Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

J.H. van Schuppen

Stochastic realization problems

Department of Operations Research, Statistics, and System Theory Report BS-R8920 October

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Research (N.W.O.).

Stochastic Realization Problems

J.H. van Schuppen
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

The stochastic realization problem asks for the existence and the classification of all stochastic systems for which the output process equals a given process in distribution or almost surely. This is a fundamental problem of system and control theory. The stochastic realization problem is of importance to modelling by stochastic systems in engineering, biology, economics etc. Several stochastic systems are mentioned for which the solution of the stochastic realization problem may be useful. As an example recent research on the stochastic realization problem for the Gaussian factor model and a Gaussian factor system is discussed.

This paper is dedicated to J.C. Willems on the occasion of his fiftieth birthday.

1980 Mathematics Subject Classification (1985 Revision): 93B15, 93E03.
Key Words and Phrases: system theory, stochastic realization, Gaussian factor analysis, Gaussian factor system, errors-in-variables model.

1. Introduction

The purpose of this paper is to introduce the reader to stochastic realization theory. This will be done by presentation of a verbal introduction, a survey of Gaussian stochastic realization theory, formulation of open stochastic realization problems, and a discussion of the stochastic realization problem for Gaussian factor models. This tutorial and survey-like paper is written for researchers in system and control theory, but may also be of interest to researchers dealing with mathematical models in engineering, biology, and economics.

The Kalman filter and stochastic control algorithms have proven to be very useful for those control and signal processing problems in which there is a considerable amount of noise in the observation processes. Examples of such problems are: minimum variance control of a paper machine, access control of communication systems, and prediction of water levels. The solution of stochastic control and filtering problems depends crucially on the availability of a model in the form of a stochastic system in state space form. There is thus a need for modelling and realization of noisy processes by stochastic systems. Stochastic realization theory addresses this modelling problem.

System and control theory is the subject within engineering and mathematics that deals with modelling and control problems for dynamic processes or phenomena. Such a phenomenon may initially be described by specifying the observation process or trajectories, which description will be termed the *external description.* For reasons of modelling and control it is often better to work with an *internal description.* The form of such an internal description depends on the properties of the observation process. For deterministic linear systems it may be a description in state space form. The state of such a system at any particular time contains all information from the past necessary to determine the future behavior of the state and output process. For stochastic systems the internal description is a stochastic system in state space form. Here the state is that amount of information that makes the past and the future of the observations and the state process conditionally independent. For a vector valued random variable one may consider the internal description of a Gaussian factor model, see section 5. For models of images and spatial phenomena in the form of random fields, other internal descriptions are

Report BS-R8920
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

needed.

The *realization problem* of system theory can then be formulated as how to determine an internal description of a model given an external description. Motivation for this problem comes from engineering, in particular from system identification and signal processing, from biology, and from econometrics. In these subject areas one may want to estimate parameters of the internal description from observations. The question should then be posed whether these parameters can be uniquely determined from the observations, that is whether they are identifiable. This question may be resolved by solution of the realization problem. First one must impose the condition that the model is minimal in some sense. The concept of minimality will depend on the class of internal descriptions. Secondly, there is in general no unique internal description for a phenomenon given an external description. The realization problem therefore also asks for a classification of all minimal internal descriptions that correspond to a given external description. Such internal descriptions may be called equivalent. Once the equivalence class has been determined one may choose a canonical form for it. From that point on standard techniques from system identification and statistics may be used to determine the internal description of the model. The part of system theory that deals with modelling questions is referred to as *realization theory*. It treats topics such as transformations between representations, parametrization of model classes, identifiability questions, and approximate modelling.

A brief description of this paper's content follows. Section 2 contains a verbal introduction to the modelling procedure of system theory. In section 3 a tutorial is presented on Gaussian stochastic realization theory. Several examples of stochastic systems for which the stochastic realization problem is open and relevant for engineering and economics, are mentioned in section 4. As an example the stochastic realization problem for the Gaussian factor analysis model is discussed in section 5, and for Gaussian factor systems or error-in-variables systems in section 6.

2. Modelling and system theory

2.1. Introduction

As identified in the previous section there is a need for stochastic models of engineering and economic phenomena. The purpose of this section is to describe the modelling procedure of system and control theory. Particular attention will be devoted to modelling of economic processes.

2.2. The modelling procedure

It is assumed that data, possibly in the form of time series, are available for the modeller. It is well-recognized that useful data are easy to obtain in the technical sciences but hard to obtain in economics. One reason is that economics is not a laboratory science; experiments are often impossible or if possible cannot be repeated. Also data gathering is much more expensive in economics than in the technical sciences.

The objective of modelling is to obtain a model for a phenomenon that is realistic and of low complexity. A model is called *realistic* if its observed behaviour is in close agreement with the phenomenon. A measure of fit for this agreement has to be formulated. The term *low complexity* should be considered as in ordinary use. A mathematical definition of this term is very much model dependent. Models of high complexity are mathematically not well analyzable and computationally not feasible. The two modelling objectives mentioned are conflicting. Therefore a compromise or trade-off between these objectives is necessary.

The preferred modelling procedure consists of the following two steps:

- selection of a model class;
- selection of an element in the model class involving the above mentioned trade-off.

This procedure must be applied in an iterative fashion. If the selected element in the model class is not a realistic model then the model class may be adjusted. The two steps of this procedure will now be discussed separately.

2.3. Selection of a model class

In the selection of a model class one has to keep in mind the objectives of a realistic model and a model of low complexity. The selection procedure demands application of concepts and results both from the research area of the object to be modelled, and from system and control theory.

The formulation of realistic economic models is difficult for several reasons. One reason is that economic transactions involve multiple decisionmakers compared with a single decisionmaker in most engineering problems. The appropriate mathematical models are therefore game and team models and their dynamic counterparts. The status of dynamic game and team theory is not yet at a level at which a body of results is available for applications. A second reason, closely related to the first, is that a decisionmaker must also model the decisionmaking process of the other decisionmakers. This remark is well-known in the literature on stochastic dynamic games. The discussion about rational expectation also illustrates this point. A third reason is that the rules of the economic process change quickly compared with the periods over which economic data are available. Assumptions of time-invariance or stationarity are often unrealistic.

In system theory a formalism has been developed for the formulation of mathematical models of dynamic phenomena and for a modelling procedure. For a dynamic phenomenon in the form of a time series a preferred deterministic model is called a *dynamic system* in state space form. One distinguishes *inputs* and *outputs* of such a system, and a state process. The *state* of a dynamic system at any particular time is that amount of information that together with the future inputs completely determines the future outputs. The trajectories of the input, output and state process are the basic objects of a dynamical system. The reader is referred to [78] for material on linear systems.

Stochastic systems have proven to be useful models in several areas of engineering such as signal processing, communications and control. Within economics they are used for example in connection with portfolio theory. In stochastic system theory, probability theory is used as a mathematical model for uncertainty. A stochastic system is specified by a measure on the space of trajectories. This is a fundamental difference between deterministic and stochastic systems. For a stochastic system without inputs the state at any particular time makes the past and the future of the output and state processes conditionally independent. Despite the fact that a stochastic system is specified by a measure, the representation in terms of trajectories, for example by a stochastic differential equation, is crucial to the solution of control and filtering problems.

Why are stochastic models realistic in certain cases? Within economics reasons for this are that such modelling involves:

- aggregation over many decisionmakers;
- uncertainty over future actions of other decisionmakers;
- uncertainty in the measurement process, due to vague definitions and averaging.

Remark that the costs involved often prevent the gathering of full information. Therefore aggregation must be used. The variability of the data then suggests a stochastic model. This author is not optimistic about the applicability of stochastic models to economic phenomena. Reasons for this are the relatively short time series and the frequent change in structural relations.

Should one use a deterministic or a stochastic model class to model a certain phenomenon? What is needed is a criterion to decide whether for a specific phenomenon the class of deterministic systems or that of stochastic systems is the appropriate model class.

A crucial observation from system theory is that the *choice of model class* is all-important. Of course, a model must be realistic and of low complexity. But within these constraints there is left some freedom in the mathematical formulation of the model. Given this freedom it is advisable to choose a model class for which the motivating control problem is analytically tractable. An example of such a choice is the Gaussian system that leads to the Kalman filter. Filtering theory was formulated by N. Wiener and A.N. Kolmogorov for stationary Gaussian processes. R.E. Kalman restricted attention to a particular class of stationary Gaussian processes, those generated by linear stochastic systems driven by white noise. For this class of systems the solution of the filtering problem has proven to be straightforward. That this class may be extended to include non-stationary processes is then a useful corollary. How is this observation to be used in economic modelling? As suggested by R.E. Kalman, a detailed study must be made of economic models that are published in the literature to see whether changes in the mathematical formulation of these models are advantageous for the solution of control problems. The selection of the model class seems a creative process that involves knowledge of both the research area of the phenomenon to be modelled and of system theory.

For stochastic processes indexed by the real line the model class of stochastic systems seems an appropriate model. See section 3.1 for a definition of this concept. For a vector of random variables the model class of Gaussian factor models may be useful, see section 5. For random fields it is not yet clear what the appropriate model class should be.

Once the model class has been determined, the modelling procedure prescribes the solution of the stochastic realization problem. In section 3 this problem is formulated and the solution shown for the case of Gaussian processes.

2.4. Selection of an element in the model class

Given the data and the model class, the problem arises of how to select an element in the model class. As indicated earlier, the selection of a model is a trade-off between the objective of a realistic model and the objective of a model with low complexity. For deterministic dynamical systems results on the selection of an element in the model class are reported in [35, 79].

For stochastic systems a formalism for the selection of an element in the class of stochastic systems is described below. Consider first a measure of fit between the observations of the phenomenon and the external behaviour of a stochastic system. Recall that the observations consist of numbers while the external behaviour consists of a measure on the sample space of observation trajectories. The way to proceed is to use the observations, the numbers, to estimate the measure on the sample space of observation trajectories. In case this measure is Gaussian and the observation process is stationary it suffices to estimate the mean and covariance function of this measure.

One can define a measure of fit between the measure for the output trajectories estimated from observations and the measure associated with the external description of the system. Examples of such a measure are the Kullback-Leibler measure and the Hellinger measure; see section 3.7.

For stochastic systems one also needs a measure of complexity. A stochastic complexity measure introduced by J. Rissanen [60-64] seems the appropriate tool for this purpose. Stochastic complexity is based on A.N. Kolmogorov's complexity theory. Since this subject is well covered elsewhere the reader is referred to the indicated references.

The actual selection procedure given data, a model class, and measures of fit and complexity, consists then of a combination of analysis and numerical minimization. The details of this will not be discussed here.

3. Gaussian stochastic realization

The purpose of this section is to present the modelling procedure for Gaussian processes. In this tutorial part of the paper results for the Gaussian stochastic realization problem are summarized. For a reference on the weak Gaussian stochastic realization problem see the book [24] and for a shorter introduction in the English language [23]. For a survey of the strong Gaussian stochastic realization problem see [47].

Notation

The following notation is used. $\mathbb{N}=\{0,1,2,\cdots\}$. $\mathbb{Z}_+=\{1,2,\cdots\}$. $\mathbb{Z}=\{\cdots,-1,0,1,\cdots\}$. $\mathbb{Z}_k=\{1,2,\cdots,k\}$. $\mathbb{R}$ denotes the set of real numbers, and $\mathbb{R}_+=[0,\infty)$. For a probability space (Ω,F,P) consisting of a set Ω, a σ-algebra F and a probability measure P, denote

$$L^+(F) = \{x:\Omega\rightarrow\mathbb{R}_+ \mid x \text{ is a random variable measurable with respect to } F\}.$$

$x\in G(0,Q)$ denotes that the random variable x has a Gaussian distribution with mean zero and variance Q.

For a stochastic process $y:\Omega\times T\rightarrow\mathbf{R}^k$ the following notation is used for the σ-algebra's generated by the process $F_t^y=F_t^{y-}=\sigma(\{y(s),\forall s\leq t\})$ and $F_t^{y+}=\sigma(\{y(s),\forall s\geq t\})$.

DEFINITION 3.0.1. *The σ-algebra's F_1,F_2 are called* conditionally independent *given the σ-algebra G if*

$$E[z_1z_2\,|\,G]=E[z_1\,|\,G]E[z_2\,|\,G]$$

for all $z_i\in L^+(F_i)$. The notation

$$(F_1,F_2\,|\,G)\in CI$$

will be used to denote that F_1,F_2 are conditionally independent given G and CI will be called the conditional independence relation.

3.1. Stochastic systems and Gaussian systems

The purpose of this section is to define stochastic dynamic systems. Attention is restricted to discrete-time stochastic dynamic systems. Stochastic systems with inputs will not be considered here.

A motivation for the definition of a discrete time stochastic dynamic system follows. Consider the object that is usually called a stochastic system,

$$x_{t+1}=Ax_t+Mv_t,\ x_0, \tag{3.1.1}$$

$$y_t=Cx_t+Nv_t, \tag{3.1.2}$$

where $x_0:\Omega\rightarrow\mathbf{R}^n, x_0\in G(m_0,Q_0)$, $v:\Omega\times T\rightarrow\mathbf{R}^m$ is a Gaussian white noise process with $v_t\in G(0,V)$, F^{x_0},F^v_∞ are independent σ-algebras, $A\in\mathbf{R}^{n\times n}$, $M\in\mathbf{R}^{n\times m}$, $C\in\mathbf{R}^{p\times n}$, $N\in\mathbf{R}^{p\times m}$, $x:\Omega\times T\rightarrow\mathbf{R}^n$ and $y:\Omega\times T\rightarrow\mathbf{R}^p$ defined by the above equations. It may be shown that this object is equivalent with the object specified by:

$$x_0\in G(m_0,Q_0); \tag{3.1.3}$$

$$E[\exp(iu^Tx_{t+1}+iw^Ty_t)\,|\,F_t^{x-}\vee F_{t-1}^{y-}]=\exp\left(i\begin{bmatrix}u\\w\end{bmatrix}^T\begin{bmatrix}Ax_t\\Cx_t\end{bmatrix}-\tfrac{1}{2}\begin{bmatrix}u\\w\end{bmatrix}^TS\begin{bmatrix}u\\w\end{bmatrix}\right), \tag{3.1.4}$$

for all $t\in T$ and some $S\in\mathbf{R}^{(n+p)\times(n+p)}$. Observe that the conditional characteristic function of (x_{t+1},y_t) given $(F_t^{x-}\vee F_{t-1}^{y-})$ depends only on the random variable x_t. It then follows that

$$E[\exp(iu^Tx_{t+1}+iw^Ty_t)\,|\,F_t^{x-}\vee F_{t-1}^{y-}]=E[\exp(iu^Tx_{t+1}+iw^Ty_t)\,|\,F^{x_t}] \tag{3.1.5}$$

for all $t\in T$. A stochastic dynamic system could now be defined as a state process x and an output process y such that for all $t\in T$ there is a map

$$x_t\mapsto \textit{distribution of}\,(x_{t+1},y_t)$$

This definition may be found in [42; p. 5]. Below a different definition will be adopted. It may be shown that (3.1.5) is equivalent with the condition that for all $t\in T$

$$(F_t^{y+}\vee F_t^{x+},F_{t-1}^{y-}\vee F_t^{x-}\,|\,F^{x_t})\in CI,$$

where $F_t^{x+}=\sigma(\{x_s,\forall s\geq t\})$, $F_t^{x-}=\sigma(\{x_s,\forall s\leq t\})$, and similar definitions for F_t^{y+},F_t^{y-}. The property that the past and future of the state and output process are conditionally independent given the current state will be taken as the definition of a stochastic dynamic system.

DEFINITION 3.1.1. *A discrete-time* stochastic dynamic system *is a collection*

$$\sigma=\{\Omega,F,P,T,Y,B_Y,X,B_X,y,x\},$$

where

$\{\Omega,F,P\}$ *is a complete probability space;*

$T=\mathbb{Z}$, to be called the time index set;
(Y,B_Y) is a measurable space, to be called the output space;
(X,B_X) is a measurable space, to be called the state space;
$y:\Omega\times T\rightarrow Y$ is a stochastic process, to be called the output process;
$x:\Omega\times T\rightarrow X$ is a stochastic process, to be called the state process;
such that for all $t\in T$

$$(F_t^{y+}\vee F_t^{x+},F_{t-1}^{y-}\vee F_t^{x-}\mid F^{x_t})\in CI. \tag{3.1.6}$$

A stochastic dynamic system on $T\subset\mathbb{Z}$ is defined analogously. The class of stochastic systems is denoted by $S\Sigma$.

The above definition of a stochastic dynamic system is based on related concepts given in [48, 52, 72, 73].

