stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE BESLISKUNDE BW 108/80 MAART
(DEPARTMENT OF OPERATIONS RESEARCH)

J.C. WILLEMS & J.H. VAN SCHUPPEN

STOCHASTIC SYSTEMS AND THE PROBLEM OF STATE SPACE REALIZATION

Preprint

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O).

AMS(MOS) subject classification scheme (1979): 93E03, 60G15, 60A10

First printing 1979
Second printing 1980

Stochastic Systems and the Problem of State Space Realization*)

by

J.C. Willems**) & J.H. van Schuppen

ABSTRACT

The purpose of this paper is to give an exposition of an approach to the problem of stochastic realization theory. We will introduce this problem through the concept of splitting relations and splitting random variables and show in detail how one can construct all minimal splitters for gaussian random vectors. With these ideas in mind, we then introduce the relevant definitions of (autonomous) stochastic dynamical systems and the problem of stochastic realization theory and of white noise representation as they arise naturally in this context. The case of gaussian random processes is then worked out in detail.

KEY WORDS & PHRASES: *Stochastic realization theory, Dynamical systems, Stochastic processes, Conditional independence.*

---

*) For presentation at the NATO Advanced Study Institute and AMS Summer Seminar in Applied Mathematics on *Algebraic and Geometric Methods in Linear System Theory,* Harvard University, Cambridge, Mass., USA, June 18-29, 1979.

**) Advisor, Mathematical Centre. *Permanent address:* Mathematics Institute, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands.

This report will be submitted for publication elsewhere

1. INTRODUCTION

It may be argued that from the theoretical (and certainly from the pedagogical) point of view one of the most outstanding contributions of mathematical system theory has been the axiomatization of the concept of an abstract dynamical system and of the concept of state, in the context of systems with inputs and outputs. This has not only provided a long overdue generalization of the autonomous case, which has been studied in great detail in topological dynamics, but it also gives a very nice and useful axiomatic framework for the study of many problems in control theory, (recursive) signal processing, digital computation and automata theory, etc.

One of the outstanding and completely new problems which has arisen quite naturally in this framework is the so-called *problem of state space realization*. This concerns the question of representation of an input/output map as a system in state space form. For deterministic systems it may be argued that the theory (far from being closed) is quite advanced and completely worked out, both on the abstract and on the algorithmic level, for linear time-invariant systems. However, this is not the case for stochastic systems where other than some partial results for finite state stochastic automata and a fairly complete theory for finite dimensional gaussian processes very little research has been done on these problems. In fact, a conceptual framework in which to treat these questions is still very much absent.

In the present paper we will attempt to give a systematic exposition of some of the main probems and results in this area. In view of the space limitation, we are unable to include proofs. These are either well-documented in the literature, or will appear elsewhere. We have concentrated our efforts for a great deal on some original aspects which involve introducing these problems via deterministic relations and splitting random variables and giving some general definitions of the abstract notion of a stochastic dynamical system and introducing the realization problem from this point of view. We will also give a rather complete description of the situation with gaussian processes. Unfortunately, due to space limitations we have not been able to include a review of some recent results on finite state processes (see [1] for a good exposition and [2,3] for some more recent results), nor have we been able to cover some recent work on a σ-algebraic approach to these problems [4].

Especially the realization of gaussian processes has been given a great deal of attention in the recent system theory literature. Although DOOB [5] already posed some questions in this direction, it is particularly since the work of KALMAN [6] that one has seen some significant progress in this area. Particularly important in this development has been the work of FAURRE [7,8,9] and of ANDERSON [10] who basically solved what we will call the 'weak' or 'measure theoretic' version of this problem and also showed its relation to the classical problem of spectral factorization. There has been some recent progress in this area through the work of LINDQUIST and PICCI [11,12,13,14,15] and of RUCKEBUSCH [16,17,18,19] which has culminated in a solution of what we will call the 'strong' or the 'output-induced' stochastic realization problem. In addition, the neat geometric approach followed by these authors has provided a very important and useful approach to this problem. In closing this introduction, we would like to point out some related work by AKAIKE [20] which fits rather well in the approach which we take in our paper. For a discussion of the relevance of stochastic realization theory in Kalman filtering, see [9 and 21].

ACKNOWLEDGEMENT. *The authors would like to acknowledge the help of Cees van Putten.*

2. SPLITTING RELATIONS AND SPLITTING RANDOM VARIABLES

It turns out that many of the problems which one encounters in stochastic realization theory are already apparent in the seemingly trivial case of a time index set $T = \{1,2\}$. The idea of 'state' then becomes that of a splitting random variable and the basic problem is to find an efficient way for constructing splitting random variables. Moreover, as far as we are aware of, it is not known that these concepts and problems also have very natural analogues in the context of deterministic systems. We will introduce the problems from this point of view since we believe that it gives one a very clear and sharp introduction to this area.

2.1. Relations

A relation is simply a subset of a product set. In our context it is best to think of the two components of this product set as representing the 'past' and the 'future' of some dynamical phenomenon. In a splitting relation one should think of the splitting variable as the state which contains the information in the past which is relevant for the future (and vice-versa!). With this intuitive picture in mind, we now proceed with the formal development:

A *relation* R on the product space $Z := Z_1 \times Z_2 \times \ldots \times Z_n$ is simply a subset $R \subset Z$. The subset of Z_i defined by

$$\{z_i \in Z_i \mid \exists z_1,\ldots,z_{i-1},z_{i+1},\ldots,z_n \text{ such that } (z_1,\ldots,z_{i-1},z_i,z_{i+1},\ldots,z_n) \in R\}$$

is called the *projection* of R on Z_i and will be denoted by $P_{Z_i}R$. The relation on $Z_1 \times \ldots \times Z_{i-1} \times Z_{i+1} \times \ldots \times Z_n$ defined by

$$R_{\{z_i=a_i\}} := \{(z_1,\ldots,z_{i-1},z_{i+1},\ldots,z_n) \mid (z_1,\ldots,z_{i+1},a_i,z_{i+1},\ldots,z_n) \in R\}$$

is called the *relation* R *conditioned by* $\{z_i=a_i\}$. Obvious generalization to the case $z_i \in A_i \subset Z_i$, or conditioning to more than one of the z_i's presents no difficulties. The following notions are the deterministic analogues of 'white noise' and of a 'Markov process'. The relation R on Z is said to be a *product relation* if $R = \prod_i P_{Z_i}R$. We will say that z_i *splits* R if

$$R_{\{z_i=a_i\}} = P_{Z_1 \times \ldots \times Z_{i-1}} R \times P_{Z_{i+1} \times \ldots \times Z_{i+1}} R.$$

If z_i is splitting for all i then we will call R *Markovian*. Using these notions one can develop a systematic and novel approach to deterministic system theory and its realization problems which is a bit more general and in some applications much more appropriate than the existing input/output approach. However in the present paper we will only pursue the realization problem for $T = \{1,2\}$. Then we have the following definitions:

Let R_e be a given relation on $Y_1 \times Y_2$. If R is a relation on $Y_1 \times X \times Y_2$ such that

(i) x splits (for simplicity we call R then *splitting*), and

(ii) $R_e = P_{Y_1 \times Y_2} R$,

then we will call R a (splitting) *realization* of R.

Many of the qualitative notions of classical deterministic and stochastic realization theory admit very natural generalizations to this framework: Let R_1 and R_2 be two realizations of the same R_e with respective splitting spaces X_1 and X_2. Consider now the pre-ordering $R_1 > R_2$ defined by

$$\{R_1 > R_2\} :\Longleftrightarrow \{\exists \text{ a partial surjective set to point map } f\colon X_1 \to X_2 \text{ such that } \{(y_1,x_2,y_2) \in R_2\} \Rightarrow \{\exists x_1 \in f^{-1}(x_2) \text{ such that } (y_1,x_1,y_2) \in R_1\}\}.$$

Two realizations will be called *equivalent* if there exists a bijection $f\colon X_1 \to X_2$ such that $\{(y_1,x_1,y_2) \in R_1\} \Longleftrightarrow \{(y_1,f(x_1),y_2) \in R_2\}$. A realization R of R_e is said to be *irreducible* if any other realization $R' < R$ is necessarily equivalent to R'. It is said to be *attainable* if for all $x \in X$ there is $(y_1,y_2) \in R_e$ such that $(y_1,x,y_2) \in R$ and such that $\{(y_1,x',y_2) \in R\} \Rightarrow \{x'=x\}$. It is said to be *observable* if $a \mapsto P_{Y_2} R_{\{x=a\}}$ is injective as a map from X to 2^{Y_2}. It is said to be *reconstructible* if $a \mapsto P_{Y_1} R_{\{x=a\}}$ is injective. It is easy to show that irreducibility implies attainability, observability and reconstructibility, but as we shall see shortly, the converse is not necessarily true.

