component of $\hat{z}_t$, $\hat{z}_{1,t} = \hat{x}_t$ say, denote the "true" input; then using the x_j in (4.4), $j = 2 \cdots n$, we have, for the "true" outputs,

$$\hat{z}_{j,t} = \hat{y}_{j-1,t} = k_{j-1}(z)\hat{x}_t, \qquad j = 2 \cdots m, \tag{5.3}$$

where $k_{j-1}(e^{-i\lambda}) = ((-x_j))_j$. As is easily seen for rational Σ, the transfer functions k_j are also rational.

Note that once a restricted estimate of Σ (satisfying the conditions of Theorem 3.1) is available then the results of Section 4 together with the results above allow us to construct directly the corresponding systems and in particular the (for given choice of inputs) unique system corresponding to $m = m^* = m - 1$. It is also easily seen that the constructions are continuous, so that a consistent estimator for Σ will give a consistent estimator for the underlying "true" system.

We want to emphasize again that we have not restricted ourselves to causal transfer functions k_j. It should be noted that in general the system corresponding to $m = m^*$ will not be unique (see, e.g., [DA2], where $m^* = 1$ is considered).

A different, but mathematically equivalent, formulation to the EV model (1.1)–(1.3) is given by the *factor model*, where $\hat{\Sigma}$ is represented as

$$\hat{\Sigma} = l \cdot l^*, \qquad l(z) = \sum_{j=0}^{\infty} L_j z^j, \qquad L_j \in \mathbb{R}^{n \times (n-m)}, \tag{5.4}$$

and $(\hat{z}_t)$ is represented as

$$\hat{z}_t = l(z) \cdot f_t. \tag{5.5}$$

The matrix $l(z)$ is called the factor loading matrix and (f_t) is the factor process which here is assumed to be white noise with $Ef_t f_t' = 2\pi I$. From a formal point of view the only difference to the EV model is that the restrictions on $(\hat{z}_t)$ are expressed by the range of a matrix, namely $l(z)$, rather than by the kernel of a matrix, namely $w(z)$. However, the interpretation is different since in the factor model the main interest is in the dynamic system represented by the factor loading matrix $l(z)$ and its unobserved inputs, the factors f_t.

Clearly, $m = n - 1$ corresponds to the single-factor case, where f_t is one dimensional.

References

[AD1] B. D. O. Anderson and M. Deistler, Identifiability in dynamic errors-in-variables models, *J. Time Ser. Anal.*, 5 (1984), 1–13.

[AD2] B. D. O. Anderson and M. Deistler, Dynamic errors-in-variables systems with three variables, *Automatica*, **23** (1987), 611–616.

[AR] T. W. Anderson and H. Rubin, Statistical inference in factor analysis, in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (J. Neyman, ed.), pp. 11–150, University of California Press, Berkeley, CA, 1956.

[BL] P. Bekker and J. de Leeuw, The rank of reduced dispersion matrices, *Psychometrika*, **52** (1987), 125–135.

[D] M. Deistler, Symmetric modelling in system identification, in *Three Decades of Mathematical System Theory* (H. Nijmeijer and J. M. Schumacher, eds.), pp. 129–147, Lecture Notes in Control and Information Sciences, Vol. 135, Springer-Verlag, Berlin, 1989.

[DA1] M. Deistler and B. D. O. Anderson, Linear dynamic errors-in-variables models, some structure theory, *J. Econometrics*, **41** (1989), 39–63.

[DA2] M. Deistler and B. D.O. Anderson, Identification of dynamic systems from noisy data: the case $m^* = 1$, in *Mathematical System Theory: The Influence of R. E. Kalman* (A. C. Antoulas, ed.), pp. 423–435, Springer-Verlag, Berlin, 1991.

[EW] R. Engle and M. Watson, A one-factor multivariate time series model of metropolitan wage rates, *J. Amer. Statist. Assoc.*, **76** (1981), 774–781.

[G] J. F. Geweke, The dynamic factor analysis of economic time series, *Latent Variables in Socioeconomic Models* (D. Aigner and A. Goldberger, eds.), pp. 365–383. North-Holland, Amsterdam, 1977.

[H] S. Haykin, ed., *Array Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1985.

[K] R. E. Kalman, System identification from noisy data, in *Dynamic Systems II, a University of Florida International Symposium* (A. Bednarek and L. Cesari, eds.), pp. 135–164, Academic Press, New York, 1982.

[S] C. Spearman, General intelligence, objectively determined and measured, *Amer. J. Psychology*, **15** (1904), 201–293.

[W] J. C. Willems, From time series to linear systems, Part 1, *Automatica*, **22** (1986), 561–580.