From the definition of a stochastic dynamic system one obtains that the state process satisfies the condition

$$(F_t^{x+},F_t^{x-}\mid F^{x_t})\in CI$$

for all $t\in T$. This is equivalent with x being a Markov process. Markov processes are thus also stochastic dynamic systems, and the latter class thus contains the classical model of state processes.

The defining condition of a stochastic dynamic system is more or less symmetric with respect to time in the past and future of the state and output process. This is an advantage over the asymmetric formulation given in the representation (3.1.1) and (3.1.2).

The condition (3.1.6) is asymmetric with respect to the output process. This is a convention. A priori there are four possible conditions for a stochastic dynamic system which are listed below:

$$(F_t^{y+}\vee F_t^{x+},F_t^{y-}\vee F_t^{x-}\mid F^{x_t})\in CI\ \ \forall\, t\in T; \tag{3.1.7.1}$$

$$(F_{t+1}^{y+}\vee F_t^{x+},F_{t-1}^{y-}\vee F_t^{x-}\mid F^{x_t})\in CI\ \ \forall\, t\in T; \tag{3.1.7.2}$$

$$(F_t^{y+}\vee F_t^{x+},F_{t-1}^{y-}\vee F_t^{x-}\mid F^{x_t})\in CI\ \ \forall\, t\in T; \tag{3.1.7.3}$$

$$(F_{t+1}^{y+}\vee F_t^{x+},F_t^{y-}\vee F_t^{x-}\mid F^{x_t})\in CI\ \ \forall\, t\in T. \tag{3.1.7.4}$$

Condition (3.1.7.1) and a property of conditional expectation imply that

$$F^{y_t}\subset(F_t^{y+}\vee F_t^{y-})\subset F^{x_t}$$

which fact is not compatible with the intuitive concept of state in that the output is in general not part of the state. Condition (3.1.7.2) is not suitable because it would allow examples that are counter-intuitive to the concept of state, see example 3.1.6. The conditions (3.1.7.3) and (3.1.7.4) thus remain, of which condition 3 has been chosen. This is a convention. Condition (3.1.7.4) results in the representation

$$x_{t+1} = Ax_t + Mv_t,$$

$$y_{t+1} = Cx_t + Nv_t,$$

which form is inconsistent with the system theoretic convention of (3.1.1 & 3.1.2). The option of taking condition (3.1.7.3) or (3.1.7.4) in the definition of a stochastic dynamic system is related to the option of considering Moore or Mealey machines in automata theory, see [50; I.A.2].

The definition of a stochastic system is formulated in terms of σ-algebras rather than in terms of stochastic processes. This is a geometric formulation in which emphasis is put on spaces and subspaces rather than on the variables or processes that generate those spaces.

DEFINITION 3.1.2. *Given a stochastic dynamic system*

$$\sigma = \{\Omega, F, P, T, Y, B_Y, X, B_X, y, x\} \in S\Sigma.$$

This system is called:

a. stationary *or* time-invariant *if* (x,y) *is a jointly stationary process;*

b. Gaussian *if* $Y = \mathbf{R}^p$, $X = \mathbf{R}^n$ *for certain* $p, n \in \mathbf{Z}_+$, $B_Y = B^p$ *and* $B_X = B^n$ *are Borel* σ*-algebras on* Y *respectively* X, *and if* (x,y) *is a jointly Gaussian process; by way of abbreviation, a Gaussian stochastic dynamic system will be called a* Gaussian system *and the class of such systems is denoted by* $GS\Sigma$;

c. finite *if* Y, X *are finite sets and* B_Y, B_X *are the* σ*-algebras on* Y, X *generated by all subsets; by way of abbreviation a finite stochastic dynamic system will be called a* finite stochastic system *and the class of such systems is denoted by* $FS\Sigma$.

PROPOSITION 3.1.3. *Consider a collection*

$$\{\Omega, F, P, T, Y, B_Y, X, B_X, y, x\}$$

as defined in 3.1.1 but without condition (3.1.6). The following statements are equivalent:

a. *for all* $t \in T$

$$(F_t^{y+} \vee F_t^{x+}, F_{t-1}^{y-} \vee F_t^{x-} \mid F^{x_t}) \in CI;$$

b. *for all* $t \in T$

$$(F^{y_t} \vee F^{x_{t+1}}, F_{t-1}^{y-} \vee F_t^{x-} \mid F^{x_t}) \in CI;$$

c. *for all* $t \in T$

$$(F_t^{y+} \vee F_t^{x+}, F^{y_{t-1}} \vee F^{x_{t-1}} \mid F^{x_t}) \in CI.$$

The following result is a useful sufficient condition for a stochastic dynamic system.

PROPOSITION 3.1.4. *Consider the collection*

$$\sigma = \{\Omega, F, P, T, Y, B_Y, X, B_X, y, x\}$$

as defined in 3.1.1 but without condition (3.1.6). If for all $t \in T$

1. $(F_t^{y+}, F_\infty^{x-} \vee F_{t-1}^{y-} \mid F^{x_t}) \in CI$;

2. $(F_t^{x+}, F_t^{x-} \vee F_{t-1}^{y-} \mid F^{x_t}) \in CI$;

then $\sigma \in S\Sigma$.

Below two examples of stochastic dynamic systems are presented.

EXAMPLE 3.1.5. Consider a Gaussian system representation

$$x_{t+1} = Ax_t + Mv_t, \tag{3.1.8}$$

$$y_t = Cx_t + Nv_t, \tag{3.1.9}$$

with the conventions given below (3.1.1 & 3.1.2). As indicated there this representation is equivalent with

$$E[\exp(iu^T x_{t+1} + iw^T y_t) \mid F_t^{x-} \vee F_{t-1}^{y-}]$$
$$= \exp(i \begin{bmatrix} u \\ w \end{bmatrix}^T \begin{bmatrix} Ax_t \\ Cx_t \end{bmatrix} - ½ \begin{bmatrix} u \\ w \end{bmatrix}^T S \begin{bmatrix} u \\ w \end{bmatrix}),$$

for all $t \in T$ and $x_0 \in G$. This and a property of conditional independence imply that

$$(F^{x_{t+1}} \vee F^{y_t}, F^{y-}_{t-1} \vee F^{x-}_t \mid F^{x_t}) \in CI, \ \forall\, t \in T,$$

and from 3.1.3 then follows that, with x,y specified by (3.1.8 & 3.1.9),

$$\sigma = \{\Omega, F, P, T, \mathbf{R}^p, B^p, \mathbf{R}^n, B^n, y, x\} \in S\Sigma.$$

From properties of Gaussian random variables follows that (x,y) is a jointly Gaussian process, hence σ is a Gaussian system or $\sigma \in GS\Sigma$. In the following (3.1.8 & 3.1.9) will be called a forward representation of a Gaussian system.

EXAMPLE 3.1.6. Let $v:\Omega \times T \to \mathbf{R}$ be a standard Gaussian white noise process. Define $y:\Omega \times T \to \mathbf{R}$, $x:\Omega \times T \to \mathbf{R}$ by

$$x_t = v_{t-1}, \ y_t = x_t + v_t = v_{t-1} + v_t.$$

Then the following hold.

a. For all $t \in T$ $(F^{y+}_{t+1}, F^{y+}_{t-1} \mid N) \in CI$, where $N \subset F$ is the trivial σ-algebra. Thus the process y is the output process of a stochastic dynamic system according to (3.1.7.2) with a trivial state space.

b. For all $t \in T$

$$E[\exp(iuy_t) \mid F^{y+}_{t-1}]$$

is nondeterministic, indicating that the process y has some kind of memory.

c.

$$(F^{y+}_t \vee F^{x+}_t, F^{y-}_{t-1} \vee F^{x-}_t \mid F^{x_t}) \in CI$$

for all $t \in T$, hence

$$\sigma = \{\Omega, F, P, T, Y, B, X, B, y, x\} \in GS\Sigma.$$

3.2. Forward and backward representations of Gaussian systems

The purpose of this subsection is to show that a Gaussian system has both a forward and a backward representation, and to derive relations between these representations.

PROPOSTION 3.2.1. *Let*

$$\sigma = \{\Omega, F, P, T, \mathbf{R}^p, B^p, \mathbf{R}^n, B^n, y, x\} \in GS\Sigma$$

be a Gaussian system. Assume that for all $t \in T$ $E[x_t]=0$, $E[y_t]=0$ *and that* $Q:T \to \mathbf{R}^{n \times n}, Q(t)=E[x_t x_t^T]>0$.

a. *The Gaussian system has what will be called a* forward representation *given by*

$$x_{t+1} = A^f(t)x_t + Mv^f_t, \ x_0, \tag{3.2.1}$$

$$y_t = C^f(t)x_t + Nv^f_t, \tag{3.2.2}$$

where $v^f:\Omega \times T \to \mathbf{R}^{n+k}$ *is a Gaussian white noise process with intensity* V^f. *Given* σ *then*

$$A^f(t) = E[x_{t+1}x_t^T]Q(t)^{-1},$$

$$C^f(t) = E[y_t x_t^T]Q(t)^{-1},$$

$$V^f(t) = \begin{bmatrix} Q(t+1) & E[x_{t+1}y_t^T] \\ E[y_t x_{t+1}^T] & E[y_t y_t^T] \end{bmatrix} - \begin{bmatrix} A^f(t) \\ C^f(t) \end{bmatrix} Q(t)^{-1} \begin{bmatrix} (A^f(t))^T & (C^f(t))^T \end{bmatrix},$$

$$M = (I_n \ 0) \in \mathbf{R}^{n \times (n+p)}, \ N = (0 \ I_p) \in \mathbf{R}^{p \times (n+p)}.$$

Conversely, given a forward representation with A^f, C^f, V^f, M, N *functions and* x,y *defined by the above* *forward representation (3.2.1 & 3.2.2), then σ is a Gaussian system.*

b. *The given Gaussian system has also a* backward representation *given by*

$$x_{t-1} = A^b(t)x_t + Mv_t^b,\ x_0, \tag{3.2.3}$$

$$y_{t-1} = C^b(t)x_t + Nv_t^b, \tag{3.2.4}$$

where $v^b:\Omega\times T\to\mathbb{R}^{n+k}$ is a Gaussian white noise process with intensity V^b. Given σ

$$A^b(t) = E[x_{t-1}x_t^T]Q(t)^{-1}, \tag{3.2.5}$$

$$C^b(t) = E[y_{t-1}x_t^T]Q(t)^{-1}, \tag{3.2.6}$$

$$V^b(t) = \begin{bmatrix} Q(t-1) & E[x_{t-1}y_t^T] \\ E[y_t x_{t-1}^T] & E[y_t y_t^T] \end{bmatrix} - \begin{bmatrix} A^b(t) \\ C^b(t) \end{bmatrix} Q(t)^{-1} \begin{bmatrix} (A^b(t))^T & (C^b(t))^T \end{bmatrix}, \tag{3.2.7}$$

$$M = (I_n\ 0),\ N = (0\ I_p).$$

Conversely, given a backward representation with A^b, C^b, V^b, M, N and x, y as defined by the above backward representation, then σ is a Gaussian system.

c. *The relation between the forward and backward representation of a Gaussian system is given by*

$$A^f(t)Q(t) = Q(t+1)(A^b(t+1))^T, \tag{3.2.8}$$

$$C^b(t)Q(t) = C^f(t-1)Q(t-1)(A^f(t-1))^T + NV^f(t-1)M^T, \tag{3.2.9}$$

$$C^f(t)Q(t) = C^b(t+1)Q(t+1)(A^b(t+1))^T + NV^b(t+1)M^T. \tag{3.2.10}$$

d. *Assume that the given Gaussian system is stationary. Then $A^f, C^f, V^f, A^b, C^b, V^b$, do not depend explicitly on $t\in T$ and $Q(t)=Q\in\mathbb{R}^{n\times n}$, $Q=Q^T>0$. The relation between the forward and backward representation is then given by*

$$A^f = Q(A_b)^T Q^{-1}, \tag{3.2.11}$$

$$A^b = Q(A^f)^T Q^{-1}, \tag{3.2.12}$$

$$C^b = C^f Q(A^f)^T Q^{-1} + NV^f M^T Q^{-1} = C^f A^b + NV^f M^T Q^{-1}, \tag{3.2.13}$$

$$C^f = C^b Q(A^b)^T Q^{-1} + NV^b M^T Q^{-1} = C^b A^f + NV^b M^T Q^{-1}. \tag{3.2.14}$$

In the following the superscripts f and b will be omitted when it is clear from the context which representation is referred to.

3.3. Stochastic observability and stochastic reconstructibility

The theorem on the characterization of minimality of a stochastic realization makes use of the concepts of stochastic observability and stochastic reconstructibility. Below these concepts are introduced.

DEFINITION 3.3.1. Consider a stochastic system

$$\sigma = \{\Omega, F, P, T, \mathbb{R}^p, B^p, \mathbb{R}^n, B^n, y, x\} \in S\Sigma.$$

a. This system is called *stochastically observable on the interval* $\{t, t+1, \ldots, t+t_1\}$ if the map

$$x(t) \mapsto E[\exp(i\sum_{s=0}^{t_1} u(s)^T y(t+s)) \mid F^{x(t)}]$$

from $x(t)$ to the conditional characteristic function of $\{y(t), y(t+1), \ldots, y(t+t_1)\}$ given $x(t)$ is injective on the support of $x(t)$.

b. Assume that the system σ is stationary. Then it is called *stochastically observable* if there exists a t, $t_1\in T$, $0<t_1<\infty$, such that it is stochastically observable on the interval $\{t, t+1, \ldots, t+t_1\}$ as defined above. By stationarity this then holds for all $t \in T$.

The interpretation of a stochastically observable stochastic system is that if one knows the conditional distribution of $\{y(t), y(t+1), \ldots, y(t+t_1)\}$ given $x(t)$, then one can uniquely determine the value of $x(t)$. Note that the conditional distribution of $\{y(t), \ldots, y(t+t_1)\}$ given $x(t)$ can in principle be determined from measurements.

THEOREM 3.3.2. *Consider the Gaussian system*

$$\sigma = \{\Omega, F, P, T, \mathbb{R}^p, B^p, \mathbb{R}^n, B^n, y, x\} \in GS\Sigma,$$

with forward representation

$$x(t+1) = A^f(t)x(t) + Mv(t),$$
$$y(t) = C^f(t)x(t) + Nv(t),$$

with $v(t) \in G(0, V(t))$.

a. *The system* σ *is stochastically observable on the interval* $\{t, t+1, \ldots, t+t_1\}$ *iff*

$$ker(S^f(t)Q(t)) = ker(Q(t)), \qquad (3.3.1)$$

iff $F^{x(t)} = F^{S^f(t)x(t)}$, *where* $x(t) \in G(0, Q(t))$ *and*

$$S^f(t) = \begin{bmatrix} C^f(t) \\ C^f(t+1)\Phi^f(t+1,t) \\ \cdots \\ C^f(t+t_1)\Phi^f(t+t_1,t) \end{bmatrix}.$$

b. *Assume that the system is stationary with forward representation*

$$x(t+1) = A^f x(t) + Mv(t),$$
$$y(t) = C^f x(t) + Nv(t),$$

with $v(t) \in G(0, V)$. *This implies that the matrix* A^f *is exponentially stable. Let* $x(t) \in G(0, Q)$ *in which* $Q \in \mathbb{R}^{n \times n}$ *is the solution of the Lyapunov equation*

$$Q = A^f Q (A^f)^T + MVM^T.$$

Then this system is stochastically observable iff

$$ker(S^f Q) = ker(Q), \qquad (3.3.2)$$

iff $F^{x(t)} = F^{S^f x(t)}$ *for some* $t \in T$, *where*

$$S^f = \begin{bmatrix} C^f \\ C^f A^f \\ \cdots \\ C^f (A^f)^{n-1} \end{bmatrix}.$$

c. *Consider the assumptions of b. Let* $G \in \mathbb{R}^{n \times n}$ *be such that* $GG^T = MVM^T$. *Assume that* (A^f, G) *is a reachable pair. Then the system is stochastically observable iff* (A^f, C^f) *is an observable pair.*

PROOF. a. Let $t, t_1 \in T$, $t > 0$,

$$\bar{y} = \begin{bmatrix} y_t \\ y_{t+1} \\ \cdots \\ y_{t+t_1} \end{bmatrix}, \quad u = \begin{bmatrix} u_t \\ u_{t+1} \\ \cdots \\ u_{t+t_1} \end{bmatrix}.$$