Let R be a realization of R_e. Then it is said to be *output induced* if there exists $f\colon Y_1 \times Y_2 \to X$ such that $\{(y_1,x,y_2) \in R\} \Longleftrightarrow \{x = f(y_1,y_2)\}$; it is said to be *past (future) output induced* if $f\colon Y_1 \to X$ ($f\colon Y_2 \to X$). (Actually output induced realizations have multiplicity one. The *multiplicity* of a point $(y_1,y_2) \in R_2$ in the realization R is the cardinality of the set $\{a \in X \mid (y_1,y_2) \in R_{\{x=a\}}\}$. In output induced realizations every point of R_e is thus covered exactly once.)

Geometric Illustration: Let R_e be a subset of $\mathbb{R}^2$. A realization of R_e is simply a family (parametrized by elements of X) of rectangles which together cover R_e exactly. If every point is covered once then this realization is output induced. It is irreducible if no non-trivial recombinations or deletion of these rectangles results in a new realization. Thus the problem of finding an irreducible realization is the problem of filling up a given set by an (in this sense) minimal number of (non-overlapping) rectangles. We can also view the realization R as a relation on $\mathbb{R} \times X \times \mathbb{R}$ which has rectangular x-level sets and which projected down along X yields R_e. This realization is output induced if R is a 'surface' with 'global chart' $\mathbb{R}^2$.

The following proposition links some of the concepts introduced above:

PROPOSITION. *Let* R *be a realization of* R_e. *Then*

(i) {R *is irreducible*} $\Rightarrow$ {R *is attainable, observable, and reconstructible*};

(ii) {R *is irreducible*} $\Longleftrightarrow$ {R *is attainable*} *and*

{{$R_{\{x \in A\}}$ *is rectangular*} $\Longleftrightarrow$ {A *consists of at most one point*}}.

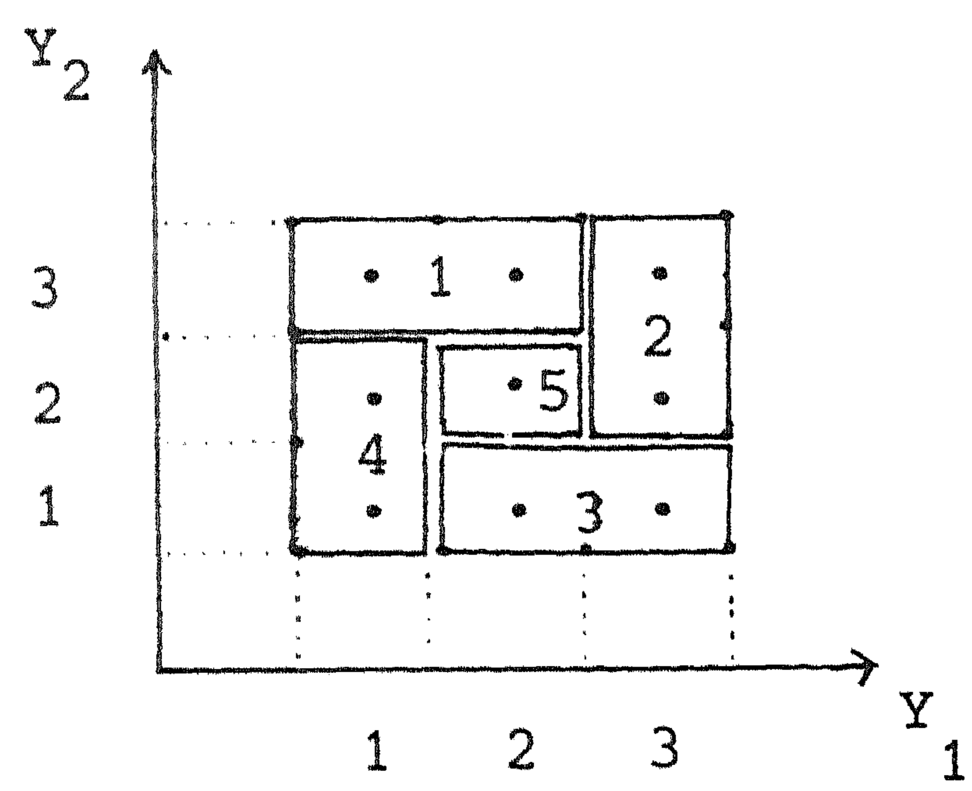
It would be of much interest to give this last property in (ii) a satisfactory system theoretic interpretation. Actually the above proposition falls considerably short from the results of the classical literature on deterministic realization theory. This is due to the fact that the realizations considered there are all past output induced (actually in that context it is better to speak of 'past input induced').

PROPOSITION. *Let* R *be a past output induced realization of* R_e. *Then* {R *is irreducible*} $\Longleftrightarrow$

(i) {R *is attainable*} *(which in this case means that for all* $x \in X$ *there exists* $(y_1, y_2) \in R_e$ *such that* $(y_1, x, y_2) \in R_e$, *i.e., reachability), and*

(ii) {R *is observable*}.

An analogous proposition holds for future induced realizations. It would be of interest to discuss the cases in which {irreducibility} $\Longleftrightarrow$ {readability, observability, and reconstructibility}, a situation which we shall have in the case of gaussian random variables. In general however ($\Leftarrow$) does not hold, not even for output induced realizations, as the following picture decisively illustrates:



$Y_1 = Y_2 = \{1,2,3\}$
$X = \{1,2,3,4,5\}$

$R_e = Y_1 \times Y_2$
R: see picture:
$(3,1,1) \in R$, $(3,1,2) \in R$, etc.

In trying to construct realizations there are three constructions which appear natural:

(i) by defining an equivalence relation E_1 on Y_1 defined by $\{y_1' E_1 y_1''\} :\Longleftrightarrow \{\{(y_1',y_2) \in R_e\} \Longleftrightarrow \{(y_1'',y_2) \in R_e\}\}$ and taking for $X = Y_1/E_1$ and defining a (past output induced) realization from there in the obvious way as $R^+ = \bigcup_{(y_1,y_2)\in R_e} (y_1, y_1 (\mathrm{mod}\ E_1), y_2)$. This realization will be called the *forward canonical realization*;

(ii) using the same idea on the set Y_2, thus obtaining R^-, the (future output induced) *backward canonical realization*;

(iii) defining an equivalence relation E_{12} defined as the coarsest refinement of E_1 and E_2, i.e., the equivalence relation on $Y_1 \times Y_2$ defined by $\{(y_1',y_2')E_{12}(y_1'',y_2'')\} :\Longleftrightarrow \{y_1' E_1 y_1''$ and $y_2' E_2 y_2''\}$ and proceeding in a similar fashion. The ensuring realization will be denoted by $R^{\pm}$.

The idea behind constructing R^+ is thus to view R_e as a map, f, from Y_1 into 2^{Y_1} and to define the equivalence relation as the kernel of f. Hence every splitting element in R^+ can either be identified by a subset of Y_1 (the elements of the partition induced by f) or a subset of Y_2 (the elements of the range of f). The family of subsets of Y_1 form a partition and are thus non-overlapping while the family of subsets of Y_2 need not have such structure.

It is easy to see that R^+ and R^- are refinements of the realization in (iii), and thus in general this realization will not be irreducible. We have the following result which is a rather nice generalization of what can be obtained in the classical case:

PROPOSITION.

(i) *The forward canonical realization* R^+ *is minimal in the class of past output induced realizations (in the sense that every other past output induced realization* R *satisfies* $R^+ < R$) *and thus irreducible.*

(ii) *All irreducible past output induced realizations are equivalent to* R^+ *and thus minimal and pairwise equivalent.*

(iii) *(see the previous proposition) A past output induced realization is irreducible iff it is reachable and observable.*

Needless to say that a similar proposition holds for R^-. Unfortunately general statements regarding the structure of the other (output induced) irreducible realizations appear hard to come by. We pose the following

Research Problem. Investigate whether every irreducible (output induced) realization R may be obtained from $R^{\pm}$ in the sense that $R < R^-$ and describe an effective procedure by which all irreducible realizations may be obtained from $R^{\pm}$.

2.2. Splitting Random Variables

The situation with random variables in much like the one with relations as explained in the previous section but where instead of having a yes-no situation on the elements in the relation one has a probability measure on the product space which expresses how likely two elements will be related. Of course, all notions such as independence then need to be interpreted in a measure theoretic sense. However, in the context of realization theory problems, a new dimension is added in the problem. This is akin to the problem of output induced versus not output induced realizations as discussed in the previous sections, but is also related to the common dichotomy in probability theory which is concerned with the question whether the probability space $\{\Omega,\mathcal{A},P\}$ is given (an impression which one gets from studying modern mathematical probability theory) or whether it is to be constructed (a point of view which appears much closer to what one needs in applications). In this section we will describe the random variable approach and the next section is devoted to the measure approach.

For a brief review of some relevant notions from probability theory, the reader is referred to the Appendix.