Then

$$y(t+s) = C^f(t+s)\Phi^f(t+s,t)x(t) + \\ + \sum_{\tau=t}^{t+s-1} [C^f(t+s-1)\Phi^f(t+s-1,\tau)M(\tau)v(\tau)] + N(t+s)v(t+s),$$

$$x(t) \mapsto E[\exp(iu^T\bar{y}) \,|\, F^{x(t)}] = \exp(iu^T S^f x(t) - \tfrac{1}{2}u^T L u)$$

for some deterministic matrix L. Now this map is injective on the support of x iff the map $x(t) \mapsto S^f(t)x(t)$ is injective on the support of $x(t)$. The support of $x(t)$ is $im(Q(t))$. Thus the map $x(t) \mapsto S^f(t)x(t)$ is injective on the support of $x(t)$ iff for all $w \in \mathbf{R}^n$ $S^f(t)Q(t)w=0$ implies that $Q(t)w=0$, which is true iff $ker(S^f(t)Q(t)) \subset ker(Q(t))$ iff $ker(S^f(t)Q(t))=ker(Q(t))$, since $ker(Q(t)) \subset ker(S^f(t)Q(t))$ always holds.
b. If σ is stochastically observable then there exists a $t_1>0$ such that for any $t \in T$ it is stochastically observable on the interval $\{t,t+1,\ldots,t+t_1\}$. From a then follows that with

$$S_1^f = \begin{bmatrix} C^f \\ C^f A^f \\ \cdots \\ C^f (A^f)^{t_1} \end{bmatrix}$$

$ker(S_1^f Q)=ker(Q)$. If $t_1<n$ then (3.3.2) holds, else the same conclusion is reached by applying the Cayley-Hamilton theorem. Conversely, the condition (3.3.2) and a imply that σ is stochastically observable on the interval $\{t,t+1,\ldots,t+n-1\}$. Because condition (3.3.2) does not depend on $t \in T$, it is equivalent to $F^{x(s)}=F^{S^f x(s)}$ for some $s \in T$. By stationarity this then holds for all $s \in T$.
c. It follows from standard stochastic theory results and the assumption that (A^f,G) is a reachable pair that the solution $Q \in \mathbf{R}^{n \times n}$ of the Lyapunov equation is such that $Q>0$. Hence the condition $ker(S^f Q)=ker(Q)$ is equivalent to $ker(S^f)=0$, to $rank(S^f)=n$, and to (A^f,C^f) an observable pair. □

Definition 3.3.3. Consider the stochastic system

$$\sigma = \{\Omega,F,P,T,\mathbf{R}^p,B^p,\mathbf{R}^n,B^n,y,x\} \in S\Sigma.$$

a. This system is called *stochastically reconstructible* on the interval $\{t-1,t-2,\ldots,t-t_1\}$ if the map

$$x(t) \mapsto E[\exp(i\sum_{s=1}^{t_1} u(s)^T y(t-s)) \,|\, F^{x(t)}]$$

is injective on the support of $x(t)$.
b. Assume that the system is stationary. Then it is called *stochastically reconstructible* if there exist $t,t_1 \in T$, $0<t_1<\infty$, such that it is stochastically reconstructible on the interval $\{t-1,\ldots,t-t_1\}$. By stationarity this then holds for any $t \in T$.

Theorem 3.3.4. *Consider the Gaussian system*

$$\sigma = \{\Omega,F,P,T,\mathbf{R}^p,B^p,\mathbf{R}^n,B^n,y,x\} \in GS\Sigma$$

with backward representation

$$x(t-1) = A^b(t)x(t) + Mv(t),$$
$$y(t-1) = C^b(t)x(t) + Nv(t),$$

with $v(t) \in G(0,V(t))$.
a. *The system σ is stochastically reconstructible on the interval $\{t-1,t-2,\ldots,t-t_1\}$ iff*

$$ker(S^b(t)Q(t))=ker(Q(t)), \tag{3.3.3}$$

iff $F^{x(t)} = F^{S^b(t)x(t)}$, *where* $x(t) \in G(0, Q(t))$ *and*

$$S^b(t) = \begin{bmatrix} C^b(t) \\ C^b(t-1)\Phi^b(t-1,t) \\ \cdots \\ C^b(t-t_1)\Phi^b(t-t_1,t) \end{bmatrix}.$$

and where Φ^b *represents the backward state transition matrix associated with the above backward representation.*

b. *Assume that the system* σ *is stationary with backward representation*

$$x(t-1) = A^b x(t) + Mv(t),$$
$$y(t-1) = C^b x(t) + Nv(t),$$

with $v(t) \in G(0, V)$. *Stationarity implies that* A^b *is exponentially stable. Let* $x(t) \in G(0, Q)$ *in which* $Q \in \mathbb{R}^{n \times n}$ *is the solution of the discrete-time Lyapunov equation*

$$Q = A^b Q (A^b)^T + MVM^T.$$

Then this system is stochastically reconstructible iff

$$ker(S^b Q) = ker(Q), \quad (3.3.4)$$

iff $F^{x(t)} = F^{S^b x(t)}$ *for some* $t \in T$, *in which*

$$S^b = \begin{bmatrix} C^b \\ C^b A^b \\ \cdots \\ C^b (A^b)^{n-1} \end{bmatrix}.$$

c. *Consider the assumptions of b. Let* $G \in \mathbb{R}^{n \times n}$ *be such that* $GG^T = MVM^T$. *Assume that* (A^b, G) *is a reachable pair. Then the system is stochastically reconstructible iff* (A^b, C^b) *is an observable pair.*

PROOF. The proof of this result is analogous to that of 3.3.2. □

Note that the condition (3.3.2) is expressed in terms of the matrices (A^f, C^f) of the forward representation of the Gaussian system and the condition (3.3.4) is expressed in terms of the matrices (A^b, C^b) of the backward representation. See section 3.2 for the way the matrices of the forward and backward representation are related.

3.4. The weak Gaussian stochastic realization problem

Attention is again directed to the problem of modelling by a stochastic system. So, one is given a measure on the observed process that has been estimated from the data. One is asked to determine a stochastic system in the model class such that the measure restricted to the observation process equals the given measure.

PROBLEM 3.4.1. *The* weak Gaussian stochastic realization problem *for a stationary Gaussian process is, given a stationary Gaussian process on* $T = \mathbb{Z}$ *taking values in* $(\mathbb{R}^p, B^p)$ *having mean value function zero and covariance function* $W: T \to \mathbb{R}^{p \times p}$, *to solve the following subproblems.*

a. *Does there exist a stationary Gaussian system*

$$\sigma = \{\Omega, F, P, T, \mathbb{R}^p, B^p, \mathbb{R}^n, B^n, y, x\} \in GS\Sigma$$

such that the output process y of this system equals the given process in distribution. This means that these processes have the same family of finite dimensional distributions. Effectively this means that the *covariance function of the output process must be equal to the given covariance function W because both processes are Gaussian. If such a system exists, then one calls σ a* weak Gaussian stochastic realization *of the given process, or, if the context is known, a* stochastic realization.

b. *Classify all minimal stochastic realizations of the given process. A weak Gaussian stochastic realization is called* minimal *if the dimension of the state space is minimal. The following subproblems must be solved:*
   1. *characterize those stochastic realizations that are minimal;*
   2. *obtain the classification as such;*
   3. *indicate the relation between two minimal stochastic realizations;*
   4. *produce an algorithm that constructs all minimal weak Gaussian stochastic realizations of the given process.*

In problem 3.4.1 one is given a stationary Gaussian process with zero mean value function. Such a process is thus completely characterized by its covariance function. In part a. of this problem the question is whether the given process can be the output of a stationary Gaussian system. Because by definition such a Gaussian system has a finite-dimensional state space, not all stationary Gaussian processes can be the output process of a Gaussian system. The question should therefore be interpreted as to determine a necessary and sufficient condition on the given process, or its covariance function, such that it can be the output process of a Gaussian system.

In part b. of problem 3.4.1 a classification is asked for. This question arises because a stochastic realization, if it exists, is in general nonunique. This will be indicated below. The dimensions of the state space of two stochastic realizations may also be different in general. For system theoretic reasons, such as identifiability, one should restrict attention to those stochastic realizations for which the dimension of the state space is minimal. Such a realization is called *minimal.* In general minimal stochastic realizations are also nonunique. A classification of all minimal stochastic realizations is then useful for the solution of the identifiability question. The above defined problem is related to the problem of determining spectral factorizations of the spectral density of the given process.

Below a notation is used for the parameters of a time-invariant finite-dimensional linear system of the form

$$x(t+1) = Ax(t) + Bu(t),$$
$$y(t) = Cx(t) + Du(t),$$

with $U=\mathbb{R}^m$, $X=\mathbb{R}^n$, $Y=\mathbb{R}^p$, $u:T\to U$, $x:T\to X$, $y:T\to Y$. The notation is then

$$pls = \{p,n,m,A,B,C,D\} \in L\Sigma P.$$

In the formulation of theorem 3.4.2 use is made of the set $\mathbf{Q}_{\overline{pls}}$. The definition of this set is given in subsection 3.5.

THEOREM 3.4.2. *Consider the weak Gaussian stochastic realization problem for a stationary Gaussian process as posed in 3.4.1. Assume that* $\lim_{t\to\infty} W(t)=0$ *and that* $W(0)>0$.

a. *There exists a weak Gaussian stochastic realization of the given process iff there exists a* $pls=\{p,n,p,F,G,H,J\} \in L\Sigma P$ *with* $J=J^T$ *such that*

$$W(t) = \begin{cases} HF^{t-1}G, & \text{if } t>0, \\ 2J, & \text{if } t=0, \\ G^T(F^T)^{-t-1}H^T, & \text{if } t<0. \end{cases} \quad (3.4.1)$$

(a function having the form (3.4.1) will be called a discrete-time Bohl function; *the right hand side of (3.4.1) will be called a* covariance realization *of the covariance function W.)*
iff

$$\hat{W}(\lambda) = \sum_{t \in \mathbb{Z}} W(t)\lambda^{-|t|} \qquad (3.4.2)$$

is a rational function. The dimension n in the covariance realization (3.4.1) is also called the McMillan degree *of the covariance function.*

b. *A weak Gaussian stochastic realization is minimal iff it is stochastically observable and stochastically reconstructible.*

c. *A minimal weak Gaussian stochastic realization is nonunique in two ways.*

1. *If* $pgs_1 = \{p,n,m,A,C,M,N,V\} \in GS\Sigma P$ *are the parameters of a forward representation of a minimal stochastic realization, and if* $S \in \mathbb{R}^{n \times n}$ *is nonsingular, then* $pgs_2 = \{p,n,m,SAS^{-1},CS^{-1},SM,N,V\} \in GS\Sigma P$ *are also the parameters of a forward representation of a minimal stochastic realization.*
2. *Fix the parameters of a minimal covariance realization as given in a. above,*

$$pls = \{p,n,p,F,G,H,J\} \in L\Sigma P_{\min}.$$

Denote the parameters of a forward representation of a minimal Gaussian stochastic realization by $\{p,n,A,C,V\}$ *and the set of such parameters by* $WGSRP_{\min}$. *Define the classification map*

$$c_{pls}: \mathbf{Q}_{\overline{pls}} \to WGSRP_{\min},\ c_{pls}(Q) = \{p,n,A,C,V\}, \qquad (3.4.3)$$

by $A = F$, $C = H$,

$$V = \overline{V}(Q) = \begin{bmatrix} Q - FQF^T & G - FQH^T \\ G^T - HQF^T & 2J - HQH^T \end{bmatrix}.$$

Then, for fixed $pls \in L\Sigma P_{\min}$ *is* c_{pls} *a bijection. Thus all minimal weak Gaussian stochastic realizations are classified by the elements of* $\mathbf{Q}_{\overline{pls}}$.

d. *The stochastic realization algorithm as defined in 3.4.3 below is well defined and constructs all minimal weak Gaussian stochastic realizations.*

ALGORITHM 3.4.3. *The stochastic realization algorithm for weak Gaussian stochastic realizations of stationary Gaussian processes.*

Data: given a stationary Gaussian process with zero mean value function and covariance function $W: T \to \mathbb{R}^{p \times p}$. *Assume that the condition of 3.4.2.a. holds.*

1. *Determine a minimal covariance realization of W via a realization algorithm for time-invariant finite-dimensional linear systems, or* $pls = \{p,n,p,F,G,H,J\} \in L\Sigma P_{\min}$, *such that*

$$W(t) = \begin{cases} HF^{t-1}G, & \text{if } t > 0, \\ 2J, & \text{if } t = 0, \\ G^T(F^T)^{-t-1}H^T, & \text{if } t < 0. \end{cases} \qquad (3.4.4)$$

For algorithms for this step see books on linear system theory.

2. *Determine a* $Q \in \mathbf{Q}_{\overline{pls}}$, *or a* $Q \in \mathbb{R}^{n \times n}$ *satisfying* $Q = Q^T \geqslant 0$,

$$\begin{bmatrix} Q - FQF^T & G - FQH^T \\ G^T - HQF^T & 2J - HQH^T \end{bmatrix} \geqslant 0. \qquad (3.4.5)$$

3. *Let*

$$A = F,\ C = H,\ M = (I_n\ 0) \in \mathbb{R}^{n \times (n+p)},\ N = (0\ I_p) \in \mathbb{R}^{p \times (n+p)},$$

$$V = \overline{V}(Q) = \begin{bmatrix} Q - FQF^T & G - FQH^T \\ G^T - HQF^T & 2J - HQH^T \end{bmatrix} \in \mathbb{R}^{(n+p) \times (n+p)},$$

construct a probability space by

$$\Omega = (\mathbf{R}^{(n+p)})^T,\ F = \Pi_T \otimes B^{(n+p)},\ v:\Omega\times T \to \mathbf{R}^{(n+p)},\ v(\omega,t) = \omega(t),\ P:F\to[0,1]$$

a probability measure such that v is a Gaussian white noise process with intensity V, $x:\Omega\times T\to\mathbf{R}^n$ $y:\Omega\times T\to\mathbf{R}^p$ defined by

$$x_{t+1} = Ax_t + Mv_t,\ x_{-\infty}=0, \tag{3.4.6}$$

$$y_t = Cx_t + Nv_t. \tag{3.4.7}$$

Then

$$\sigma = \{\Omega,F,P,T,\mathbf{R}^p,B^p,\mathbf{R}^n,B^n,y,x\} \in GS\Sigma \tag{3.4.8}$$

is a minimal weak Gaussian stochastic realization of the given process, meaning that the output process y is a Gaussian process with covariance function equal to the given covariance function W.

A mistake that is sometimes made is the following. Consider the following forward representation of a Gaussian system

$$x_{t+1} = Ax_t + Mv_t,$$

$$y_t = Cx_t + Nv_t,$$

with $v_t \in G(0,V)$. A statement is that if the pair of matrices $(A,MV^{1/2})$ is a reachable pair and if (A,C) is an observable pair, that then the stochastic realization described by the above system representation is a minimal realization of the output process. This statement is false as the following example shows.

EXAMPLE 3.4.4. Consider the Gaussian system

$$\sigma = \{\Omega,F,P,T,\mathbf{R},B,\mathbf{R},B,y,x\} \in GS\Sigma$$

with forward representation

$$x_{t+1} = ax_t + bv_t,$$

$$y_t = x_t + v_t,$$

with $v_t \in G(0,1)$, $a\in(-1,+1)$, $a\neq 0$, $b=(a^2-1)/a$.
a. Then (a,b) is a reachable pair and $(a,1)$ is an observable pair.
b. The system σ is a nonminimal realization of its output process.

It is possible to interpret certain stochastic realizations as a Kalman filter but this will not be done here. For a reference see [24].

The implication of the weak Gaussian stochastic realization problem for the identifiability question is illustrated by the following example.

EXAMPLE 3.4.5. Consider the time-invariant Gaussian system

$$\sigma = \{\Omega,F,P,T,\mathbf{R},B,\mathbf{R},B,y,x\} \in GS\Sigma$$

with forward representation

$$x_{t+1} = ax_t + (1\ 0)v_t, \tag{3.4.9}$$

$$y_t = cx_t + (0\ 1)v_t, \tag{3.4.10}$$

with $v_t \in G(0,V)$,

$$V = \begin{pmatrix} v_{11} & 0 \\ 0 & v_{22} \end{pmatrix}. \tag{3.4.11}$$

Consider the asymptotic Kalman filter for the Gaussian system (3.4.9 & 3.4.10)

$$\hat{x}_{t+1} = a\hat{x}_t + k(y_t - c\hat{x}_t), \quad (3.4.12)$$

$$\overline{v}_t = y_t - c\hat{x}_t, \quad (3.4.13)$$

in which $\overline{v}:\Omega\times T\rightarrow\mathbf{R}$ is a Gaussian white noise process with $\overline{v}_t \in G(0,r)$. This asymptotic Kalman filter may be rewritten as

$$\hat{x}_{t+1} = a\hat{x}_t + k\overline{v}_t = a\hat{x}_t + (1\ 0)v_1(t), \quad (3.4.14)$$

$$y_t = c\hat{x}_t + \overline{v}_t = c\hat{x}_t + (0\ 1)v_1(t), \quad (3.4.15)$$

in which $v_1:\Omega\times T\rightarrow\mathbf{R}^2$ is a Gaussian white noise process with $v_1(t)\in G(0,V_1)$,

$$V_1 = \begin{bmatrix} k^2r & kr \\ kr & r \end{bmatrix} = \begin{bmatrix} k \\ 1 \end{bmatrix} r(k\ 1). \quad (3.4.16)$$

From these forward representations one deduces that (3.4.9 & 3.4.10) and (3.4.14 & 3.4.15) are both weak Gaussian stochastic realizations of the output process y. This may be verified by computing the covariance function of the output process. This example shows that one may not be able to uniquely determine the parameters of the noise process of a Gaussian system, here (3.4.11) and (3.4.16), from the covariance function of the output process. For results on the parametrization of Gaussian systems see [34].

Attention has also been devoted to the partial weak Gaussian stochastic realization problem in which one is not given a covariance function on all of $T=\mathbf{Z}$ but only on a finite time set, say $T=\{-t_1,-t_1+1,\ldots,-1,0,1,\ldots,t_1\}$. The motivation for this problem is that in practice one can estimate from a finite time series only the covariance function on a finite time set.

3.5. The dissipation matrix inequality

In subsection 3.4 it has been stated that the minimal weak Gaussian stochastic realizations are classified by the set $\mathbf{Q}_{\overline{pls}}$. In this section the set $\mathbf{Q}_{\overline{pls}}$ and its dual $\mathbf{Q}_{pls}$ will be considered. Throughout this section $J=J^T$. The results of this subsection may be found in [23, 24].

DEFINITION 3.5.1. Let $pls=\{p,n,p,F,G,H,J\}\in L\Sigma P$ with $J\geq 0$ and

$$\mathbf{Q}_{pls} = \{Q\in\mathbf{R}^{n\times n} \mid Q=Q^T\geq 0,\ V(Q) = \begin{bmatrix} Q-F^TQF & H^T-F^TQG \\ H-G^TQF & 2J-G^TQG \end{bmatrix} \geq 0\}, \quad (3.5.1)$$

and for $\overline{pls}=\{p,n,p,F^T,H^T,G^T,J\}\in L\Sigma P$

$$\mathbf{Q}_{\overline{pls}} = \{Q\in\mathbf{R}^{n\times n} \mid Q=Q^T\geq 0,\ \overline{V}(Q) = \begin{bmatrix} Q-FQF^T & G-FQH^T \\ G^T-HQF^T & 2J-HQH^T \end{bmatrix} \geq 0\}. \quad (3.5.2)$$

PROBLEM 3.5.2. Given $pls\in L\Sigma P$ and $\mathbf{Q}_{pls}$.

a. Classify all elements of $\mathbf{Q}_{pls}$.

b. Determine an algorithm that constructs all elements of $\mathbf{Q}_{pls}$.

PROPOSITION 3.5.3. *Consider* $pls=\{p,n,p,F,G,H,J\}\in L\Sigma P_{\min}$ *and* $\mathbf{Q}_{pls}$. *Assume that* $\mathbf{Q}_{pls}\neq\emptyset$, *and that* $J>0$. *Then* $\mathbf{Q}_{pls}$ *is a convex, closed and bounded set, and there exists a* Q^-, $Q^+\in\mathbf{Q}_{pls}$ *such that for any* $Q\in\mathbf{Q}_{pls}$, $Q^-\leq Q\leq Q^+$.

DEFINITION 3.5.4.

a. *The* regular part of $\mathbf{Q}_{pls}$ *is defined as*

$$\mathbf{Q}_{pls,r} = \{Q \in \mathbf{Q}_{pls} \mid 2J - G^TQG > 0\}.$$

The set $\mathbf{Q}_{pls}$ *will be called* regular *if* $\mathbf{Q}_{pls} = \mathbf{Q}_{pls,r}$.

b. *For* $Q \in \mathbf{R}^{n\times n}$ *with* $Q = Q^T$ *and* $2J - G^TQG > 0$ *define*

$$D(Q) = Q - F^TQF - [H^T - F^TQG][2J - G^TQG]^{-1}[H^T - F^TQG]^T. \tag{3.5.3}$$

c. *Correspondingly define*

$$\mathbf{Q}_{\overline{pls},r} = \{Q \in \mathbf{Q}_{\overline{pls}} \mid 2J - HQH^T > 0\},$$
$$\overline{D}(Q) = Q - FQF^T - [G - FQH^T][2J - HQH^T]^{-1}[G - FQH^T]^T, \tag{3.5.4}$$

and $\mathbf{Q}_{\overline{pls}}$ *is regular if* $\mathbf{Q}_{\overline{pls}} = \mathbf{Q}_{\overline{pls},r}$.

PROPOSITION 3.5.5. *Let* $pls = \{p,n,p,F,G,H,J\} \in L\Sigma P$. *Let* $Q \in \mathbf{R}^{n\times n}$, $Q = Q^T$.

a. *Assume that* $2J - G^TQG > 0$, *and let*

$$T = \begin{bmatrix} I & 0 \\ -[2J - G^TQG]^{-1}[H - G^TQF] & I \end{bmatrix} \in \mathbf{R}^{(n+p)\times(n+p)}. \tag{3.5.5}$$

Then

$$\begin{bmatrix} D(Q) & 0 \\ 0 & 2J - G^TQG \end{bmatrix} = T^TV(Q)T, \tag{3.5.6}$$

and

$$V(Q) = T^{-T}\begin{bmatrix} D(Q) & 0 \\ 0 & 2J - G^TQG \end{bmatrix}T^{-1}, \tag{3.5.7}$$

where $V(Q)$ *is as defined in 3.5.1.*

b. *Assume that* $2J - G^TQG > 0$. *Then* $V(Q) \geq 0$ *iff* $D(Q) \geq 0$. *Also* $V(Q) > 0$ *iff* $D(Q) > 0$. *In fact,* $rank\,(V(Q)) = rank\,(D(Q)) + p$.

c.

$$\mathbf{Q}_{pls,r} = \{Q \in \mathbf{R}^{n\times n} \mid Q = Q^T \geq 0,\ 2J - G^TQG > 0,\ D(Q) \geq 0\}.$$

Notation for the boundary of $\mathbf{Q}_{pls}$ will be needed. The following notation will be used in the sequel,

$$\|Q\|_2 = \left[sup_{x \in \mathbf{R}^n, x \neq 0} \frac{x^TQ^TQx}{x^Tx} \right]^{1/2}, \tag{3.5.8}$$

$$B(Q,\epsilon) = \{S \in \mathbf{R}^{n\times n} \mid \|S - Q\|_2 \leq \epsilon\}. \tag{3.5.9}$$

DEFINITION 3.5.6. Let $pls \in L\Sigma P$ and consider $\mathbf{Q}_{pls}$. Define the *boundary* of $\mathbf{Q}_{pls}$ as the set

$$\partial\mathbf{Q}_{pls} = \{Q \in \mathbf{Q}_{pls} \mid \forall\, \epsilon \in \mathbf{R},\ \epsilon > 0,\ \exists S \in B(Q,\epsilon) \text{ such that } S = S^T,\ S \neq Q,\ S \notin \mathbf{Q}_{pls}\},$$

and the *interior of* $\mathbf{Q}_{pls}$ as the set

$$int\,(\mathbf{Q}_{pls}) = \mathbf{Q}_{pls} \cap (\partial\mathbf{Q}_{pls})^c.$$

PROPOSITION 3.5.7. *Let* $pls = \{p,n,p,F,G,H,J\} \in L\Sigma P$.

a. $Q \in \partial\mathbf{Q}_{pls}$ *iff* $V(Q)$ *is singular.* $Q \in int\,(\mathbf{Q}_{pls})$ *iff* $V(Q) > 0$.

b. *Assume that* $\mathbf{Q}_{pls}$ *is regular. Then* $Q \in \partial\mathbf{Q}_{pls}$ *iff* $D(Q)$ *is singular; and* $Q \in int\,(\mathbf{Q}_{pls})$ *iff* $D(Q) > 0$.

DEFINITION 3.5.8. *Let* $pls=\{p,n,p,F,G,H,J\}\in L\Sigma P$ *and consider* $\mathbf{Q}_{pls}$.
a. *The set of* singular boundary points of $\mathbf{Q}_{pls}$ *is defined as*

$$\partial\mathbf{Q}_{pls,s} = \{Q\in\partial\mathbf{Q}_{pls}\,|\,rank\,(V(Q)) = rank\,(2J-G^TQG)\}.$$

b. *The set of* singular boundary points of the regular part of $\mathbf{Q}_{pls}$ *is defined as*

$$\partial\mathbf{Q}_{pls,r,s} = \{Q\in\mathbf{Q}_{pls,r}\cap\partial\mathbf{Q}_{pls}\,|\,rank(V(Q))=p\}.$$

THEOREM 3.5.9. *Let* $pls=\{p,n,p,F,G,H,J\}\in L\Sigma P_{\min}$. *Assume that* $\mathbf{Q}_{pls}\neq\varnothing$ *and that it is regular. Let*

$$F^- = F - G[2J-G^TQ^-G]^{-1}[H^T-FQ^-G]^T.$$

Then $Q^-+\Delta Q\in\mathbf{Q}_{pls}$ *and* $\Delta Q>0$ *iff*
1. $\Delta Q\in\mathbb{R}^{n\times n}$, $\Delta Q>0$;
2.

$$(\Delta Q)^{-1} - F^-(\Delta Q)^{-1}(F^-)^T - G[2J-G^TQ^-G]^{-1}G^T - S = 0, \tag{3.5.10}$$

for some $S\in\mathbb{R}^{n\times n}$, $S=S^T\geq 0$;
3. $sp(F^-)\subset C^-$.

3.6. The strong Gaussian stochastic realization problem

PROBLEM 3.6.1. *The* strong Gaussian stochastic realization problem *for a stationary Gaussian process is, given a probability space* (Ω,F,P), *a time index set* $T=\mathbb{Z}$ *and a stationary Gaussian process* $z:\Omega\times T\to\mathbb{R}^p$ *having zero mean value function and covariance function* $W:T\to\mathbb{R}^{p\times p}$, *to solve the following subproblems.*
a. *Does there exist a stationary Gaussian system*

$$\sigma = \{\Omega,F,P,T,\mathbb{R}^p,B^p,\mathbb{R}^n,B^n,y,x\}\in GS\Sigma$$

with forward representation

$$x_{t+1} = Ax_t + Mv_t,\ x_0,$$
$$y_t = Cx_t + Nv_t,$$

such that
1. $y_t=z_t$ *a.s. for all* $t\in T$;
2. $F^{x_t}\subset F^y_\infty$ *for all* $t\in T$.
If such a system exists then one calls σ *a* strong Gaussian stochastic realization *of the given process, or, if the context is known, a* stochastic realization.

b. *Classify all minimal stochastic realizations of the given process. A strong Gaussian stochastic realization is called* minimal *if the dimension of the state space is minimal.*

The difference between the weak and the strong Gaussian stochastic realization problems is that the given process and the output process of the Gaussian stochastic system are equal in the sense of the family of finite-dimensional distributions respectively equal in the sense of almost surely. For the strong Gaussian stochastic realization problem this requires that the stochastic system is constructed on the same probability space as the given process. Therefore the state process has to be constructed from the given process, and this explains condition 2 of problem 3.6.1.a.

For a survey of the strong Gaussian stochastic realization problem the reader is referred to the paper [47].

3.7. Pseudo-distances on the set of probability measures

The purpose of this subsection is define distances on the set of probability measures as a preparation for the approximate stochastic realization problem to be discussed in the next subsection.

DEFINITION 3.7.1. *Let X be a set. A* pseudo-distance *is a function* $d:X \times X \to \mathbf{R}$ *such that*
1. $d(x,y) \geq 0$ *for all* $x,y \in X$;
2. $d(x,y)=0$ *iff* $x=y$.

If a pseudo-distance is not symmetric then one may construct its symmetrized version. A pseudo-distance need not satisfy the triangle inequality.

DEFINITION 3.7.2. *Let*

$$F_{2s} = \{f:\mathbf{R}_+ \to \mathbf{R} \mid f \in C^2, f(1)=0, \forall x \in (0,\infty), f''(x)>0\}.$$

DEFINITION 3.7.3. *Given a measurable space* (Ω,F), *let*

$$\underline{P} = \{P:F \to \mathbf{R}_+ \mid P \text{ is a probability measure }\}.$$

For $f \in F_{2s}$ *define the* pseudo-distance $d_f:\underline{P} \times \underline{P} \to \mathbf{R}$ *on the set of probability measures* $\underline{P}$ *on* (Ω,F) *by*

$$d_f(P_1,P_2) = E_Q[f(\frac{r_1}{r_2})r_2] = E_{P_2}[f(\frac{r_1}{r_2})]$$

where Q is a σ-finite measure on (Ω,F) *such that*

$$P_1 < Q \text{ with } \frac{dP_1}{dQ} = r_1,\ P_2 < Q \text{ with } \frac{dP_2}{dQ} = r_2.$$

The pseudo-distance d_f *is also called the* f-information measure, *the* f-entropy *or the* f-divergence.

A σ-finite measure Q as mentioned above always exists, for example $Q=P_1+P_2$ will do. In case $(\Omega,F)=(\mathbf{R},B)$ one may sometimes take Q to be Lebesgue measure. Because $r_2>0$ a.s. P_2 the above expression is well defined. The above definition has been given in [1].

PROPOSITION 3.7.4. *[1].*
a. *The function* d_f *defined in 3.7.3. is a pseudo-distance.*
b. *The pseudo-distance* d_f *does not depend on the choice of the σ-finite measure Q.*

DEFINITION 3.7.5. *The* Kullback-Leibler *pseudo-distance is defined as* $d_{f_1}:\underline{P} \times \underline{P} \to \mathbf{R}$ *with*

$$f_1:\mathbf{R}_+ \to \mathbf{R},\ f_1(x) = \begin{cases} x\ln(x), & x>0, \\ 0, & x=0, \end{cases}$$

$$d_{f_1}(P_1,P_2) = E_{P_2}[f_1(\frac{r_1}{r_2})] = E_Q[f_1(\frac{r_1}{r_2})r_2] = E_Q[r_1 \ln(\frac{r_1}{r_2}) I_{(r_2>0)}].$$

DEFINITION 3.7.6. *The* Hellinger *pseudo-distance is defined as* $d_{f_2}:\underline{P} \times \underline{P} \to \mathbf{R}$ *with*

$$f_2:\mathbf{R}_+ \to \mathbf{R},\ f_2(x) = (\sqrt{x}-1)^2,$$

$$d_{f_2}(P_1,P_2) = E_{P_2}[(\sqrt{\frac{r_1}{r_2}}-1)^2] = E_Q[(\sqrt{r_1}-\sqrt{r_2})^2].$$

The Hellinger pseudo-distance is symmetric.

Consider the set of functions on $T=\mathbb{Z}$ with values in $\mathbb{R}^k$. Let P be the set of Gaussian measures on this space that make the underlying process a stationary Gaussian process with zero mean value function. An expression for the Kullback-Leibler pseudo-distance on this set was derived in [43].

PROPOSITION 3.7.7. *Let P_1,P_2 be two probability measures on the set of functions defined on $T=\mathbb{Z}$ with values in $\mathbb{R}^k$. Assume that these measures are such that the underlying process is Gaussian, stationary, has zero mean value function, and covariance functions W_1,W_2 respectively. Moreover, assume that these covariance functions admit spectral densities $\hat{W}_1,\hat{W}_2$ respectively and that they satisfy condition C of [43]. Then the Kullback-Leibler pseudo-distance is given by the expression*

$$d_{KL}(P_1,P_2) = \frac{1}{4\pi}\int_{-\pi}^{\pi}[\, tr(\hat{W}_1^{-1}(\lambda)[\hat{W}_2(\lambda)-\hat{W}_1(\lambda)]) - \ln(\hat{W}_1^{-1}(\lambda)\hat{W}_2(\lambda))\,]d\lambda.$$

3.8. The approximate weak Gaussian stochastic realization problem

How to fit to data a model in the form of a Gaussian system? In engineering, in biology and in economics there are many modelling problems for which an answer to this question is useful. As indicated in section 2, from data one may estimate a measure on the set of observation trajectories. In case that one models the observations as a sample function of a Gaussian process, one may estimate its covariance function. Suppose further that one wants to model the observations as the output process of a stationary Gaussian system. Such a system has a finite-dimensional state space. In theorem 3.4.2 it has been shown that a covariance function has a stochastic realization as a Gaussian system only if it has a covariance realization as indicated or if it is rational. Now an arbitrary covariance function obtained from data may not correspond to such a covariance function. Therefore one has to resort to approximation.