Let $\{\Omega,\mathcal{A},P\}$ be a probability space, $(Y_1,\mathcal{Y}_1)$, $\{Y_2,\mathcal{Y}_2\}$, and $\{X,\mathcal{X}\}$ be measurable space, and $y_1,y_2,x\colon \Omega \to Y_1,Y_2,X$ random variables on Ω. We say that x *splits* y_1 and y_2 if y_1 and y_2 are independent given x, in which case we say that (y_1,x,y_2) realizes (y_1,y_2). A realization is said to be *irreducible* if

(i) $\{X,\mathcal{X}\}$ is Borel and x is surjective (see Appendix) (in which case we call the realization *attainable*), and

(ii) if $\{X',\mathcal{X}',x'\}$ is any other realization for which there exists a surjection $f\colon X \to X'$ such that the scheme

$$\begin{tikzcd} & X \arrow[dd, "f"] \\ \Omega \arrow[ur, "x"] & \\ & X' \arrow[ul, "x'"] \end{tikzcd}$$

commutes, then f is injective (see Appendix).

Two realizations with respective splitting spaces X_1 and X_2 are said to be *equivalent* if there exists a bijection $f\colon X_1 \to X_2$ such that $x_2 = f(x_1)$. A realization is said to be *observable* if $x \mapsto P(y_2|x)$ is injective and *reconstructible* if $x \mapsto P(y_1|x)$ is injective. It is easy to see that irreducibility implies attainability, observability, and reconstructibility, but the converse is in general not true, unless the random variables involved are all gaussian random vectors.

A realization is called (past, future) *output induced* if there exists $(f\colon Y_1 \to X,\ f\colon Y_2 \to X)$ $f\colon Y_1\times Y_2 \to X$ such that $(f(y_1) = x,\ f(y_2) = x)$ $f(y_1,y_2) = x$. In an output induced realization the subset of $Y_1\times X\times Y_2$ defined by

$$\{(y_1',x_1',y_2') \mid \exists\omega \text{ such that } (y_1',x',y_2') = (y_1(\omega),x(\omega),y_2(\omega))\}$$

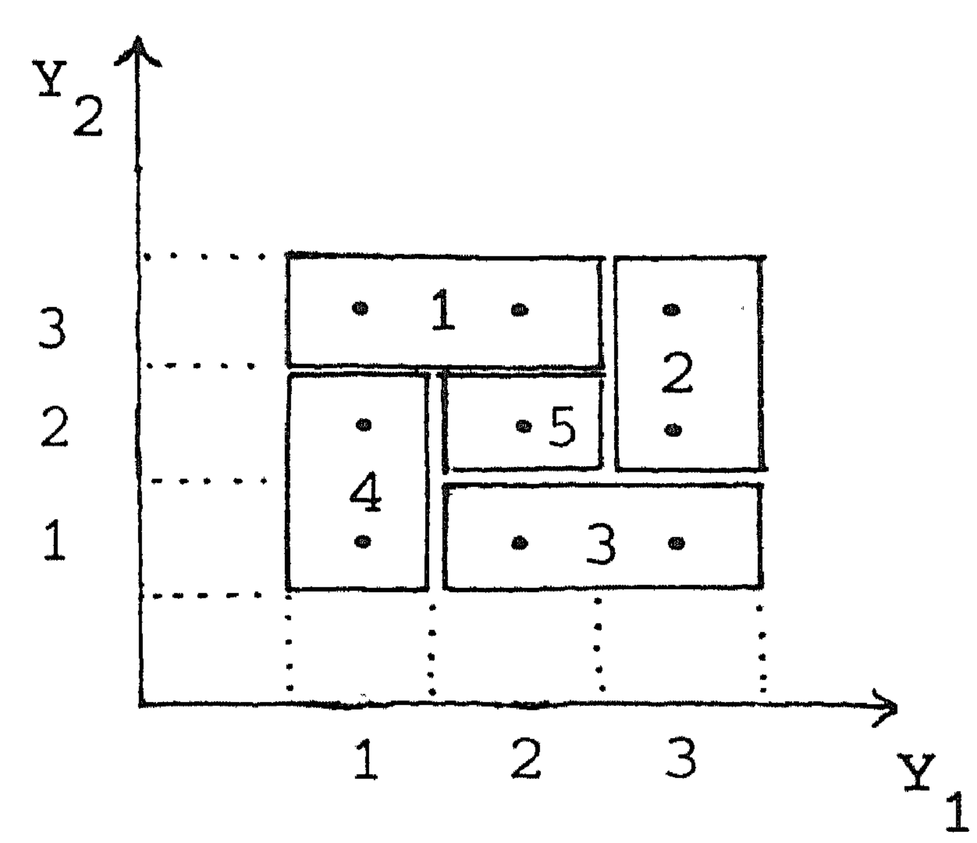
is a surface parametrised by ω or by the 'global chart' $Y_1\times Y_2$.

REMARK. The problem of finding a past output induced splitter is very akin to the Bayesian idea of sufficient statistic [22]. Formally: Let y_1,y_2 be random variables and assume that $f\colon Y_1 \to X$ is measurable. Then $x := f(y_1)$ is said to be a *sufficient statistic* for the estimation of y_2 through y_1 if x splits y_1 and y_2. Thus the problem of finding a sufficient statistic is the same as finding a past output based realization.

PROPOSITION. *Let* (y_1, x, y_2) *be a past (future) output induced realization of* (y_1, y_2). *Then*

$\{$*irreducibility*$\} \Longleftrightarrow \{$*attainability and observability (reconstructibility)*$\}$.

In general no such proposition is true, not even for output induced realizations. However for gaussian random variables we will show that $\{$irreducibility$\} \Longleftrightarrow \{$attainability, observability, and reconstructibility$\}$. An example which shows where things can go wrong is the following:



$$\Omega = Y_1 \times Y_2$$
$$Y_1 = Y_2 = \{1,2,3\}$$
$$p(y_1, y_2) = \frac{1}{9} \text{ for all } y_1, y_2$$
$$X = \{1,2,3,4,5\}$$

The realization is the output induced realization defined by $f\colon Y_1 \times Y_2 \to X$ $:= \{1,3\} \mapsto 1$, $\{2,3\} \mapsto 1$, $\{3,3\} \mapsto 2$, $\{3,2\} \mapsto 2$, $\{3,1\} \mapsto 3$, $\{2,1\} \mapsto 3$, $\{1,1\} \mapsto 4$, $\{1,2\} \mapsto 4$, $\{2,2\} \mapsto 5$.
It is easily shown that it is attainable, observable, and reconstructible. However the realization is not irreducible since $X' = \{1\}$ and $f\colon X \to X'$, defined by $f\colon x \mapsto \{1\}$, yields a reduced realization.

Similarly as in the deterministic case with relations one may construct the canonical past output induced realization R^+, the canonical future output induced realization R^-, and the join of both, $R^{\pm}$.

In constructing the forward canonical realization R^+, one considers the equivalence realation E_1 on Y_1, defined by

$$\{y_1' E_1 y_1''\} :\Longleftrightarrow \{P(y_2 | y_1') = P(y_2 | y_1'')\}.$$

Defining now

$$X = Y_1/E_1, \quad \mathcal{X} = \mathcal{Y}_1/E_1 \quad \text{and} \quad x\colon \omega \mapsto y_1(\omega)/E_1$$

leads to a past output induced realization of (y_1, y_2). Note that one may identify elements of X with subsets of Y_1 (those given by the partition of X induced by E_1) or with 'random' probability measures on Y_2 (given by $P(y_2|y_1)$). Note however that in the second parametrization it is a bit more difficult to give the exact nature of the subset of probability measures on Y_2 which are thus obtained. It is possible to formulate a proposition which reads identical to the last proposition in Section 2.1. Its proof presents no difficulties, at least in countable or smooth finite dimensional case. The technical details however still need to be worked out. We formulate this as a

<u>Research Problem</u>. Prove the stochastic analogon of the last proposition of Section 2.1 in the case that Y_1 and Y_2 are arbitrary measurable spaces. Investigate whether every irreducible (output induced) realization may be obtained from $R^{\pm}$ in a similar manner (a surjective set to point mapping on the splitting spaces) as will be the case for (deterministic) relations.

2.3. <u>Splitting measures</u>

Much of what has been said in Section 2.2 may be repeated for the case in which a probability measure is given on $Y_1 \times Y_2$ or on $Y_1 \times X \times Y_2$ directly. We will not give all the relevant definitions in detail but restrict ourselves to the definition of a realization.

Let $\{Y_1, \mathcal{Y}_1\}$ and $\{Y_2, \mathcal{Y}_2\}$ be measurable spaces and let P_e be a probability measure on $Y_1 \times Y_2$. A *realization* of P_e is defined by a measurable space $\{X, \mathcal{X}\}$ and a probability measure on $Y_1 \times X \times Y_2$ which induces P_e on $Y_1 \times Y_2$ and which is such that x splits y_1 and y_2.