The approximate stochastic realization problem is then to determine a stochastic system in a specified class such that the measure on the output process of this system approximates the measure on the same space determined from the data. Attention below will be restricted to the class of stationary Gaussian systems with dimension of the state space less or equal to $n\in\mathbb{Z}_+$. As a measure of fit the Kullback-Leibler pseudo-distance will be taken as mentioned in subsection 3.7. A measure of complexity will not be considered here; it may be based on stochastic complexity as indicated in section 2.

PROBLEM 3.8.1. Approximate weak Gaussian stochastic realization problem. *Let Y^T denote the set of time series defined on $T=\mathbb{Z}$ with values in $\mathbb{R}^p$, and let $P(Y^T)$ denote the set of probability measures on Y^T. Given is a Gaussian measure $P_0\in P(Y^T)$ such that the underlying process corresponds to a stationary Gaussian process with zero mean function. Given is also an integer $n\in\mathbb{Z}_+$ and let $GS\Sigma(n)$ be the set of Gaussian systems with state space dimension $\leq n$. Solve the optimization problem*

$$inf_{\sigma\in GS\Sigma(n)}\, d_{KL}(P_0,P(\sigma))$$

where d_{KL} is the Kullback-Leibler pseudo-distance on the set of probability measures on $P(Y^T)$, and $P(\sigma)\in P(Y^T)$ is the probability measure on Y^T associated with the Gaussian system $\sigma\in GS\Sigma(n)$.

As indicated in 3.7.7, if the pseudo-distance on the set of Gaussian measures is the Kullback-Leibler measure then the pseudo-distance may be expressed as a pseudo-distance on the set of covariance functions

$$d_{KL}(P_0,P(\sigma)) = d_1(W_0,W(\sigma))$$

where W_0 is the covariance function associated with the Gaussian measure P_0 and $W(\sigma)$ the covariance function associated with the Gaussian measure $P(\sigma)$. Note that the covariance function $W(\sigma)$ is a rational function with McMillan degree less or equal to n because it corresponds to a Gaussian system of state space dimension less or equal than n. The approximate weak Gaussian stochastic realization problem may therefore be considered as an approximation problem for a covariance function. In this problem the approximant $W(\sigma)$ has to be a rational function of McMillan degree at most n while the given covariance function W_0 may neither be rational nor of finite McMillan degree.

The approximate stochastic realization problem 3.8.1 is unsolved. Approaches along three different lines have been investigated.

Approach 1. Given any pseudo-distance d_1, problem 3.8.1 can be reformulated as an approximation problem for covariance functions with the criterion

$$d_1(W_0, W(\sigma))$$

where W_0 is the covariance function associated with the Gaussian measure P_0 and $W(\sigma)$ the covariance function associated with the Gaussian measure $P(\sigma)$ related to $\sigma \in GS\Sigma$.

PROBLEM 3.8.2. *Given a covariance function* $W_0 : T \rightarrow \mathbb{R}^{p \times p}$ *solve*

$$inf_{\sigma \in GS\Sigma_n} d_1(W_0, W(\sigma)).$$

The pseudo-distance d_1 on the set of covariance functions may be taken to be the Hankel norm or the H-infinity norm. Possibly the $\mathcal{L}_2$-norm is suitable.

The above problem may be rephrased as, given a not necessarily rational covariance function, to determine a rational covariance function that approximates the given covariance with respect to an approximation criterion. Note that a function is a covariance function iff it is anti-symmetric and a positive definite function.

It seems that a Hankel norm approximation of a covariance function is not itself a covariance function. The positive definiteness of a covariance function is therefore an essential constraint. References on this approach are [28, 29, 31, 38, 51, 65].

There is a related approach in which one first determines a spectral factor of the given covariance function and then a rational approximation of the spectral factor. This approach seems too restrictive to start with, although it may be the solution to some approximation criterion.

Of course, given any rational approximation of the covariance function one will still have to determine a state space realization for it.

Approach 2. By analogy with the approximate prediction problem for finite-dimensional Gaussian random variables, algorithms have been proposed for the approximate weak Gaussian stochastic realization problem.

ALGORITHM 3.8.3. Let be given a covariance function W_0.

1. Solve an approximate prediction problem. Fix $t \in T$. Let

$$y^+(t) = \begin{bmatrix} y_t \\ y_{t+1} \\ \ldots \\ y_{t+r} \end{bmatrix}, \; y^-(t) = \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \ldots \\ y_{t-s} \end{bmatrix}.$$

The variance of the pair $(y^+(t), y^-(t))$ may be computed from the covariance function W_0. Let $n \in \mathbb{Z}_+$. Determine a matrix $S \in \mathbb{R}^{n \times s}$ such that with $x(t) = Sy^-(t)$ the following prediction criterion is minimized

$$inf_{S \in \mathbb{R}^{n \times s}} \; tr(E[(y^+(t) - E[y^+(t) \,|\, F^{x(t)}])(y^+(t) - E[y^+(t) \,|\, F^{x(t)}])^T]).$$

2. Determine a Gaussian system via regression by proceeding as follows,

$$\begin{bmatrix} x(t+1) \\ y(t) \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} x(t) + v(t), \;\; v(t) \in G(0, V),$$

where

$$\begin{bmatrix} A \\ C \end{bmatrix} = E[\begin{bmatrix} x(t+1) \\ y(t) \end{bmatrix} x(t)^T](E[x(t)x(t)^T])^{-1},$$

$$v(t) = \begin{bmatrix} x(t+1) \\ y(t) \end{bmatrix} - \begin{bmatrix} A \\ C \end{bmatrix} x(t).$$

Finally, replace the Gaussian process v with a Gaussian white noise process w with variance V.

The above algorithm in a somewhat different form appeared first in a paper of H. Akaike [3]. Other references are [11, 12, 44-46, 75, 76]. These papers differ mainly in the way they perform step 1 of the above algorithm. For canonical correlation analysis and the prediction problem see [27, 57].

It is not clear in what sense the Gaussian system determined in step 2 of the above algorithm is a good approximation to the given Gaussian process. In other words, the approximation criterion, although inspired by the static approximate prediction problem, is never mentioned. The replacement of the process v by a Gaussian white noise process is also unmotivated.

Approach 3. Canonical correlation analysis for finite-dimensional Gaussian random variables has been generalized to infinite-dimensional Hilbert spaces in [36, 37, 49]. One has investigated approximate prediction problems for time series by canonical correlation analysis techniques. Approximation bounds have been derived [30]. It remains to be seen whether this approach is useful in practice.

Approach 4. Inspired by the above mentioned second approach to the approximate weak Gaussian stochastic realization problem yet another approach has been formulated. This approach has been worked out by M. Stöhr at the Centre for Mathematics and Computer Science. The following results up to the end of section 3 are due to M. Stöhr and are as of yet unpublished.

NOTATION 3.8.4. *Let $k_1, k_2, n \in \mathbb{Z}_+$, $k = k_1 + k_2$. Recall that $G(0,Q)$ denotes a Gaussian measure, say on $\mathbb{R}^k$, with zero mean and variance Q. For $Q \in \mathbb{R}^{k \times k}$ the decomposition*

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}^T & Q_{22} \end{bmatrix}$$

will be used in which $Q_{11} \in \mathbb{R}^{k_1 \times k_1}$, $Q_{22} \in \mathbb{R}^{k_2 \times k_2}$, and $Q_{12} \in \mathbb{R}^{k_1 \times k_2}$, Let

$$\mathbb{Q}(n) = \{Q \in \mathbb{R}^{k \times k} \mid Q = Q^T \geqslant 0,\ rank(Q_{12}) \leqslant n\}.$$

PROBLEM 3.8.5. The static approximate weak Gaussian stochastic realization problem. *Given are $k_1, k_2, n \in \mathbb{Z}_+$, $k = k_1 + k_2$, and a Gaussian measure $G(0,Q_0)$ with $Q_0 = Q_o^T > 0$. Let d_{KL} be the Kullback-Leibler pseudo-distance on the set of Gaussian measures on $\mathbb{R}^k$. Solve*

$$inf_{G(0,Q_1),\ Q_1 \in \mathbb{Q}(n)}\ d_{KL}(G(0,Q_0), G(0,Q_1)).$$

One may interpret the above problem in the light of approach 2 indicated above. Associate the space $\mathbb{R}^{k_1}$ with the past of the observations, and the space $\mathbb{R}^{k_2}$ with the future of the observations. The Gaussian measure $G(0,Q_0)$ may then be associated with that derived from the data. In problem 3.8.5 one is asked to determine the measure $G(0,Q_1)$ with $Q_1 \in \mathbb{Q}(n)$. The latter condition implies that the dimension of the state space associated with $G(0,Q_1)$ is less or equal to n. Therefore the essential constraint on the dimension of the state space is taken care of.

PROPOSITION 3.8.6. *Consider problem 3.8.5. The Kullback-Leibler measure of two Gaussian measures $G(0,Q_0)$ and $G(0,Q_1)$ on $\mathbb{R}^k$ is given by the expression*

$$d_{KL}(G(0,Q_0), G(0,Q_1)) = ½[\ tr(Q_1^{-1}Q_0) - \ln(\det(Q_1^{-1}Q_0)) - k\ ]$$

$$= \tfrac{1}{2}[\sum_{i=1}^{k}(\lambda_i(Q_0,Q_1) - \ln(\lambda_i(Q_0,Q_1))) - k],$$

where $\{\lambda_i(Q_0,Q_1), i\in\mathbf{Z}_k\}$ *are the generalized eigenvalues of* Q_0 *with respect to* Q_1*, here defined as the zeroes of* $det(Q_1\lambda - Q_0)=0$.

It can be shown that the generalized eigenvalues are real and satisfy $\lambda_i(Q_0,Q_1)\geqslant 0$, for $i\in\mathbf{Z}_k$.

NOTATION 3.8.7. *For* $Q_0\in\mathbf{R}^{k\times k}$, $Q_0=Q_0^T>0$, $n\in\mathbf{Z}_+$ *let*

$$\Lambda(Q_0,n) = \left\{\begin{array}{c}\lambda\in\mathbf{R}_+^k \mid \exists\ Q\in\mathbf{Q}(n) \textit{ such that generalized eigenvalues} \\ \textit{of } Q_0 \textit{ with respect to } Q \textit{ are } \{\lambda_1,\ldots,\lambda_k\}\end{array}\right\}.$$

and for $\lambda\in\mathbf{R}_+^k$ *let*

$$\mathbf{Q}_s(Q_0,n,\lambda) = \left\{\begin{array}{c}Q\in\mathbf{Q}(n) \mid \textit{generalized eigenvalues} \\ \textit{of } Q_0 \textit{ with respect to } Q \textit{ are } \{\lambda_1,\ldots,\lambda_k\}\end{array}\right\}.$$

$$f:\mathbf{R}_+^k\rightarrow\mathbf{R}_+,\ f(\lambda) = \tfrac{1}{2}[\sum_{i=1}^{k}(\lambda_i - ln(\lambda_i)) - k].$$

It may be shown that the function f is convex. There are results on the structure of the matrices in the set $\mathbf{Q}_s(Q_0,n,\lambda)$.

PROBLEM 3.8.8. *Consider problem 3.8.5 and the notation 3.8.7. Solve*

$$inf_{\lambda\in\Lambda(Q_0,n)}\ f(\lambda).$$

Suppose that there exists a $\lambda^*\in\Lambda(Q_0,n)$ such that

$$f(\lambda^*) = inf_{\lambda\in\Lambda(Q_0,n)}\ f(\lambda).$$

The solution set of problem 3.8.5 is then given by $\mathbf{Q}_s(Q_0,n,\lambda^*)$. Note that problem 3.8.8 is the infimization of a convex function over the set $\Lambda(Q_0,n)$. The latter set is a cone. It is conjectured that it is a polyhedral cone. It may be shown that the optimal solution of problem 3.8.8 is such that $\sum_{i=1}^{k}\lambda_i=k$. This property simplifies the function f. If this constraint is taken into account then the set $\Lambda(Q_0,n)$ is reduced to a shifted simplex. It is not yet known whether problem 3.8.8 admits an explicit expression as solution or whether one has to resort to numerical minimization.

The hope is that the solution of problem 3.8.5 provides information on the solution of the approximate weak Gaussian stochastic realization problem 3.8.1.

4. SPECIFIC OPEN STOCHASTIC REALIZATION PROBLEMS

The purpose of this section is to present several stochastic systems and processes for which the solution to the stochastic realization problem may be useful for engineering, economics etc. The presentation of these models is brief. The tutorial and survey-like character of this paper may make it useful to mention these models.

Gaussian systems
The approximate weak Gaussian stochastic realization problem, as described in subsection 3.8, is unsolved. For Gaussian systems there are unsolved problems for specific subclasses of systems that may be of interest to specific application areas. Some of these problems and models are described below.

The co-integration and the error correction model. As a model for economic processes that move about an equilibrium, C.W.J. Granger [32] has proposed a model that is known as the *co-integration model.*

The components of a vector valued process $y:\Omega \times \mathbf{Z} \to \mathbf{R}^k$ are said to be *co-integrated of order* 1,1 if

1. after differencing once ($\nabla y(t)=y(t)-y(t-1)$) the resulting process has a stationary invertible AutoRegressive-Moving-Average (ARMA) representation without deterministic component;
2. there exists a vector $\alpha \in \mathbf{R}^k$, $\alpha \neq 0$, such that $z(t)=\alpha^T y(t)$ has again a stationary invertible ARMA representation without deterministic component.

The interpretation of this model is that the economic process that is modelled consists of a trend and stationary fluctations, but is such that a linear combination of the process is stationary. The linear combination should be associated with some difference of economic processes, say income minus consumption. According to the model this difference fluctuates around some equilibrium value and it may be considered as forced towards this equilibrium by economic forces. A generalization of this model has been proposed, see [22]. That paper also reports on the suitability of the co-integration model for economic processes.

A vector valued proces $y:\Omega \times T \to \mathbf{R}^k$ is said to have an *error correction representation,* see [22], if it can be expressed as:

$$A(B)(1-B)y(t) = -\gamma z(t-1) + u(t)$$

in which u is a stationary process representing a disturbance, $A(.)$ is a matrix polynomial with $A(0)=I$, B is the delay operator defined by $By(t)=y(t-1)$, there exists a $\alpha \in \mathbf{R}^k$ such that $z(t)=\alpha^T y(t)$ and $\gamma \in \mathbf{R}^k$, $\gamma \neq 0$.

The interpretation of an error correction model is that the disequilibrium of one period, $z(t-1)$, is used to determine the economic process in the next period.

For recent work on the co-integration and error correction model see a special issue of *Journal of Economic Dynamics and Control* that is opened by the special editor M. Aoki with the paper [8]. In that issue there is another paper by M. Aoki [9] in which he shows that the co-integration model may be obtained from a Gaussian system representation under a condition on the poles of the system. In that approach a co-integration vector is not assumed, nor are assumptions needed on trends or periods.

An approach to the stochastic realization problem for the co-integration model and the error correction model may be based on stochastic realization theory for a particular class of Gaussian systems.

Gaussian systems with inputs. A time-invariant Gaussian system with inputs has a forward representation of the form

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) + Mv(t), \\ y(t) &= Cx(t) + Du(t) + Nv(t), \end{aligned}$$

where $u:\Omega \times T \to \mathbf{R}^m$ is an input process, and $v:\Omega \times T \to \mathbf{R}^k$ is a Gaussian white noise process. Such systems are used in stochastic control. The stochastic realization problem for this class of systems has not yet been treated. It is motivated by stochastic control theory. An unsolved question is whether such a stochastic system is a minimal realization of the measure on the observation processes of output y and input u. The conditions for minimality should be related to the solvability conditions of the linear-quadratic-Gaussian stochastic control problem.

For this class of systems one has also to investigate the stochastic realization problem associated with the solution to the linear-exponential-quadratic-Gaussian stochastic control problem [14, 77]. This solution is related to recent results in H-infinity theory.

The Gaussian factor model
This model and the associated stochastic realization problem are discussed in section 5 of this paper.

Factor systems
These systems and the associated stochastic realization problem are discussed in section 6.