The problem of realizing two given random variables (y_1, y_2) may hence be interpreted in the sense of the notions defined in Section 2.2 or in the sense of the above definition where we take the given measure on $Y_1 \times Y_2$ which is to be realized to be the one induced on $Y_1 \times Y_2$ by the probability measure on Ω. We now formalize these two possible type of realization of given random variables.

Let $\{\Omega_1,\mathcal{A}_1,P_1\}$ and $\{\Omega_2,\mathcal{A}_2,P_2\}$ be two probability spaces, $\{Y,\mathcal{Y}\}$ a measurable space, and z_1: $\Omega_1 \to Y$ and z_2: $\Omega_2 \to Y$ be two random variables on Ω_1 and Ω_2 respectively. We will say that z_1 and z_2 are *equivalent* if the measure induced by z_1 on Y is the same as the one induced by z_2 on Y.

PROBLEM 1 (*the strong realization problem*). Let (y_1,y_2) be given random variables defined on a probability space $\{\Omega,\mathcal{A},P\}$. The strong realization problem consists in finding the measurable spaces $\{X,\mathcal{X}\}$ and the random variable x: $\Omega \to X$ such that (y_1,x,y_2) is a realization of (y_1,y_2).

PROBLEM 2 (*the weak realization problem*). Let (y_1,y_2) be given random variables defined on a probability space $\{\Omega,\mathcal{A},P\}$. The weak realization problem consists in finding a probability space $\{\Omega',\mathcal{A}',P'\}$, the measurable space $\{X,\mathcal{X}\}$, and the random variables y_1',x,y_2': $\Omega' \to Y_1,X,Y_2$ such that x splits y_1' and y_2' and such that (y_1',y_2') is equivalent to (y_1,y_2).

It is clear from these problem statements that Problem 2 is actually a problem which involves finding a splitting measure and it is best to think about it in these terms, without involving Ω at all, but starting from the measures induced on $Y_1 \times Y_2$. In considering Problem 1 it is unclear what Ω should be and usually one would take $\Omega = Y_1 \times Y_2$. The problem then becomes precisely the problem of finding output induced realizations as discussed in Section 2.2. A possible and meaningful generalization on which very little work has been done so far is to start with three random variables y_1, y_2, and z, defined on Y_1, Y_2, and Z respectively, taking $\Omega = Y_1 \times Y_2 \times Z$ and finding strong realizations of (y_1,y_2). One could also ask the question if there are z-induced realizations. In this context z would thus be the random variable which carries the information on which the splitter x has to be based.

As we already see in the next section there is very much of a difference in the specific solutions of the strong and the weak realization problems.

2.4. The gaussian case

In this section we will give a rather complete picture of the problems formulated in the previous sections in an important particular case, namely when all the random variables are jointly gaussian. We will thus assume that

(y_1,y_2) is a real zero mean gaussian random vector with $y_1 \in \mathbb{R}^{n_1}$ and $y_2 \in \mathbb{R}^{n_2}$. Also, we will be looking for realizations (y_1,x,y_2) which are real zero mean gaussian random vectors with $x \in \mathbb{R}^n$. The measure of (y_1,x,y_2) is hence completely specified by

$$\Sigma = \left[\begin{array}{c|c|c} \Sigma_{11} & \Sigma_{1x} & \Sigma_{12} \\ \hline \Sigma_{x2} & \Sigma_{xx} & \Sigma_{x2} \\ \hline \Sigma_{21} & \Sigma_{2x} & \Sigma_{22} \end{array}\right],$$

where $\Sigma_{11} = E\{y_1y_1^T\}$, $\Sigma_{1x} = E\{y_1x^T\}$, etc.

The conditional independence condition is specified in the following:

<u>PROPOSITION</u>. *The following conditions are equivalent:*

(i) x *splits* y_1 *and* y_2*;*

(ii) $y_1 - E\{y_1|x\}$ *and* $y_2 - E\{y_2|x\}$ *are independent;*

(iii) *(if* $\Sigma_{xx} > 0$*)* $\Sigma_{12} = \Sigma_{1x}\Sigma_{xx}^{-1}\Sigma_{x2}$.

Also the conditions for the irreducibility, attainability, etc., of a realization are easily established:

<u>PROPOSITION</u>. *Assume that* (y_1,x,y_2) *is a realization of* (y_1,y_2)*. Then*

(i) $\dim x =: n \geq \text{Rank } \Sigma_{12} =: n_{12}$;

(ii) $n = n_{12}$ *iff the realization is irreducible;*

(iii) $\Sigma_{xx} > 0$ *iff the realization is attainable;*

(iv) $\text{Rank } \Sigma_{2x} = n_{12}$ *iff the realization is observable;*

(v) $\text{Rank } \Sigma_{1x} = n_{12}$ *iff the realization is reconstructible;*

(vi) $n = n_{12}$ *iff* $\Sigma_{xx} > 0$ *and* $\text{Rank } \Sigma_{2x} = \text{Rank } \Sigma_{1x} = n_{12}$.

As a consequence of property (ii) it is natural to call irreducible realizations *minimal*. Notice that for this gaussian case, even for realizations which are <u>not</u> output induced, we obtain the equivalence {irreducibility} $\Longleftrightarrow$ {attainability, observability, and reconstructibility}. Note also that here irreducibility means that whenever a surjective matrix S is such that $\{y_1,Sx,y_2\}$ is also a realization of (y_1,y_2), then S is necessarily square and invertible.

In most of the problems of realization theory the choice of the bases is immaterial and we may thus choose them to our convenience. The choice of the bases of the vector spaces in which y_1 and y_2 lie will be choosen so as give us the canonical variable representation, as introduced by HOTELLING [23].

LEMMA. *There exist nonsingular matrices* S_1 *and* S_2 *such that the covariance matrix of* $\bar{y}_1 := S_1 y_1$ *and* $\bar{y}_2 := S_2 y_2$, *defined by*

$$\bar{\Sigma} := \left[\begin{array}{c|c} \bar{\Sigma}_{11} & \bar{\Sigma}_{12} \\ \hline \bar{\Sigma}_{21} & \bar{\Sigma}_{22} \end{array}\right]$$

with $\bar{\Sigma}_{11} := E\{\bar{y}_1\bar{y}_1^T\}$, $\bar{\Sigma}_{12} = \bar{\Sigma}_{21}^T := E\{\bar{y}_1\bar{y}_2^T\}$, *and* $\bar{\Sigma}_{22} := E\{\bar{y}_2\bar{y}_2^T\}$ *takes the form*

$$\left[\begin{array}{cccc|cccc} I & 0 & 0 & 0 & I & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 & \Lambda & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline I & 0 & 0 & 0 & I & 0 & 0 & 0 \\ 0 & \Lambda & 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

identical components
correlated components
independent components
zero components

where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{n_{21}})$ *with* $1 > \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{n_{21}} > 0$.

We will denote the various components of $\bar{y}_1$ and $\bar{y}_2$ in this basis by $\bar{y}_{11}$, $\bar{y}_{12}$, etc. and their dimension by n_{11}, n_{12}, etc. Moreover, since we will assume that this basis transformation has been carried out we will drop the bars on the y's.

The components of y_1 and y_2 in this representation are called *canonical variables*. They are very useful in statistical analyses. They are unique modulo the following transformation: an orthogonal transformation on y_{11} and y_{21}, one on y_{13}, one on y_{23}, one on y_{14}, and one on y_{24}. Moreover, if in the sequence of λ_i's there is equality:

$$\lambda_{i-1} > \lambda_i = \lambda_{i+1} = \ldots = \lambda_{i+k} > \lambda_{i+k+1},$$

then one can also apply an orthogonal transformation on the component of y_{21} and y_{22} corresponding to these equal λ_i's.

The above lemma shows how the basis for y_1 and y_2 is choosen. We choose the basis of x as follows:

PROPOSITION. *Assume that the* (y_1,x,y_2) *is irreducible. Then we may always choose the basis for* x *such that* $E\{y_2|x\} = x$. *This implies together with conditional independence that* $\Sigma_{2x} = \Sigma_{xx}$ *and* $\Sigma_{1x} = \Sigma_{12}$.

The following two theorems are the main results of this section. The first theorem solves the 'weak' realization problem for gaussian random vectors, while the second theorem solves the 'strong' realization problem. The interpretation of the output induced irreducible realizations in terms of canonical variables in rather striking.

THEOREM. *In the bases given,* (y_1,x,y_2) *will be a minimal weak realization of* (y_1,y_2) *iff the correlation matrix of* (y_1,x,y_2) *takes the form*

$$\left[\begin{array}{cccc|cc|cccc}
I & 0 & 0 & 0 & I & 0 & I & 0 & 0 & 0 \\
0 & I & 0 & 0 & 0 & \Lambda & 0 & \Lambda & 0 & 0 \\
0 & 0 & I & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
I & 0 & 0 & 0 & I & 0 & I & 0 & 0 & 0 \\
0 & \Lambda & 0 & 0 & 0 & \Sigma & 0 & \Sigma & 0 & 0 \\
\hline
I & 0 & 0 & 0 & I & 0 & I & 0 & 0 & 0 \\
0 & \Lambda & 0 & 0 & 0 & \Sigma & 0 & I & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & I & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right]$$

with Σ *any matrix satisfying* $\boxed{\Lambda^2 \leq \Sigma \leq I}$

The above theorem seems more complicated than it is because we have taken a completely general case for (y_1,y_2). The point however is that in this choice of the bases it is exceedingly simple to see how the correlation matrix of (y_1,x,y_2) can look like. Note that the components which are common to y_1 and y_2 will appear in every realization as components of x. The uncorrelated components on the other hand donot influence x.

THEOREM. *Assume that* (y_1,x,y_2) *is a minimal output induced realization of* (y_1,y_2). *Then there exists a choice of canonical variables for* y_1 *and* y_2 *such that in a suitable basis* x *is given by* $x = (y_{11}=y_{21},z_1,z_2)$ *with* z_1 *a vector consisting of some components of* y_{21} *and* z_2 *a vector consisting of the other components of* y_{22}. *Conversely, for every choice of the canonical variables and every such choice of* z_1 *and* z_2, x *will be an output induced minimal realization of* (y_1,y_2).

Let us call two gaussian realizations (y_1,x_1,y_2) and (y_1,x_2,y_2) *equivalent* if there exists a non-singular matrix S such that $x_2 = Sx_1$. Otherwise they will be called *distinct*. From the above theorems the following corollary is immediate:

COROLLARY. *The number of distinct weak realizations is one in the case* $n_{21} = n_{22} = 0$ *and non-denumerably infinite otherwise. The number of distinct output induced minimal realizations is* $2^{n21} = 2^{n22}$ *if* $1 > \lambda_1 > \lambda_2 >\ldots> \lambda_{n_{21}} > 0$, *and non-denumerably infinite otherwise.*

We close this section with some remarks:

1. Much of the structure of the realizations shown in the above theorem may by found in one form or another in the work of RUCKEBUSH [see e.g. 16,17,18,19].
2. In order to generate x in a weak realization, one has to add a source of randomness which is external to (y_1,y_2). In fact, one needs exactly $\mathrm{rank}(\Sigma-\Lambda^2) + \mathrm{rank}(I-\Sigma) - n_{12}$ additional independent random variables to achieve minimal a weak realization.
3. It is of interest to develop the above theory for given gaussian vectors (y_1,y_2,z) and requiring x to be z-induced in the realization (y_1,x,y_2) of (y_1,y_2).

3. STOCHASTIC DYNAMICAL SYSTEMS

We will in this paper exclusively be concerned with stochastic systems without external inputs (with this we mean that the dynamics are influenced by a chance variable but not by other 'external' inputs). We will therefore introduce the relevant concepts on this level of generality. Properly speaking we are concerned with *autonomous* stochastic systems. These are described by a stochastic process.

3.1. Basic definitions

DEFINITION. A *stochastic system (in output form)*, Σ_e, is defined by

(i) a probability space $\{\Omega, A, P\}$,

(ii) a *time index set* $T \subset \mathbb{R}$,

(iii) a measurable space $\{Y, \mathcal{Y}\}$ called the *output space*, and

(iv) a stochastic process $y: T\times\Omega \to Y$.

It is said to be *time-invariant* if T is an interval in $\mathbb{R}$ of Z, and

(v) y is stationary.

Many of our comments are in the first place relevant to the time-invariant case.

Let $z: \Omega\times T \to Z$ be a process. We will denote by $z_t^- := \{z(\tau), \tau<t\}$ the *past* and by $z_t^+ := \{z(\tau), \tau>t\}$ the *future* of z. Let $z,r: \Omega\times T \to Z,R$ be two processes. We will say that r *splits* z if r(t) splits z_t^- and $(z_t^+, z(t))$ for all t. Notice that this definition is not symmetric in time.

DEFINITION. Let Σ be a stochastic system on $X\times Y$. Then it will be said to be in *state space form* with *state space* X and *output space* Y if x splits (x,y). The external *behaviour* of Σ is simply the process y which we may of course consider to be a stochastic system in its own right. We will denote it by Σ_e and say that Σ *realizes* Σ_e, denoted by $\Sigma \Rightarrow \Sigma_e$.

REMARKS.

1. It is obvious from the above definition that x will itself be a Markov process. Let R denote the time-reversal operator, i.e., $Rz(t) := z(-t)$. Note that $R\Sigma$ will in general not be a system in state space form. Thus, contrary to Markov processes, a state space system forward in time need not be a state space system backwards in time. However, in the continuous time case there is a weak condition under which we will have this time-reversibility, i.e., when the σ-algebras induced by $\{y(\tau), \tau<t\}$ and $\{y(\tau), \tau\leq t\}$ are equal for all t. This will be the case whenever the sample paths of y are smooth in some appropriate sense. Actually, in that case

there exists a map $f: X \times T \to Y$ such that $y(t) = f(x(t),t)$.

2. It is likely that there are many applications (e.g. in recursive signal processing and in stochastic control) where the relevant property is that the process x splits y, and that the state property expressed in the fact that x splits (x,y) is not as crucial as we have learned to think.

The problem of (state space) realization is then simply the following: *Given* Σ_e, *find* Σ *such that* $\Sigma \Rightarrow \Sigma_e$. This problem is the strong realization problem. The alternative approach to stochastic systems where one starts with measures leads to the following

DEFINITION. A *stochastic system (in output form) in terms of its measures* is defined by giving for all $t_1, t_2, \ldots, t_n \in T$ a probability measure on Y^n satisfying the usual compatibility conditions. Such a system defined on $X \times Y$ is said to be in *state space form* if for all $t_1 \leq t_2 \leq \ldots \leq t_n \leq t$ and for every bounded real measurable function f there holds
$E\{f(x(t),y(t)) \mid (x(t_i),y(t_i)),\ i = 1,2,\ldots,n\} = E\{f(x(t),y(t)) \mid x(t_n)\}$.

Instead of specifying the measures directly as in the above definition one can do this in terms of appropriate kernels which express how a given state $x(t_0)$ will result in a state/output pair $(x(t_1),y(t_1))$ at $t_1 \geq t_0$. This construction requires only a slight extension from the usual construction of Markov kernels and we will not give them in detail here.

From the above definition it should be clear what one means by the externel behaviour of a system defined in terms of its measures and hence by the weak version of the state space realization problem. If we also consider the fact that for every process defined in terms of its marginal probability laws one can construct a probability space and an equivalent stochastic process on it, then we see that this weak version of the realization problem may be expressed in the following, albeit somewhat indirect, way:

DEFINITION. Two stochastic systems Σ_e' and Σ_e with the same time index set and output space are said to be *equivalent* if the defining processes are equivalent in the sense that they have the same marginal measures. We will call Σ' a *weak realization* of Σ_e if Σ' is in state space form and if Σ_e' is equivalent to Σ_e.

All of the definitions (irreducibility, attainability, equivalence, observability, and reconstructibility) have obvious generalizations to the problem at hand. We will not give them explicitely. Instead we turn to a topic which we will only touch on very briefly but which fits very well in an exposition on the representation of stochastic systems.

3.2. White noise representation

Often, particularly in the engineering literature, one starts with a stochastic system which is actually defined in terms of a deterministic system driven by "white noise". This starting point may be introduced in our framework as follows:

DEFINITION. Let T be a (possibly infinite) interval in $\mathbb{Z}$. A *discrete time white noise driven stochastic system* (denoted by Σ_r: r stands for 'recursive') is defined by

(i) three processes, $w,x,y: T \to W,X,Y$ with w white noise and (x,y) a stochastic system in state space form, such that

(ii) x_t^- and $(w(t),w_t^+)$ are independent for all t, and

(iii) two maps $f,r: X\times W\times T \to X,Y$ called respectively the *next state map* and the *read-out map*, such that

$$x(t+1) = f(x(t),w(t),t)$$
$$y(t) = r(x(t),w(t),t).$$

There is clearly an analogue of this definition for systems defined in terms of their measures. We will denote the system Σ in state space form induced in the obvious way by Σ_r by $\Sigma_r \Rightarrow \Sigma$. The *white noise representation problem* is the problem of finding for a given stochastic system in state space form Σ a white noice driven system Σ_r such that $\Sigma_r \Rightarrow \Sigma$. Thus in this problem one is asked to construct the white noise process w and maps f and g. There is also an obvious 'weak' version of this problem.