Positive stochastic linear systems
A stochastic system in which the state and observations process take values in the vector space $\mathbb{R}_+$ will be called *a positive stochastic system.* The gamma distribution is an example of a probability distribution on $\mathbb{R}_+$. Such systems may be appropriate stochastic models in economics, biology, and communication systems where the state variables are economic quantities, concentrations etc. Examples from biology may be found in [56]. Several examples of such systems follow.

Portfolio models. A portfolio model is a dynamic model for the growth of assets such as shares, bonds and money in savings accounts. After the fall of share prices in October 1987 there is a renewed interest in portfolio models.

A stochastic portfolio model may be specified by

$$dp(t) = ap(t)dt + p(t)dv(t), p(0),$$

where $p:\Omega \times T \rightarrow \mathbb{R}$ represents the price of the asset, $a \in \mathbb{R}$ represents a growth trend and $v:\Omega \times T \rightarrow \mathbb{R}$ represents random fluctuations. More refined models can be defined to account for control of buying and selling, and for switch-over costs. A realistic portfolio model would require a realistic macro economic model for short-term and long-term economic growth, preferably on an international scale.

The portfolio model should be seen as a special case of a growth model. In addition, growth models that exhibit saturation should be investigated in connection with market saturation effects.

The realization problem for the stochastic portfolio model would have to deal with questions as whether the trends and variances of these models can be determined from observed prices. This problem becomes more interesting if, for example, the price of a share is related to development of the markets in which the company is active, to its management structure, and to long-term growth of the economy.

The Gale model and a Leontieff system. For production planning of firms a model proposed by D. Gale is used. For references on this model see the book by V.I. Arkin and I.V. Evstigneev [10]. The classical Leontieff model is a matrix relation between inputs and outputs of an economic unit. A dynamic version of this model has been proposed, it will be called a *Leontieff system.*

The Gale model is specified by

$$z(t) = \begin{bmatrix} x(t-1) \\ y(t) \end{bmatrix}, \quad x,y:T \rightarrow \mathbb{R}_+^n \tag{4.1}$$

satisfying

$$z(t) \in Q(t), \tag{4.2}$$

$$y(t) \geqslant x(t), \tag{4.3}$$

where $Q(t) \in \mathbb{R}_+^{2n}$ is a convex set. Here $x(t-1)$ is called the *input,* and $y(t)$ the *output* in period $(t-1, t]$, and $z(t)$ the *technological process* at time $t \in T$. Condition (4.2) is a technological feasibility condition; condition (4.3) implies that the input at any time step cannot exceed the output of the previous step. A parametric form of this model is given in subsection 1.1.8 of [10].

There is also a stochastic version of the Gale model, see the subsections 2.4.1 and 2.4.7 of [10].

Optimal control problems for the Gale model are treated in [10]. The results are maximum principles and turnpike theorems.

Finite stochastic systems
In section 3 a finite stochastic system has been defined. It consists of an output process taking values in a finite set and a finite-state Markov process. The stochastic realization problem for this class of systems is then to classify all minimal stochastic systems such that the output process of such a system equals a given process either in distribution or almost surely. The motivation of this problem comes from the use of finite stochastic systems as models for communication or computers systems. For such technical problems, stochastic models with discrete variables arise naturally or are useful approximate models. The stochastic realization problem was formulated in 1957 in a paper by Blackwell and Koopmans [15]. During the 1960's several publications appeared that provide a necessary and sufficient condition for the existence of a finite stochastic realization. For references see [52]. Unsolved questions are the characterization of minimality of the state space and the classification of all minimal stochastic realizations. The main bottleneck is currently the characterization of the minimality of the state space. This question leads to a basic problem for positive linear algebra, that is, linear algebra over $\mathbb{R}_+$.

Counting process systems
An example of a counting process system is a continuous-time stochastic system of which the output process is a counting process with stationary increments and in which the intensity process of the counting process is a finite-state Markov process. The stochastic realization problem for this class of systems is unsolved.

The motivation for this stochastic realization problem comes from the use of counting process models in communication, queueing theory, computer science, and biology. The observation process may often be taken as a counting process with stationary increments.

The above mentioned class of stochastic systems has been investigated in [68, 69]. The question of characterizing the minimal size of the state space is closely related to the same question for the finite stochastic realization problem.

Gaussian random fields
For this class of stochastic objects new mathematical models are needed.

5. FACTOR ANALYSIS
In this section the stochastic realization problem for the Gaussian factor analysis model will be formulated and analyzed.

The factor analysis model was proposed early this century. For references on the factor analysis model see [7, 74]. Factor analysis is used as a quantitative model in sociology and psychology. R. Frisch has suggested the factor analysis model as a way to determine relations among random variables [25]. R.E. Kalman has emphasized this model and formulated the associated stochastic realization problem [39-41]. Since then several researchers have considered the stochastic realization problem for this model class. This problem is still unsolved. Below one finds a problem formulation, questions, partial results and conjectures for this stochastic realization problem. For recent publications on this problem see the special issue of *J. of Econometrics* that is opened by the paper [2].

Problem formulation
From economic data that exhibit variability one may estimate a covariance. Suppose that this data vector may be modelled by a Gaussian random variable. Effectively one is thus given a Gaussian measure, say on $\mathbb{R}^k$. The initial problem may then be stated as: how to represent this measure such that the dependencies between the components of the vector are exhibited? The factor analysis model will be used to describe these dependencies.

DEFINITION 5.1. *A* Gaussian factor analysis model *or a* Gaussian factor model *is defined by the specification*

$$y = Hx + w, \qquad (5.1)$$

or

$$y_i = H_i x + w_i, \ i=1,\ldots,k, \tag{5.2}$$

where $x:\Omega \to \mathbb{R}^n$, $x \in G(0,Q_x)$ *is called the* factor, $w:\Omega \to \mathbb{R}^k$, $w \in G(0,Q_w)$ *is called the* noise, $y:\Omega \to \mathbb{R}^k$, $y \in G(0,Q_y)$ *is called the* observation vector, $H \in \mathbb{R}^{k \times n}$ *is called the* matrix of factor loadings, Q_w *is a diagonal matrix, and* (x,w) *are independent random variables.*

The interpretation of the Gaussian factor analysis model (5.2) is that each component of the observation vector consists of a systematic part $H_i x$ and a noise part w_i. Observe that the condition that Q_w is diagonal is equivalent to the condition that $(w_1, \ldots, w_k)$ are independent random variables. A generalization of the above definition may be given to the case in which Q_w is block diagonal. The Gaussian factor model in rudimentary form goes back to [67]. The Gaussian factor analysis model is equivalent to the *confluence analysis model* introduced by R. Frisch [25]. In this model the representation of the observation vector is specified by

$$y = u + w, \ Au = 0,$$

in which $A \in \mathbb{R}^{(k-n) \times k}$ u,w are independent random variables, and Q_w is a diagonal matrix. For other references on this approach see the publications of O. Reiersøl [58, 59].

The Gaussian factor analysis model, or, equivalently, the confluence analysis model, has been suggested as an alternative to regression analysis. Strong pleas for this approach are the introduction of the book by R. Frisch [25], and the papers of R.E. Kalman [39-41]. Within economic and statistical literature the questions regarding regression and factor models have been recognized, see for example [7, 66, 70, 80].

PROBLEM 5.2. *The* weak stochastic realization problem for a Gaussian factor model *is given a Gaussian measure* $G(0,Q)$ *on* $\mathbb{R}^k$ *to solve the following subproblems.*

a. *Determine a Gaussian factor model, say*

$$y = Hx + w,$$

such that the measure of y equals the given measure or

$$y \in G(0,Q_y) = G(0,Q).$$

If such a Gaussian factor model exists then it is called a weak stochastic realization *of the given measure.*

b. *Determine the minimal dimension* $n^*(Q)$ *of the factor* x *in a weak stochastic realization of the given measure* $G(0,Q)$. *Call a weak stochastic realization* minimal *if the dimension of the factor systems equals* $n^*(Q)$.

c. *Classify all minimal weak stochastic realizations of the given measure.*

Part a. of problem 5.2 is equivalent to: determine $(n, Q_x, Q_w, H) \in \mathbb{N} \times \mathbb{R}^{n \times n} \times \mathbb{R}^{k \times k} \times \mathbb{R}^{k \times n}$ such that

$$Q = HQ_xH^T + Q_w,$$

where $Q_x = Q_x^T \geqslant 0$, $Q_w = Q_w^T \geqslant 0$, and Q_w is diagonal. Part a of the above problem is trivial, the hard parts of the problem are b and c.

Corresponding to problem 5.2 there is a *strong stochastic relization problem for a Gaussian factor model.* In this problem one is given a probability space (Ω, F, P) and a Gaussian distributed random variable $z \in G(0,Q)$. The problem is then to construct a Gaussian factor model

$$y = Hx + w$$

on the given probability space such that

$$z = y \ \ a.s.$$

and to classify all minimal models of this type. This problem has been defined in [54], where a generalization of the Gaussian factor model for Hilbert spaces is introduced. The strong stochastic realization problem will not be discussed in detail here.

What is the main characteristic of the Gaussian factor model? To answer this question one has to introduce the following concept.

DEFINITION 5.3. *The σ-algebra's $F_1, F_2, \ldots, F_m$ are called* conditionally independent *given the σ-algebra G if*

$$E[z_1 \cdots z_m \mid G] = E[z_1 \mid G] \cdots E[z_m \mid G]$$

for all $z_i \in L^+(F_i)$. The notation

$$(F_1, F_2, \ldots, F_m \mid G) \in CI$$

will be used to denote that $F_1, \ldots, F_m$ are conditionally independent given G and CI will be called the multivariate conditional independence relation.

The following elementary result then establishes the relation between the Gaussian factor model and the conditional independence relation.

PROPOSITION 5.4. *Let $y_i : \Omega \to \mathbb{R}$, $i = 1, 2, \ldots, k$, $x : \Omega \to \mathbb{R}^n$. The following statements are equivalent:*
a. *The random variables $(y_1, \ldots, y_k, x)$ are jointly Gaussian with zero mean and satisfy*

$$(F^{y_1}, \ldots, F^{y_k} \mid F^x) \in CI.$$

b. *The random variables y, x satisfy the conditions of the Gaussian factor analysis model of 5.1 with the representation*

$$y = Hx + w.$$

The conditional independence property of a Gaussian factor model is now seen to be its main characteristic. It will be called the *factor property* of a Gaussian factor model. It allows extensions to non-Gaussian random variables. Such extensions have been considered in the literature, see for references [74]. The factor property is a generalization of the concept of state for a stochastic system. In such a system the future of the state and output process on one hand, and the past of the state and output process on the other hand are conditionally independent given the present state. The analogy is such that the state corresponds to the factor and the output process to the observation vector of the factor model. The factor property or the conditional independence property occurs in many mathematical models in widely different application areas.

Below the stochastic realization problem 5.2 will be discussed, first in terms of the external description and then in terms of the internal description.

The stochastic realization problem in terms of the external description.
In this subsection one is assumed to be given a Gaussian measure $G(0, Q_y)$. The weak stochastic realization problem for a Gaussian factor model specializes in this case to the following question.

QUESTION 5.5. *Given a Gaussian measure $G(0, Q_y)$.*
a. *What is the minimal dimension $n^*(Q_y)$ of the factor in a stochastic realization of $G(0, Q)$?*
b. *What is the classification of all minimal stochastic realizations of $G(0, Q)$, or all decompositions of the form*

$$Q_y = Q_1 + Q_w$$

in which $Q_1 = Q_1^T \geqslant 0$, $Q_w = Q_w^T \geqslant 0$ is diagonal and $\mathrm{rank}(Q_1) = n^*(Q_y)$.

NOTATION 5.6.
a. *If* $Q \in \mathbb{R}^{k \times k}$ *then*

$$D(Q) \in \mathbb{R}^{k \times k}$$

is a diagonal matrix with on the diagonal the elements of the diagonal of the matrix Q.
b. *If* $Q \in \mathbb{R}^{k \times k}$ *then the matrix* $OD(Q) \in \mathbb{R}^{k \times k}$, *called the* off-diagonal part of Q, *is defined by*

$$OD(Q)_{ii} = 0, \; OD(Q)_{i,j} = Q_{i,j}, \text{ for all } i,j \in \mathbb{Z}_k, \; i \neq j.$$

c.

$$\mathbb{Q}(Q_y,k,n) = \left\{ \begin{matrix} (Q_1,Q_w) \in \mathbb{R}^{k \times k} \times \mathbb{R}^{k \times k} \mid Q_1 = Q_1^T \geq 0, \; rank(Q_1) = n, \\ Q_w = Q_w^T \geq 0, \; Q_w \text{ diagonal}, \; Q_y = Q_1 + Q_w \end{matrix} \right\}$$

d.

$$n^*(Q_y) = \min\{n \in \mathbb{N} \mid \exists \, (Q_1,Q_w) \in \mathbb{Q}(Q_y,k,n)\}$$

It turns out to be useful to work with a standard form for the variance matrix, a canonical form.

DEFINITION 5.7. *One says that the matrices* $Q_1, Q_2 \in \mathbb{R}^{k \times k}$, *that are assumed to be strictly positive definite, are* equivalent *if there exists a diagonal matrix* $D \in (0,\infty)^{k \times k}$ *such that*

$$Q_1 = DQ_2D.$$

A canonical form with respect to this equivalence relation is then such that $D(Q) = I$. An investigation should be made of another equivalence relation defined as in 5.7 in which negative elements are also admitted on the diagonal.

Question 5.5.a is still unsolved. Characterizations of $n^*(Q_y)$ are known in the two extreme cases of $n^*(Q_y) = 1$ and $n^*(Q_y) = k-1$. These results are stated below. The characterization for $n^*(Q_y) = 1$ may go back to C. Spearman and co-workers. The formulation given here is from [13].

THEOREM 5.8. [13]. *Given* $Q_y \in \mathbb{R}^{k \times k}$, $Q_y = Q_y^T > 0$. *Assume that* $k \geq 4$, $Q_y \in (0,\infty)^{k \times k}$, *and that* Q_y *is irreducible. Then* $n^*(Q_y) = 1$ *iff*

$$\begin{cases} q_{il}q_{jm} - q_{im}q_{jl} = 0, \; q_{il}q_{ji} - q_{ii}q_{jl} \leq 0, \\ \forall \, i,j,l,m \in \mathbb{Z}_k, \; l \neq m, j \neq l, j \neq m, i \neq j, i \neq l, i \neq m. \end{cases}$$

THEOREM 5.9. [13, 39, 58]. *Given* $Q_y \in \mathbb{R}^{k \times k}$, $Q_y = Q_y^T > 0$. *Then* $n^*(Q_y) = k-1$ *iff* Q_y^{-1} *has strictly positive elements, possibly after sign changes of rows and corresponding columns.*

What are the generic values of $n^*(Q_y)$? Below are stated the main results from a study by J.P. Dufour [20] on this question.

DEFINITION 5.10. *Let*

$$\mathbb{S}_k^+ = \{Q \in \mathbb{R}_+^{k \times k} \mid Q = Q^T\}.$$

Note that the condition of positive definiteness is not imposed in the definition of the set $\mathbb{S}_k^+$. In the following the Euclidean topology is used on the vector space $\mathbb{R}^n$.

THEOREM 5.11. [20].
a. *There exists an open and dense subset* $\mathbb{S} \subset \mathbb{S}_k^+$ *such that for all* $Q_y \in \mathbb{S}$

$$n^*(Q_y) \geq \tfrac{1}{2}(2k+1-\sqrt{1+8k}).$$

This inequality is known as the Ledermann bound.

b. *Let $Q \in \mathbb{S}$. For every Q_1 in a sufficiently small neighberhood of Q in $\mathbb{S}$ the relation*

$$n^*(Q) = n^*(Q_1)$$

holds.

c. *For any integer p such that*

$$\tfrac{1}{2}(2k+1-\sqrt{1+8k}) \leq p \leq k-1$$

there exists a $Q \in \mathbb{S}$ such that $n^*(Q)=p$.

By way of illustration there follow characterizations on the value of $n^*(Q_y)$ for variance matrices $Q_y \in \mathbb{R}^{k \times k}$ with several low values of k.

PROPOSITION 5.12. *Let $Q_y \in \mathbb{R}^{3\times 3}$, $Q_y = Q_y^T > 0$, $D(Q_y)=I$.*

a. *$n^*(Q_y)=0$ iff Q_y is diagonal.*

b. *$n^*(Q_y)=1$ iff one of the following cases applies.*

Case 1. If $q_{12}>0$, $q_{13}>0$, $q_{23}>0$ and

$$\frac{q_{12}q_{13}}{q_{23}}, \frac{q_{12}q_{23}}{q_{13}}, \frac{q_{13}q_{23}}{q_{12}} \in [0,1].$$

Case 2. If $q_{12}>0$, $q_{13}=0$, $q_{23}=0$.

Other cases are derived from the above by permutations of signs and indices.

c. *$n^*(Q_y)=2$ iff otherwise.*

For the special case in which $Q_y \in \mathbb{C}^{4\times 4}$ and $n^*(Q_y)=1$ a characterization is given in [6].

PROPOSITION 5.13. *Let $Q_y \in (0,\infty)^{4\times 4}$. Then $n^*(Q_y)=1$ iff, up to a permutation of indices,*

1. $c = \dfrac{q_{12}q_{13}}{q_{23}} = \dfrac{q_{12}q_{14}}{q_{24}} = \dfrac{q_{13}q_{14}}{q_{34}} \in (0,1];$
2. $c \geq q_{12}^2,\ c \geq q_{13}^2,\ c \geq q_{14}^2.$

Classification. In this subsubsection the classification question 5.5.b will be discussed. Thus, given $Q_y \in \mathbb{R}^{k\times k}$, the question is to classify all decompositions of the form

$$Q_y = Q_1 + Q_w$$

in which $rank(Q_1)=n^*(Q_y)$. Geometry seems the appropriate tool for this classification, in particular polyhedral cones and convex analysis. For an approach along these lines see [19]. Below another approach is indicated that combines analysis and geometry.

Remark that in the decomposition

$$Q_y = Q_1 + Q_w = HQ_xH^T + Q_w$$

the off-diagonal elements of Q_1 are equal to the off-diagonal elements of Q_y. Moreover, by convention $D(Q_y)=I$. Hence the set $\mathbb{Q}(Q_y,k,n^*(Q_y))$ may be classified by the diagonal of Q_1.

PROPOSITION 5.14. *Let*

$$\mathbb{D}(Q_y,k,n) = \left\{ D \in \mathbb{R}^{k\times k} \mid D \text{ diagonal},\ -OD(Q_y) \leq D \leq I,\ rank(D+OD(Q_y))=n \right\},$$

$$f:\mathbb{D}(Q_y,k,n)\to\mathbb{Q}(Q_y,k,n),\ f(D) = (D+OD(Q_y),I-D).$$

Then f is a bijection.

Remark that the set $\mathbb{D}(Q_y,k,n)$ without the rank condition is a closed convex set. From 5.14 and some linear algebra one obtains the following result on the classification.

THEOREM 5.15. *Let* $Q_y\in\mathbb{R}^{k\times k}$, $Q_y=Q_y^T>0$, $D(Q_y)=I$,

$$\mathbb{D}_1(Q_y,k,n) = \left\{\begin{array}{c} D_1\in\mathbb{R}^{n\times n} \mid D_1 \text{ diagonal},\ 0<D_1\leqslant I, \\ \exists \text{ permutation matrix } P \text{ such that if } PQ_yP^T=\begin{bmatrix}A & B\\ B^T & C\end{bmatrix}, \\ \text{then } D_1+OD(A)>0,\ D_2:=B^T[D_1+OD(A)]^{-1}B-OD(C) \\ \text{is diagonal and satisfies } 0\leqslant D_2\leqslant I \end{array}\right\},$$

$$g:\mathbb{D}_1(Q_y,k,n)\to\mathbb{Q}(Q_y,k,n),$$

$$g(D_1) = (P^T\begin{bmatrix}D_1+OD(A) & B\\ B^T & D_2+OD(C)\end{bmatrix}P,\ P^T\begin{bmatrix}I-D_1 & 0\\ 0 & I-D_2\end{bmatrix}P).$$

Then:

a. *g is well defined;*

b. *g is surjective;*

c. *The diagonal matrix*

$$\begin{bmatrix}D_1 & 0\\ 0 & D_2\end{bmatrix}$$

is unique up to a permutation.

The proof of the above theorem is elementary with the aid of the following lemma.

LEMMA 5.16. *Let* $k,n\in\mathbb{Z}_+$, $k>n$, $A\in\mathbb{R}^{n\times n}$, $B\in\mathbb{R}^{n\times(k-n)}$, $C\in\mathbb{R}^{(k-n)\times(k-n)}$, $A=A^T$, $C=C^T$, $rank(A)=n$,

$$Q = \begin{bmatrix}A & B\\ B^T & C\end{bmatrix}\in\mathbb{R}^{k\times k},\ T = \begin{bmatrix}A^{-1/2} & -A^{-1}B\\ 0 & I\end{bmatrix}\in\mathbb{R}^{k\times k}.$$

a. *Then*

$$T^TQT = \begin{bmatrix}I & 0\\ 0 & C-B^TA^{-1}B\end{bmatrix}.$$

b. $rank(T)=k$.

c. $rank(Q)=n$ *iff* $C-B^TA^{-1}B=0$.

d. $Q\geqslant 0$ *iff* $C-B^TA^{-1}B\geqslant 0$.

The study of the classification along the lines sketched above must proceed by an investigation of the following relations for the diagonal matrix $D_1\in\mathbb{R}^{n\times n}$:

$$D_1+OD(A)>0,$$

$$D_2 := B^T[D_1+OD(A)]^{-1}B-OD(C),\ 0\leqslant D_2\leqslant I,\ D_2 \text{ is diagonal.}$$

For the cases $n^*(Q_y)=k-1$ and $n^*(Q_y)=1$ theorem 5.15 directly yields explicit classifications. The classifications of three low-dimensional examples are listed.

PROPOSITION 5.17. *For the case $k=2$, $Q_y \in \mathbb{R}^{2\times 2}$, $n^*(Q_y)=1$ with*

$$Q_y = \begin{bmatrix} 1 & q \\ q & 1 \end{bmatrix}, \quad q \neq 0,$$

the classification, in the notation of 5.15, is given by

$$\mathbb{D}(Q_y,2,1) = \left\{ d_1 \in \mathbb{R}_+ \mid q^2 \leqslant d_1 \leqslant 1 \right\}$$

and

$$g(d_1) = (\begin{bmatrix} d_1 & q \\ q & q^2/d_1 \end{bmatrix}, \begin{bmatrix} 1-d_1 & 0 \\ 0 & 1-q^2/d_1 \end{bmatrix}).$$

PROPOSITION 5.18. *For the case $k=3$, $Q_y \in (0,\infty)^{3\times 3}$, and $n^*(Q_y)=2$ the classification according to 5.15 is given by*

$$\mathbb{D}_1(Q_y,3,2) = \left\{ \begin{array}{c} \begin{bmatrix} d_1 & o \\ 0 & d_2 \end{bmatrix} \in \mathbb{R}^{2\times 2} \mid d_1,d_2 \in [0,1],\ d_1 d_2 - q_{12}^2 \neq 0, \\ \dfrac{d_1 q_{23}^2 + d_2 q_{13}^2 - 2 q_{12} q_{13} q_{23}}{d_1 d_2 - q_{12}^2} \in [0,1] \\ \text{and conditions obtained by permutation of indices} \end{array} \right\}.$$

PROPOSITION 5.19. *For the case $k=3$, $Q_y \in (0,\infty)^{3\times 3}$, $n^*(Q_y)=1$, the decomposition is unique with*

$$Q_1 = \begin{bmatrix} \dfrac{q_{12}q_{13}}{q_{23}} & q_{12} & q_{13} \\ q_{12} & \dfrac{q_{12}q_{23}}{q_{13}} & q_{23} \\ q_{13} & q_{23} & \dfrac{q_{13}q_{23}}{q_{12}} \end{bmatrix}.$$

PROPOSITION 5.20. *For the case $k=4$, $Q_y \in (0,\infty)^{4\times 4}$, $n^*(Q_y)=2$ the classification according to 5.15 is given by*

$$\mathbb{D}_1(Q_y,2) = \left\{ \begin{array}{c} \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix} \mid d_1,d_2 \in [0,1],\ d_1 d_2 - q_{12}^2 \neq 0, \\ q_{34} = d_2 q_{13} q_{14} + d_1 q_{23} q_{24} - q_{12} q_{14} q_{23} - q_{12} q_{13} q_{24}, \\ d_2 q_{13}^2 + d_1 q_{23}^2 - 2 q_{12} q_{13} q_{23} \in [0,1], \\ d_1 q_{24}^2 + d_2 q_{14}^2 - 2 q_{12} q_{14} q_{24} \in [0,1], \\ \text{and conditions obtained by permuting the indices} \end{array} \right\}.$$

The stochastic realization problem in terms of the internal description
The specification of the Gaussian factor model as given in 5.1 will be called the *internal description.* It is called internal because the specification is in terms of the matrices (H, Q_x, Q_w) rather than in terms of Q_y. The questions for the internal desciption require one definition.

DEFINITION 5.21. *The Gaussian factor model with representation*

$$y = Hx + w$$

is called minimal *if* $n = n^*(Q_y)$ *in which* $x: \Omega \to \mathbf{R}^n$, $Q_x > 0$ *and*

$$Q_y = HQ_xH^T + Q_w.$$

Introduce the convention $Q_x = I$. The weak stochastic realization problem for a Gaussian factor model specializes in this case to the following question.

QUESTION 5.22.
a. *Which conditions on the matrices* (H, Q_x, Q_w) *are equivalent with minimality of the Gaussian factor model?*
b. *How are two minimal Gaussian factor models related?*

The above questions are still open. The minimality question 5.22.a seems most interesting because its answer will involve a new system theoretic concept like stochastic observability. To hint at what may be needed a special case is considered.

Consider a special Gaussian factor model of the form

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix} x + w$$

in which the variance Q_w is required to be block-diagonal, in particular it consists of two blocks only

$$Q_w = \begin{bmatrix} Q_{w_1} & 0 \\ 0 & Q_{w_2} \end{bmatrix}.$$

One says that this Gaussian factor model is *stochastically observable* if the map

$$x \mapsto E[\exp(iu^T y_1) \,|\, F^x]$$

is injective on the support of x. Similarly one says that the Gaussian factor model is *stochastically reconstructible* if the map

$$x \mapsto E[\exp(iu^T y_2) \,|\, F^x]$$

is injective on the support of x. It may then be proven that the Gaussian factor model is minimal iff it is stochastically observable and stochastically reconstructible iff $rank(H_1) = n = rank(H_2)$ [71].

Let's return to question 5.22.a, when is a Gaussian factor model minimal in case Q_w is restricted to be diagonal. The following conjecture comes to mind first: A Gaussian factor model is minimal iff the map

$$x \mapsto E[\exp(iuy_i) \,|\, F^x], \ \textit{for } i \in \mathbf{Z}_k,$$

is injective on the support of x for all $i \in \mathbf{Z}_k$. This conjecture is false, because the effective dimension n of x may be larger than 1. Even if $n = 1$ it is false, see 5.23 below. The special case of $k = 3$ and $n = 2$ mentioned in 5.24 shows that the equivalent condition for minimality of a Gaussian factor system needs more thinking. The minimality characterizations for the following special cases may be helpful in formulating conjectures for the general result.

PROPOSITION 5.23. *Consider a Gaussian factor model*

$$y = hx + w$$

with $k \geqslant 2$, $n=1$, $h \in \mathbf{R}^k$. *Then this model is minimal iff*

$$\exists\ i,j \in \mathbf{Z}_k,\ i \neq j, \text{ such that } h_i \neq 0 \text{ and } h_j \neq 0.$$

PROOF. The Gaussian factor model with $n=1$ is minimal iff the dimension of the factor cannot be reduced. This is true iff $n^* > 0$ or iff Q_y is non-diagonal. Note that $OD(Q_y) = OD(hh^T)$. □

PROPOSITION 5.24. *Consider the Gaussian factor model of 5.1 with* $k=3$, $n=2$,

$$H = \begin{bmatrix} h_1^T \\ h_2^T \\ h_3^T \end{bmatrix} \in \mathbf{R}^{3 \times 2},\ D(Q_y) = I.$$

Assume that $h_1^T h_2 > 0$, $h_1^T h_3 > 0$, $h_2^T h_3 > 0$. *Then this Gaussian factor model is minimal iff one of the following conditions is satisfied:*

1. $\dfrac{(h_1^T h_2)(h_1^T h_3)}{(h_2^T h_3)} \notin [0,1]$,
2. $\dfrac{(h_1^T h_2)(h_2^T h_3)}{(h_1^T h_3)} \notin [0,1]$,
3. $\dfrac{(h_1^T h_3)(h_2^T h_3)}{(h_1^T h_2)} \notin [0,1]$.

PROOF. This follows from 5.12. □

Classification of internal description

The motivating question here is whether the internal description of a Gaussian factor model is uniquely determined by the variance of the observation vector. In general such a model is not unique. This question is related to question 5.5.b. For the classification of the internal description of factor analysis models with block-diagonal structure see [53]. To structure the discussion a definition is introduced.

DEFINITION 5.25. *Two Gaussian factor models*

$$y = Hx + w$$

and

$$\overline{y} = \overline{H}\overline{x} + \overline{w}$$

are called equivalent *if*

$$HQ_xH^T + Q_w = \overline{H}Q_{\overline{x}}\overline{H}^T + \overline{Q}_w.$$

Note that the two Gaussian factor models of 5.25 that are defined to be equivalent both have the same variance matrix Q_y, since

$$Q_y = H_1 Q_{x_1} H_1^T + Q_{w_1} = H_2 Q_{x_2} H_2^T + Q_{w_2}.$$

Therefore they cannot be distinguished given Q_y. It is well-known that if (n, H, Q_x, Q_w) are the parameters of a Gaussian factor system and if $S \in \mathbb{R}^{n \times n}$ is an orthogonal matrix ($SS^T = I$), the two Gaussian factor models specified by (n, H, Q_x, Q_w) and $(n, HS, S^T Q_x S, Q_w)$ are equivalent. However, there may be other ways in which two Gaussian factor models are equivalent.

In applications of Gaussian factor analysis it has been recognized that there may be many equivalent models. To reduce the class of equivalent models practitioners fix certain elements of the matrix of factor loadings, based on prior knowledge about the observation vector or arbitrarily.

The question now is, given a Gaussian factor model, to describe the equivalence class of all Gaussian factor models that are equivalent with the given one. This question is still open.

6. Gaussian factor systems

The purpose of this section is to formulate the concept of a Gaussian factor system and to survey the preliminary results of the stochastic realization problem for this class of systems.

A motivation for the study of this class of systems is the stochastic realization problem for Gaussian systems with inputs. One would like to know whether it is possible to determine from an observed vector-valued process which components are inputs and which are outputs of a Gaussian system. Another motivation for the study of this class of systems is the exploration of the extension of Gaussian factor models to dynamic systems.

Definition 6.1. *A* Gaussian factor system, *in discrete time, is an object specified by the equations*

$$x(t+1) = Ax(t) + Bu(t),$$
$$y(t) = [Cx(t) + Du(t)] + w(t)$$

or

$$y(t) = \sum_{s \in T} H(t-s)u(s) + w(t)$$

where $u : \Omega \times T \to \mathbb{R}^p$ *is a stationary Gaussian process called the* factor process, $w : \Omega \times T \to \mathbb{R}^k$ *is a stationary Gaussian process called the* noise process, $y : \Omega \times T \to \mathbb{R}^k$ *is called the* observed process, $u, w_1, \ldots, w_k$ *are independent processes, the spectral densities of* $u, w_1, \ldots, w_k$ *are rational functions, and the Fourier transform of the transfer function* H *is rational and causal.*

A Gaussian factor system is said to have the factor property if the processes $u, w_1, \ldots, w_k$ are independent processes. This condition can also be rephrased in terms of conditional independence but this will not be done here. Note that the processes $w_1, \ldots, w_k$ need not be white noise processes.

Concepts similar to that of a Gaussian factor system have been introduced in the literature. An elementary version of a Gaussian factor system with H a constant matrix is introduced in [58]. In [26] a Gaussian factor system is defined without the rationality and causality conditions. In [21] one can find the definition 6.1 and a generalization. In [54] a generalization of 6.1 is presented in which the spectral density of the process w is not diagonal but block-diagonal and in which the transfer function H not be causal. The term *dynamic errors-in-variables systems* is used instead of Gaussian factor system in the publications of B.D.O. Anderson and M. Deistler [4-6, 16, 17]. An interpretation of this term follows.

Consider a deterministic finite-dimensional linear system in impulse response representation

$$\hat{y}(t) = \sum_{s \in T} H(t-s)\hat{u}(s).$$

Suppose that the variables of input $\hat{u}$ and output $\hat{y}$ of this system are observed with errors or noise, say by

$$u(t) = \hat{u}(t) + w_1(t), \; y(t) = \hat{y}(t) + w_2(t),$$

in which w_1, w_2 are independent Gaussian white noise processes. Combining these expressions one obtains

$$\begin{bmatrix} u(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} I\delta(t-s) \\ \sum\limits_{s \in T} H(t-s) \end{bmatrix} u(s) + \begin{bmatrix} w_1(t) \\ w_2(t) \end{bmatrix},$$

which is a Gaussian factor system except for the fact that the spectral density of the noise is not a diagonal function but block-diagonal with two blocks. The interpretation of the above defined system of which the variables are observed with error, illustrates the term errors-in-variables model.