The following definition gives a limited continuous time version of the above:

DEFINITION. Let T be a (possibly infinite) interval in $\mathbb{R}$. A continuous time *gaussian white noise driven stochastic system* (denoted by Σ_d: d stands for 'differential') is defined by

(i) three processes w,x,y: $T \to W,X,Y$ with w an m-dimensional Wiener process on T, X a subset of $\mathbb{R}^n$, and (x,y) a stochastic system in state space form, such that

(ii) x_t^- and $(w - w(t))_t^+$ are independent for all t, and

(iii) three maps f,h,r: $X \times T \to \mathbb{R}^n$, $\mathbb{R}^{n\times m}$, Y respectively called the *local drift*, the *local diffusion*, and the *read-out map*, such that

$$dx(t) = f(x(t),t)dt + h(x(t),t)dw(t),$$
$$y(t) = r(x(t),t).$$

There are obvious generalizations of these notions to independent increment processes and, more to the point, to more general stochastic differential equations admitting, for example, also jump processes.

Note that if in the discrete time (or the continuous time) case T has a lower bound, say $t = 0$, then $x(0),w,f$ and r (resp. f,h and r) determine completely x and y. A similar property occurs when T has no lower bound but when the differential equation for x has certain asymptotic stability properties. In such case a white noise representation of a realization of a given dynamical system in output form will yield maps by which $y(t)$ is expressed, in a non-anticipating manner, as a function of the independent random variables $x(0),w(0),w(1),\ldots,w(t)$.

3.3. Research problems

In this section we will indicate the sort of questions one asks in realization theory and, as far as this is known, how one expects the answers to look like. In Section 4 it will be shown how some of these questions may be answered for time-invariant finite-dimensional gaussian processes. The claims made in the various questions will all require certain regularity conditions which we will not be concerned with here.

1. Let Σ_e be a given stochastic system.
   The basic problem is to construct a state space representation of it.

Actually, if no other properties are asked then this problem is somewhat trivial. The following question however is much more useful and interesting: Find all (weak and strong) irreducible realizations of Σ_e. Give conditions under which Σ_e admits a finite state realization or a finite dimensional realization. Give algorithms for going from a numerical specification of Σ_e to a numerical specification of an irreducible realization.

2. Describe the construction of the canonical past output and future output induced realizations. Show in what sense they are unique. Prove (or disprove) that all irreducible realizations may be derived from the join of the past and the future induced realizations, and describe how all irreducible realizations could this way be constructed.

3. Let Σ_e be given and assume that z splits y. It is easily seen how to define irreducibility of z as a splitting process. It is not unreasonable to expect that an irreducible splitting process z would yield an (irreducible) realization (z,y) of y. It would appear from [15] that this will not always be true. It is of interest to settle this problem and give the additional conditions which z needs to satisfy.

4. When does a discrete time system admit a white noise representation? The expectation is that for weak representations this will always be possible but that strong white noise representability is rather special. An appropriately general concept of white noise representability must undoubtly allow the space W where the white noise takes its values to be x-dependent, pretty much like the situation with vector fields and bundles as a description of flows on manifolds. It is of much interest to clarify the probabilistic versions of such recursive stochastic models.

5. When does a continuous time system admit a white noise representation? The expectation here is that for weak representations this ought to require only some smoothness of the random process x as a function of t while for strong realizations it will also involve some sort of constant rank condition on the local covariance of x. What sort of uniqueness results can here be gotten? Some partial results for scalar equations are available in [5,24]. Extend these representation problems to

differential equations involving jump processes.

6. Clarify the relations between the stochastic realization problem and the problem of white noise representation with the problem of representing a given stochastic process as a non-anticipating function of a sequence of independent random variables. To be more specific, assume that a stochastic system has a state space realization which admits a white noise representation. This representation will often lead, as explained before, to a representation $y(t) = G_t(w(t), w(t-1), \ldots, w(0), x(0))$ when T has a lower bound or $y(t) = G_t(w(t), w(t-1), \ldots)$ when T has no lower bound. Clearly this last representation is not always possible. The question is to prove when and how it is possible: e.g., is it possible for ergodic processes? In what sense are such representations unique?
7. Show the relation and the interaction between forward time and backward time realizations and white noise representations. Investigate possible applications to filtering, prediction, smoothing, and stochastic control.
8. Generalize the concepts given here to stochastic systems which are influenced not only by a chance variable w, but also by deterministic external inputs.

4. REALIZATION OF GAUSSIAN PROCESSES

In this section we will treat gaussian processes. Most of the results which we present have been known since the work of FAURRE [7,8,9]. We will briefly touch on the output induced realizations which have been discovered rather recently by LINDQUIST & PICCI [12] and RUCKEBUSH [16]. Actually, this whole area is still in a lot of motion and it is difficult to do justice to all the ideas which have recently been put forward. The importance of these results in recursive signal processing and filtering are discussed in [9,21]. For simplicity and in view of the space limitation, we will only consider the continuous time case.

Let $y: \Omega \times \mathbb{R} \to \mathbb{R}^p$ be a zero mean stationary gaussian process, defined on a probability space $\{\Omega, \mathcal{A}, P\}$. We assume y to be mean square continuous (since, as we shall see later, we are really interested in the case that y has a rational spectral density, this implies no real loss of generality). As we have already argued in Section 3, this assumption allows us to concentrate on state space realizations of the form $y(t) = f(x(t))$ with x a Markov process. In particular, forward time realizations are thus automatically backward time realizations, a situation which is no longer valid in the discrete time case. The main problems which one considers in this area are the following:

PROBLEM 1 (*the weak realization problem*). Under what conditions does there exist a zero mean stationary $\mathbb{R}^n$-valued Gauss-Markov process x and a matrix $C: \mathbb{R}^n \to \mathbb{R}^p$ such that y is equivalent to Cx (which in this case simply means that y and Cx should have the same autocorrelation function)? What is the minimum dimension of x for which this is possible? Develop algorithms for deriving C and the parameters of the Markov process x.

PROBLEM 2 (*the strong realization problem*). If Problem 1 is solvable, will there also exist such Gauss-Markov processes x which are output induced (i.e., for which there exists a map $y \mapsto x$)? Develop algorithms for deriving the parameters of the Gauss-Markov process in this case.

PROBLEM 3 (*the white noise representation problem*). Under what conditions does there exist an m-dimensional Wiener process w, an n-dimensional Gauss-Markov process x, both defined on $\mathbb{R}$, and appropriately sized matrices A,B,C such that
(i) x_t^- is independent of $(w - w(t))_t^+$,
(ii) $dx(t) = Ax(t)dt + Bdw(t)$, $y(t) = Cx(t)$.
Develop algorithms for computing the matrices A,B,C.

The marginal probability measures of y (and thus all processes which are equivalent to it) are completely specified by its *autocorrelation* function $R: \mathbb{R} \to \mathbb{R}^{p\times p}$, defined by

$$R(t) := E\{y(t)y^T(0)\}.$$

The restriction of R to $[0,\infty)$ will be denoted by R^+. Of course, R^+ specifies R completely since $R(t) = R^T(-t)$. Let $\Phi: \mathbb{R} \to \mathbb{R}^{p\times p}$ denote the *spectral density function* of y. Very roughly speaking Φ is defined as the Fourier transform of R. The following lemma is well-known:

LEMMA. *The following conditions are equivalent:*

(i) R^+ *is Bohl (i.e., every entry of* R^+ *is a finite sum of products of a polynomial, an exponential, and a trigonometric function);*

(ii) Φ *is rational;*

(iii) *there exists matrices* $\{F,G,H\}$ *such that* $R^+(t) = He^{Ft}G$, *with* $F \in \mathbb{R}^{n\times n}$, $G \in \mathbb{R}^{n\times p}$, *and* $H \in \mathbb{R}^{p\times n}$.

In addition:

(iv) *there is a minimal* n, n_{min}, *for which the factorization in* (iii) *is possible, called the* *McMillan degree* *of* R^+ *or* Φ, *and* $n = n_{min}$ *iff* (F,G) *is controllable and* (F,H) *is observable. The triple* (F,G,H) *is then called* *minimal*;

(v) *all* (F,G,H)*'s with* $n = n_{min}$ *are obtainable from one by the transformation group* $(F,G,H) \xrightarrow[\det S\neq 0]{S} (SFS^{-1},SG,HS^{-1})$.

Assume that x is an n-dimensional zero mean stationary Gauss-Markov process with $E\{x(t)x^T(t)\} =: \Sigma > 0$. Then x is mean square continuous and hence $E\{x(t) \mid x(0)\}$ for $t \geq 0$ is of the form $e^{At}x(0)$ for some A. Moreover it is easily seen that Σ and A completely specify the marginal measures of this Markov process. Hence the marginal measures of the process Cx are completely specified by (A,Σ,C) and it makes sense to talk about this triple as defining a weak realization of y. We will call a weak realization (x,y) with x n-dimensional and n as small as possible a *minimal* realization.