PROBLEM 6.2. *The* weak stochastic realization problem for a Gaussian factor system *is to solve the following subproblems. Assume given a stationary Gaussian process with zero mean function and covariance function* Q *or spectral density* $\hat{Q}$.

a. *Find conditions under which there exists a Gaussian factor system*

$$y(t) = \sum_{s \in T} H(t-s)u(s) + w(t)$$

such that the spectral density of y *equals the given spectral density, or*

$$\hat{Q} = \hat{Q}_y = \hat{H}\hat{Q}_u\overline{\hat{H}}^T + \hat{Q}_w.$$

If such a Gaussian factor system exists then it is called a weak stochastic realization *of the given process.*

b. *Classify all minimal weak stochastic realizations of the given process. A weak stochastic realization is called* minimal *if* $rank(\hat{H}\hat{Q}_u\overline{\hat{H}}^T)$ *is minimal.*

A difficulty with the above defined problem is the definition of minimality. In addition to the concept defined in 6.2, which is minimality of the dimension of the factor process u, one could also consider minimality of the degree of $\hat{H}\hat{Q}_u\overline{\hat{H}}^T$. From a viewpoint of linear system theory the latter concept would be preferable. Possibly a mixture of both the dimension of the factor process and the degree has to be considered. Because of this difficulty the author of this paper is not yet convinced that a Gaussian factor system is a suitable model for economic and engineering practice. However, what may be of interest is the special case in which the spectral density of the noise is block-diagonal with two blocks.

The weak stochastic realization problem for Gaussian factor systems is unsolved. Only for low-dimensional cases have results been published. For the case of an observed process with two components see [4, 18, 33] and for the case with three components see [6, 18]. A discussion of the problem may be found in [17]. Questions of identifiability and problems of parameter estimation for Gaussian factor systems have been discussed in [21, 26].

A strong version of the weak stochastic realization problem of 6.2 has been proposed in [54]; see also [55]. The case in which the spectral density $\hat{Q}_w$ of the noise consists of two diagonal blocks has been treated there.

ACKNOWLEDGEMENTS

The author acknowledges J.C. Willems for his inspiring conceptual approach to system and control theory. For the material on factor analysis and factor systems the author has benefited from discussions with L. Baratchart, M. Deistler, R.E. Kalman, and G. Picci.

REFERENCES

1. N.L. AGGARWAL (1974). Sur l'information de Fisher, in *Théories de l'Information*, 111-117, Springer-Verlag, Berlin.
2. D.J. AIGNER and M. DEISTLER (1989). Latent variables models - Editor's introduction, *J. Econometrics*, 41, 1-3.
3. H. AKAIKE (1976). Canonical correlation analysis of time series and the use of an information criterion, in *System identification - Advances and case studies*, 27-96, ed. R.K. Mehra, D.G. Lainiotis, Academic Press, New York.
4. B.D.O. ANDERSON (1985). Identification of scalar errors-in-variables models with dynamics,

Automatica J.-IFAC, 21, 709-716.
5. B.D.O. ANDERSON and M. DEISTLER (1984). Identifiability in dynamic errors-in-variables models, *J. Time Series Anal.*, 5, 1-13.
6. B.D.O. ANDERSON and M. DEISTLER (1987). Dynamic errors-in-variables systems with three variables, *Automatica J.-IFAC*, 23, 611-616.
7. T.W. ANDERSON and H. RUBIN (1956). Statistical inference in factor analysis, in *Proc. Third Berkeley Symposium on Mathematical Statistics and Probability, Volume V*, 111-150, ed. J. Neyman, University of California Press, Berkeley.
8. M. AOKI (1988). Nonstationarity, cointegration, and error correction in economic modeling, *J. Econ. Dyn. Control*, 12, 199-201.
9. M. AOKI (1988). On alternative state space representations of time series models, *J. Econ. Dyn. Control*, 12, 595-607.
10. V.I. ARKIN and I.V. EVSTIGNEEV (1987). *Stochastic models of control and economic dynamics*, Academic Press, New York.
11. K.S. ARUN, D.V. BHASKAR RAO, and S.Y. KUNG (1983). A new predictive efficiency criterion for approximate stochastic realization, in *1983 Conference on Decision and Control*, 1353-1355, IEEE.
12. K.S. ARUN and S.Y. KUNG (1984). State space modeling and approximate realization methods for ARMA spectral estimation, in *IEEE International Conference on Systems, Man and Cybernetics*, IEEE.
13. P.A. BEKKER and J. DE LEEUW (1987). The rank of reduced dispersion matrices, *Psychometrika*, 52, 125-135.
14. A. BENSOUSSAN and J.H. VAN SCHUPPEN (1985). Optimal control of partially observable stochastic systems with an exponential-of-integral performance index, *SIAM J. Control Optim.*, 23, 599-613.
15. D. BLACKWELL and L. KOOPMANS (1957). On the identifiability problem for functions of finite Markov chains, *Ann. Math. Statist.*, 28, 1011-1015.
16. M. DEISTLER (1986). Linear errors-in-variables models, in *Time series and linear systems*, 37-67, ed. S. Bittanti, Springer-Verlag, Berlin.
17. M. DEISTLER and B.D.O. ANDERSON (1988). Linear dynamic errors in variables models: Some structure theory, in *Analysis and optimization of systems*, 873-883, ed. A. Bensoussan, J.L. Lions, Springer-Verlag, Berlin.
18. M. DEISTLER and B.D.O. ANDERSON (1989). Linear dynamic errors-in-variables models - Some structure theory, *J. Econometrics*, 41, 39-63.
19. B. DE MOOR (1988). *Mathematical concepts and techniques for modelling of static and dynamic systems*, Thesis, Katholieke Universiteit Leuven, Leuven.
20. J.P. DUFOUR (1983). *Resultats generiques en analyse factorielle*, Departement de Mathématiques, U.S.T.L., Montpellier.
21. R. ENGLE and M. WATSON (1981). A one-factor multivariate time series model of metropolitan wage rates, *J. Amer. Statist. Assoc.*, 76, 774-781.
22. R.F. ENGLE and C.W.J. GRANGER (1987). Co-integration and error correction: Representation estimation, and testing, *Econometrica*, 55, 251-276.
23. P. FAURRE (1976). Stochastic realization algorithms, in *System Identification - Advances and Case Studies*, 1-25, ed. R.K. Mehra, D.G. Lainiotis, Academic Press, New York.
24. P. FAURRE, M. CLERGET, and F. GERMAIN (1979). *Opérateurs rationnels positifs*, Dunod, Paris.
25. R. FRISCH (1934). *Statistical confluence analysis by means of complete regression systems*, University of Oslo Economic Institute, Oslo.
26. J.F. GEWEKE and K.J. SINGLETON (1981). Maximum likelihood 'confirmatory' factor analysis of econometric time series, *Int. Economic Rev.*, 22, 37-54.
27. R.D. GITTENS (1985). *Canonical analysis - A review with applications in ecology*, Springer-Verlag, Berlin.
28. K. GLOVER and E. JONCKHEERE (1986). A comparison of two Hankel-norm methods for approximating spectra, in *Modelling, Identification and Robust Control*, 297-306, ed. C.I. Byrnes, A. Lindquist,

Elsevier Science Publishers B.V. (North-Holland), Amsterdam.
29. A. GOMBANI and M. PAVON (1985). On the Hankel-norm approximation of linear stochastic systems, *Systems & Control Lett.*, 5, 283-288.
30. A. GOMBANI and M. PAVON (1986). On approximate recursive prediction of stationary stochastic processes, *Stochastics*, 17, 125-143.
31. A. GOMBANI, M. PAVON, and B. COPPO (1986). On Hankel-norm approximation of stationary increment processes, in *Modelling, Identification and Robust Control*, 307-323, ed. C.I. Byrnes, A. Lindquist, Elsevier Science Publishers B.V., Amsterdam.
32. C.W.J. GRANGER (1981). Some properties of time series data and their use in econometric model specification, *J. Econometrics*, 121-130.
33. M. GREEN and B.D.O. ANDERSON (1986). Identification of multivariable errors in variable models with dynamics, *IEEE Trans. Automatic Control*, 31, 467-471.
34. E.J. HANNAN and M. DEISTLER (1988). *The statistical theory of linear systems*, John Wiley & Sons, New York.
35. C. HEIJ (1988). *Deterministic identification of dynamical systems*, Thesis, University of Groningen, Groningen.
36. N.P. JEWELL and P. BLOOMFIELD (1983). Canonical correlations of past and future for time series: Definitions and theory, *Ann. Statist.*, 11, 837-847.
37. N.P. JEWELL, P. BLOOMFIELD, and F.C. BARTMANN (1983). Canonical correlations of past and future for time series: Bounds and computation, *Ann. Statist.*, 11, 848-855.
38. E.A. JONCKHEERE and J.W. HELTON (1985). Power spectrum reduction by optimal Hankel norm approximation of the phase of the outer spectral factor, *IEEE Trans. Automatic Control*, 30, 1192-1201.
39. R.E. KALMAN (1982). System identification from noisy data, in *Dynamical Systems II*, 135-164, ed. A.R. Bednarek, L, Cesari, Academic Press, New York.
40. R.E. KALMAN (1982). Identification from real data, in *Current developments in the interface: Economics, Econometrics, Mathematics*, 161-196, ed. M. Hazewinkel, A.H.G. Rinnooy Kan, D. Reidel Publishing Company, Dordrecht.
41. R.E. KALMAN (1983). Identifiability and modeling in econometrics, in *Developments in Statistics*, 97-136, ed. P.R. Krishnaiah, Academic Press, New York.
42. R.E. KALMAN, P.L.FALB, and M.A. ARBIB (1969). *Topics in mathematical systems theory*, McGraw-Hill Book Co., New York.
43. D. KAZAKOS and P. PAPANTONI-KAZAKOS (1980). Spectral distance measures between Gaussian processes, *IEEE Trans. Automatic Control*, 25, 950-959.
44. S.Y. KUNG and K.S. ARUN (1983). Approximate realization methods for ARMA spectral estimation, in *1983 IEEE International Symposium on Circuits and Systems*, 105-109, IEEE.
45. W.E. LARIMORE, S. MAHMOUD, and R.K. MEHRA (1984). Multivariate adaptive model algorithmic control, in *Proceedings 23rd Conference on Decision and Control*, 675-680, IEEE.
46. W.E. LARRIMORE (1983). System identification, reduced-order filtering and modeling via canonical variate analysis, in *Proceedings 1983 American Control Conference*, 445-451.
47. A. LINDQUIST and G. PICCI (1985). Realization theory for multivariate stationary Gaussian processes, *SIAM J. Control Optim.*, 23, 809-857.
48. A. LINDQUIST, G. PICCI, and G. RUCKEBUSCH (1979). On minimal splitting subspaces and Markovian representations, *Math. Systems Th.*, 12, 271-279.
49. M. PAVON (1984). Canonical correlations of past inputs and future outputs for linear stochastic systems, *Syst. Control Lett.*, 4, 209-215.
50. A. PAZ (1971). *Introduction to probabilistic automata*, Academic Press, New York.
51. V.V. PELLAR and S.V. KHRUSHCHEV (1982). Hankel operators, best approximations, and stationary Gaussian processes, *Russian Math. Surveys*, 37, 61-144.
52. G. PICCI (1978). On the internal structure of finite-state stochastic processes, in *Proc. of a U.S.-Italy Seminar*, 288-304, Springer-Verlag, Berlin.

53. G. PICCI (1989). Parametrization of factor analysis models, *J. Econometrics*, 41, 17-38.
54. G. PICCI and S. PINZONI (1986). Factor analysis models for stationary stochastic processes, in *Analysis and Optimization of Systems*, 412-424, ed. A. Bensoussan, J.L. Lions, Springer-Verlag, Berlin.
55. G. PICCI and S. PINZONI (1986). Dynamic factor-analysis models for stationary processes, *IMA J. Math. Control and Info.*, 3, 185-210.
56. P. PURDUE (1979). Stochastic compartmental models: A review of the mathematical theory with ecological applications, in *Compartmental analysis of ecosystem models*, 223-260, ed. J.H. Matis, B.C. Patten, G.C. White, International Co-operative Publishing House, Fairland, MD.
57. C.R. RAO (1964). The use and interpretation of principal component analysis in applied research, *Sankhya, Series A*, 26, 329-358.
58. O. REIERSØL (1941). Confluence analysis by means of lag moments and other methods of confluence analysis, *Econometrica*, 9, 1-24.
59. O. REIERSØL (1945). Confluence analysis by means of instrumental sets of variables, *Arkiv für Mathematik, Astronomi och Fysik*, 32A.
60. J. RISSANEN (1978). Modeling by shortest data description, *Automatica J. IFAC*, 14, 465- 471.
61. J. RISSANEN (1983). A universal prior for integers and estimation by minimum description length, *Ann. Statist.*, 11, 416-431.
62. J. RISSANEN (1985). *Modeling by the minimum description lenghth principle - A survey,* Preprint.
63. J. RISSANEN (1986). Stochastic complexity and modeling, *Ann. Statist.*, 14, 1080-1100.
64. J. RISSANEN (1986). Predictive and nonpredictive minimum description length principles, in *Time series and linear systems*, 115-140, ed. S. Bittanti, Springer-Verlag, Berlin.
65. J. RUCKEBUSCH (1978). *Sur l'approximation rationnelle des filtres*, Centre de Mathématiques Appliquées, Ecole Polytechnique, Palaiseau.
66. P.A. SAMUELSON (1979). A note on alternative regressions, in *The collected scientific papers of Paul A. Samuelson, Fifth Printing*, 694-697, ed. J.E. Stiglitz, M.I.T. Press, Cambridge.
67. C.A. SPEARMAN (1904). General intelligence, objectively determined and measured, *Amer. J. Psych.*, 15, 201-293.
68. P.J.C. SPREIJ (1986). *Selfexciting counting process systems with finite state space,* Report OS-R8613, Centre for Mathematics and Computer Science, Amsterdam.
69. P.J.C. SPREIJ (1987). *Counting process systems - Identification and stochastic realization,* Doctoral Thesis, University of Twente, Enschede.
70. J.H. STEIGER (1979). Factor indeterminacy in the 1930's and the 1970's some interesting parallels, *Psychometrika*, 44, 157-167.
71. C. VAN PUTTEN and J.H. VAN SCHUPPEN (1983). The weak and strong Gaussian probabilistic realization problem, *J. Multivariate Anal.*, 13, 118-137.
72. J.H. VAN SCHUPPEN (1979). Stochastic filtering theory: A discussion of concepts, methods and results, in *Stochastic control theory and stochastic differential systems*, 209-226, ed. M. Kohlmann, W. Vogel, Springer-Verlag, Berlin.
73. J.H. VAN SCHUPPEN (1982). The strong finite stochastic realization problem - Preliminary results, in *Analysis and optimization of systems*, 179-190, ed. A. Bensoussan, J.L. Lions, Springer-Verlag, Berlin.
74. J.H. VAN SCHUPPEN (1986). Stochastic realization problems motivated by econometric modeling, in *Modelling, Identification and Robust Control*, 259-275, ed. C.I. Byrnes, A. Lindquist, Elsevier Science Publishers B.V. (North-Holland), Amsterdam.
75. E. VERRIEST (1986). Projection Techniques for model reduction, in *Modelling, Identification and Robust Control*, 381-396, ed. C.I. Byrnes, A. Lindquist, Elsevier Science Publishers B.V. (North-Holland), Amsterdam.
76. J.V. WHITE (1983). Stochastic state-space models from empirical data, in *Proceedings 1983 Conference on Acoustics, Speechs and Signal Processing*, 243-246, IEEE.
77. P. WHITTLE (1981). Risk-sensitive linear/quadratic/Gaussian control, *Adv. Appl. Prob.*, 13, 764-

777.
78. J.C. WILLEMS (1986). From time series to linear systems - Part I. Finite dimensional linear time invariant systems, *Automatica J. IFAC*, 22, 561-580.
79. J.C. WILLEMS (1987). From time series to linear systems - Part III. Approximate modelling, *Automatica J. IFAC*, 23, 87-115.
80. E.B. WILSON (1929). Review of 'Crossroads in the mind of man: A study of differentiable mental abilities' by T.L. Kelley, *J. Gen. Psychology*, 2, 153-169.