PROPOSITION. *Consider the stochastic system defined by* (A,Σ,C). *Then*

(i) *it is a weak realization of* y *iff* $Ce^{At}\Sigma C^T = R(t)$ $(t \geq 0)$;

(ii) *it is attainable iff* $\Sigma > 0$;

(iii) *it is observable iff* (A,C) *is observable;*

(iv) *it is reconstructible iff* $(A^T,C\Sigma)$ *is observable;*

(v) *it is irreducible iff* $(A,\Sigma C^T,C)$ *is a minimal realization of* R^+;

(vi) *it is irreducible iff it is minimal and* n_{min} = *the McMillan degree of* R^+.

Part (i) of the proposition may be verified by direct calculation. The claim about reconstructibility may be seen from considering the stochastic system R(x,y) with R the time-reversal operator. It is easily calculated that this stochastic system has parameter matrices $(\Sigma A^T\Sigma^{-1},\Sigma,C)$. This yields, after applying (iii), the condition in (iv).

Two (gaussian) stochastic systems (x_1,y) and (x_2,y) will be called *linearly equivalent* if there exists a nonsingular matrix S such that $x_2(t) = Sx_1(t)$. The following theorem classifies all minimal weak realizations of y up to linear equivalence:

THEOREM.

(i) *There exists a finite dimensional weak realization of* y *iff* y *has rational spectral density (for various equivalent conditions, see the previous lemma);*

(ii) *all minimal realizations* (A,Σ,C) *can, up to linear equivalence, be obtained from a minimal factorization triple* (F,G,H) *of* R^+ *by taking* $A = F$, $C = H$, *and solving the equations* $\boxed{F\Sigma+\Sigma F^T \leq 0,\ \Sigma H^T = G}$ *for* $\Sigma = \Sigma^T$.

The problem thus becomes one of solving a set of inequalities. Actually, it may be shown [9,25,26] that there exist solutions Σ_-,Σ_+ such that every solution Σ satisfies $0 < \Sigma_- \leq \Sigma \leq \Sigma_+ < \infty$. Moreover the solution set in convex and compact. A great deal of additional information on the structure of the solution set of these equations may be found in the above references.

Note that choosing $A = F$ and $C = H$ in the above theorem corresponds to fixing the basis in state space. Indeed, since $E\{y(t)\,|\,x(0)\} = Ce^{At}x(0)$ for $t \geq 0$, this choice of the basis of the state space is very much alike the situation in Section 3.3 where we also fixed the basis for x this way. Once the basis has been picked, it is only the covariance of x, Σ, which remains to be choosen.

There seems to be some applications, even in filtering, where <u>any</u> (also an indefinite) solution of the equality $\Sigma H^T = G$ can be used. This point was raised in Faurre's thesis [7] but seems to have been ignored since. That may have been a pity, since it is the inequality part which makes these equations hard to solve.

The strong realizations are covered in the following theorem:

THEOREM.

(i) *There exists a finite dimensional output induced realization of* y *iff* y *has rational spectral density (for various equivalent conditions, see the previous lemma);*

(ii) *the parameter matrices* (A,Q,C) *of all minimal output induced realizations of* y *may, up to linear equivalence, be obtained from a minimal factorization triple* (F,G,H) *of* R^+ *by taking* A = F, C = H, *and solving the equations* $\boxed{F\Sigma+\Sigma F^T \leq 0;\ \mathrm{rank}(F\Sigma+\Sigma F^T) = \text{minimum};\ \Sigma H^T = G}$ *for* $\Sigma = \Sigma^T$.

The matrices Σ^- and Σ^+ mentioned above actually correspond to output induced minimal realizations, and thus satisfy the equations at the above theorem. In fact, they correspond to the (up to linear equivalence) *unique* past output and future output induced minimal realizations. Thus in the corresponding realizations (x_-,y) and (x_+,y), the state $x_-(t)$ may be viewed as the parametrization of the conditional probability measure $P(y_t^+|y_t^-)$. Of course, $x_+(t)$ admits a similar interpretation. Actually, both the solution set of the Σ's of the above theorem and the corresponding output induced realizations (x,y) have a great deal of very appealing structure. It would take us too for to explain all that here. The reader is referred to [12,26,27] for details. We just like to mention one item: namely if (x_-,y) and (x_+,y) are the canonical past and future output induced realizations and (x,y) is any other output induced minimal realization, then there will exist a projection matrix P such that $x = Px_- + (I-P)x_+$. In fact, 'generically', there are a finite number of such matrices possible and in a suitable basis this states that every output induced irreducible realization may be obtained from x_-,x_+ by picking certain components from x_- and the remaining components from x_+ (see [12,27] for a more precise statement to this effect). These results show that the situation described in Section 2.3 is rather representative for the general case.

REMARKS.

1. It is clear from the above theorem that, contrary to what one sees happening for deterministic finite dimensional time-invariant linear systems, there is no unique (up to linear equivalence) minimal realization of stochastic finite dimensional stationary gaussian processes.

However, there is one thing which one can say, namely that if there exists a finite dimensional realization, then there will also exist a finite dimensional realization (z,y) having the property that all minimal realizations (x,y) may be deduced from it by a surjective matrix $S: z \mapsto x$. Among the realizations having this property there are again irreducible (i.e., minimal) elements and it may be shown that such an irreducible realization is unique up to linear equivalence. In this case such realizations could be called *universal*. One such universal realization may be deduced from (x_-, x_+) by defining

$$z(t) := (x_-(t), x_+(t))/\mathrm{Ker}\ Q$$

with

$$Q := E\left\{\begin{pmatrix} x_-(0) \\ x_+(0) \end{pmatrix}\begin{pmatrix} x_-(0) \\ x_+(0) \end{pmatrix}^T\right\}.$$

We expect that this type of universal realizations may have some as yet undiscovered applications for example in data processing.

2. An interesting problem which has recently been investigated is how time-reversibility of y may be reflected in its realization (x,y). The basic result obtained in [28] and, independently, in [29] states that if y is equivalent to Ry and if one chooses the weak realization (x,y) right, then in a suitable basis $x = (x_1, x_2)$ with (x_1, x_2, y) equivalent to $R(x_1, -x_2, y)$. The sign reversal of x_2 is the striking element in this result. This type of time-reversibility has been referred to as 'dynamic reversibility' in the stochastic processes literature.

3. There is a very interesting duality between the problems of realization of gaussian processes and that of passive electrical network synthesis. The basic synthesis with resistors, capacitors, inductors, transformers, and gyrators is dual to the weak realization problem with minimality referring to the minimality of the number of inductors plus the number of capacitors. The output induced realization problem turns out to be dual to the synthesis with a minimal number of resistors, and the time reversibility problem discussed in the previous paragraph turns out to be dual to the gyratorless synthesis problem. It would be nice to have

a more fundamental understanding of this situation!

We close this section with an almost trivial theorem on white noise representability.

THEOREM. *Let* (x,y) *be a finite dimensional zero mean gaussian dynamical system. Then there exists a white noise representation of the form:*

$$dx = Axdt + Bdw; \quad y = Cx.$$

In fact, when $E\{x(0)x^T(0)\} =: Q$, *then* (A,Q,C) *are precisely the parameter matrices of* (x,y) *as a realization of* y *and* B *is a solution of the equation:* $\boxed{AQ + QA^T = -BB^T}$.

Moreover, if (x,y) *is a minimal realization of* y *then* $\{(A,B)$ *controllable*$\} \Longleftrightarrow \{y$ *is ergodic*$\}$.

Note that the output induced realizations are precisely those which allow a white noise representation with a minimal dimensional driving Wiener process.

5. CONCLUSION

In this paper we have attempted to give an exposition of the problems of stochastic realization theory. We have been rather detailed on the abstract setting of the problem but somehwat scant on the realization of gaussian processes for which much more is available than we have covered here. Actually, we believe that the conceptual 'set theoretic level' definitions for stochastic systems have been neglected and that putting these problems on a sound footing would provide an important pedagogical, theoretical, and practical contribution in stochastic system and control theory.

APPENDIX

Let $\{\Omega, A, P\}$ be a probability space with Ω a set, A a σ-algebra of subsets of Ω, and $P: A \to [0,1]$ a probability measure on Ω. With a *random variable* we mean a measurable mapping $f: \Omega \to F$ with $\{F, \mathcal{F}\}$ a measurable space, i.e., F is a set and $\mathcal{F}$ is a σ-algebra on F. The measure $\mu(F') := P(f^{-1}(F'))$ defined for all $F' \in \mathcal{F}$ is called the *induced measure*. Note that our random variables need not be real. If F is a topological space and $\mathcal{F}$ is the smallest σ-algebra containing all open sets, then $\{F, \mathcal{F}\}$ is called a *Borel space*. We will delete $\mathcal{F}$ in the notation of a random variable whenever it is unimportant or clear what $\mathcal{F}$ exactly is.

A random variable f on a Borel space $\{F, \mathcal{F}\}$ is said to be *surjective* if the complement of $f(\Omega)$ contains no open sets, i.e., if the closure of $f(\Omega)$ is F. Let f_1, f_2 be random variables on $\{F_1, \mathcal{F}_1\}$ and $\{F_2, \mathcal{F}_2\}$ and $h: F_1 \to F_2$ a measurable map. Let μ_1 and μ_2 be the induced measures. In this context h is said to be *injective* if $\mu_1(\{a \mid \exists b \neq a \text{ such that } h(a) = h(b)\}) = 0$.

Let f be a random variable on $\{F, \mathcal{F}\}$. The sub σ-algebra of A defined by $f^{-1}(\mathcal{F})$ will be called the *sub σ-algebra induced by* f. We call two sub σ-algebras A_1, A_2 of A *independent* if $P(A_1 \cap A_2) = P(A_1) \cdot P(A_2)$ for all $A_1 \in A_1$ and $A_2 \in A_2$. Two random variables are *independent* if their induced sub σ-algebras are independent.

Let A_1, A_2, A_3 be sub σ-algebras of A. Then we will say that A_1 and A_3 are *conditionally independent* given A_2 if for all P-integrable real random variables f_1, f_3 which are respectively A_1 and A_3-measurable, there holds

$$E(f_1 f_3 | A_2) = E(f_1 | A_2) \cdot E(f_3 | A_2).$$

The random variables f_1 and f_3 are said to be *conditionally independent* given f_2 if the induced sub σ-algebras are conditionally independent. We also say that A_2 (resp. f_2) is *splitting*.

It is well-known how conditional expectation is defined and what one means with regular versions of conditional probabilities [30,p.139]. We will without explicit mention assume that these exist whenever needed, and denote by $P(f_2 | f_1)$ the conditional measure of f_2 given f_1, where f_1 and f_2 are two random variables. We will use a similar notation for conditional expectation.

Let T be a set and $f(t) : T\times\Omega \rightarrow F$ be a family of random variables. Then the family of measures induced for all n on F^n by the random variables $(f(t_1), f(t_2), \ldots, f(t_n))$ are called the *marginal probability measures* of f. If $T \subset \mathbb{R}$ then f is called a *random process*. The smallest σ-algebra containing all the σ-algebras induced by $f(\tau)$ for $\tau < t$ is called the σ-algebra induced by the past at time t. Whenever we are conditioning or taking conditional expectation with respect to the past this should be understood in this sense. We will denote the strict past (future) at t by f_t^- (f_t^+). A random process is said to be a *white noise process* if, for all t, f_t^- and $(f(t), f_t^+)$ are independent. It is said to be a *Markov process* if, for all t, f_t^- and f_t^+ are conditionally independent given f(t). An $\mathbb{R}^n$-valued process is said to be *gaussian* if all its marginal measures are gaussian and a *Gauss-Markov process* if it is both Markov and gaussian.

Finally, let $\{\Omega_1, A_1, P_1\}$ and $\{\Omega_2, A_2, P_2\}$ be two probability spaces and $f_1, f_2 : \Omega_1, \Omega_2 \rightarrow F$ be two random variables defined on the same outcome space $\{F, \mathcal{F}\}$. Then they are said to be *equivalent* if they induce the same measure on F. Similarly two random processes are said to be *equivalent* if they induce the same marginal probability measures.

REFERENCES

[1] PICCI, G., "On the internal structure of finite-state stochastic processes", pp. 288-304 in *Recent Developments in Variable Structure Systems, Economics and Biology* (Eds. R.R. Mohler and A. Ruberti), Proc. of US-Italy Seminar, Taormina, Sicily, 1977, Springer-Verlag, Lecture Notes in Economics and Mathematical Systems, Vol. 162, 1978.

[2] BROCKETT, R.W., "Stationary covariance generation with finite state Markov processes", Proceedings 1977 IEEE Conference on Decision and Control.

[3] BROCKETT, R.W. & BLANKENSHIP, G.L., "A representation theorem for linear differential equations with Markovian coefficients", 1977 Allterton Conference on Circuits and System Theory.

[4] VAN PUTTEN, C., & VAN SCHUPPEN, J.H., "On stochastic dynamical systems", Preprint BW 101/79, Mathematical Centre, Amsterdam, to appear in the Proc. of the Fourth International Symposium on the Mathematical Theory of Networks and Systems, Delft, 1979.

[5] DOOB, J.L., "The elementary Gaussian processes", Ann. Math. Stat., 15, pp. 229-282, 1944.

[6] KALMAN, R.E., "Linear stochastic filtering - Reappraisal and outlook", Proc. Polytechnic Inst. of Brooklyn, Symposium on System Theory, pp. 197-205, 1965.

[7] FAURRE, P., *Representation of Stochastic Processes*, Thesis, Stanford University, 1967.

[8] FAURRE, P., *Réalisations Markoviennes de Processus Stationnaires*, Report IRIA No 13, 1973.

[9] FAURRE, P., CLERGET, M., & GERMAIN, F., *Opérateurs Rationnels Positifs*, Dunod, 1979.

[10] ANDERSON, B.D.O., "The inverse problem of stationary covariance generation", J. of Stat. Physics, 1, pp. 133-147, 1973.

[11] PICCI, G., "Stochastic realization of gaussian processes", Proc. IEEE, 64, pp. 112-122, 1976.

[12] LINDQUIST, A., & PICCI, G., "On the stochastic realization problem", to appear in SIAM J. Control and Optimization.

[13] LINDQUIST, A., & PICCI, G., "A state-space theory for stationary stochastic processes", Proc. 21st Midwest Symposium on Circuits and Systems, Ames, Iowa, 1978.

[14] LINDQUIST, A., & PICCI, G., "Realization theory for multivariate stationary gaussian processes I: State space construction", to appear in the Proc. of the Fourth International Symposium on the Mathematical Theory of Networks and Systems, Delft, 1979.

[15] LINDQUIST, A., PICCI, G., & RUCKEBUSCH, G., "On minimal splitting subspaces and Markovian representations", Math. Systems Theory, to appear.

[16] RUCKEBUSCH, G., *Représentations Markoviennes de Processus Gaussiens Stationnaires,* Thesis, Un. de Paris VI, 1975.

[17] RUCKEBUSCH, G., "Représentations Markoviennes de Processus Gaussiens stationnaires et applications statistiques", pp. 115-139 in *Journées de Statistique de Processus Stochastiques* (Eds. D. Dacu (Eds. D. Dacunha-Castelle and B. van Cutsem), Proc. of Grenoble Symposium, 1977, Springer-Verlag, Lecture Notes in Mathematics, Vol. 636, 1978.

[18] RUCKEBUSCH, G., "On the theory of Markovian representation", pp. 77-88 in *Measure Theory Applications to Stochastic Analysis* (Eds. G. Kallianpur and D. Kölzow), Proc. of Oberwolfach Tagung, 1977, Springer-Verlag, Lecture Notes in Mathematics, Vol. 695, 1978.

[19] RUCKEBUSCH, G., "A geometric approach to the stochastic realization problem", submitted to SIAM J. Control and Optimization.

[20] AKAIKE, H., "Markovian representation of stochastic processes by canonical variables", SIAM J. Control, 13, pp.162-173, 1975.

[21] WILLEMS, J.C., "Recursive filtering", Statistica Neerlandica, 32, 1, pp. 1-39, 1978.

[22] SKIBINSKY, M., "Adequate subfields and sufficiency", Ann. Math. Stat., 38, pp. 155-161, 1967.

[23] HOTELLING, H., "Relations between two sets of variates", Biometrika, 28, pp. 321-377, 1936.

[24] GIHMAN, I.I., & SKOROHOD, A.V., *Stochastic Differential Equations,* Springer-Verlag, 1972.

[25] WILLEMS, J.C., "Dissipative dynamical systems", Archive for Rational Mechanics and Analysis, 45, pp. 321-392, 1972.

[26] WILLEMS, J.C., "Least squares stationary optimal control and the algebraic Riccati equation", IEEE Trans. on Aut. Control, AC-16, pp. 621-634, 1971.

[27] PAVON, M., "Stochastic realization and invariant directions of the matrix Riccati equation", submitted for publication.

[28] WILLEMS, J.C., "Time reversibility in deterministic and stochastic dynamical systems", pp. 318-326, in *Recent Developments in Variable Structure Systems, Economics and Biology* (Eds. R.R. Mohler and A. Ruberti), Proc. of US-Italy Seminar, Taormina, Sicily, 1977, Springer-Verlag, Lecture Notes in Economics and Mathematical Systems, Vol. 162, 1978.

[29] ANDERSON, B.D.O., & KAILATH, T., "Forwards, backwards and dynamically reversible Markovian models of second order processes", Proc. IEEE Int. Symp. on Circuits and Systems, 1978.

[30] GIKHMAN, I.I., & SKOROKHOD, A.V., *Introduction to the Theory of Random Processes,* Saunders, 1969.