

Stochastic Realization of Gaussian Processes

GIORGIO PICCI, MEMBER, IEEE

Abstract—A Gaussian stochastic process (y_t) with known covariance kernel is given: we investigate the generation of (y_t) by means of Markovian schemes of the type

$$dx_t = F(t) x_t dt + dw_t$$

$$y_t = H(t) x_t.$$

Such a generation of (y_t) as the “output of a linear dynamical system driven by white noise” is possible under certain finiteness conditions. In fact, this was shown by Kalman in 1965. We emphasize the probabilistic aspects and obtain an intrinsic characterization of the state of the process as the state of an externally described stochastic I/O map. Realizations of (y_t) can be constructed with respect to any increasing family of σ -fields; in particular, when the family of σ -fields is induced by the process itself, the driving white noise reduces to the innovation process of (y_t) . The corresponding realization has been referred to as the “innovation representation” of (y_t) .

INTRODUCTION

THIS WORK is concerned with some probabilistic aspects of a classical problem known as the “covariance factorization problem.”

It can be viewed as an attempt to answer, at least partially, some questions raised in a famous paper by Kalman [14].

A conceptually interesting problem was posed by Kalman in that paper, namely “... to formulate the realization theory of stochastic processes in an intrinsic way as is now possible for deterministic systems ...”

At the present state of research in the field, it seems that we have achieved a fairly good knowledge of “how to compute” realizations. There is, in fact, an extensive literature on the subject, notably due to the work of Anderson [3]–[5], Faure [9], and Kailath and Geesey [10], [11]. Yet, it does not seem completely clear how to interpret these results, especially as to what kind of “stochastic inputs” one should apply to the dynamical systems in “state space form” of the above procedures. Some progress has been made recently by Akaike [1], [2], who first investigates the concept of state for stochastic systems, getting probabilistic insight into old problems like ordering of covariances etc. This paper is, in a sense, a continuation and a generalization of his work.

I. THE IDEA OF A “STOCHASTIC STATE SPACE”

In this section, we shall investigate in some depth the central issue of the realization problem for a (Gaussian) stochastic process, namely, the axiomatic characterization of the idea of state in terms of an external, or “input-output,” description of the interaction between the process and some given flow of information.

As it will appear, a very natural way to attain this objective is to approach the problem from a probabilistic standpoint, i.e., to start with the external description having the form of a conditional probability measure.

A “concrete” theory, based on effective (experimentally measureable) descriptions of the process, like cross covariance kernels, will follow quite easily by the peculiar properties of Gaussian measures. For clarity of exposition, we postpone it to Section III after some concepts have been clarified and useful representation theory introduced.

Somewhat loosely stated, the problem is the following. Given a probability space (Ω, \mathcal{A}, P) and an R^m -valued centered Gaussian stochastic process $(y_t)_{t \in T}$, with time interval $T \subseteq R$, is it possible to express y_t as an instantaneous function of a Markov process?

Let $s \in T$, we define the σ -fields

$$\mathcal{Y}_s = \bigcap_{h>0} \sigma \{y_t; t \leq s+h\}, \quad \text{the past of } y^1 \quad (1.1)$$

$$\mathcal{Y}^s = \sigma \{y_t; t \geq s\}, \quad \text{the future.} \quad (1.2)$$

The information structure available to us will, in general, be different than just the σ -fields of events relative to (y_t) (and this is actually the case in many classical estimation problems).

The information flow we assume we are provided with will be any increasing family of sub σ -fields of \mathcal{A} , $(\mathcal{B}_t)_{t \in T}$. We interpret \mathcal{B}_s as our state of knowledge about (y_t) at time s and we refer to this σ -field simply as “the past at time s .”

Consider the family of conditional probabilities

$$P(A | \mathcal{B}_s), \quad A \in \mathcal{Y}^s \quad (1.3)$$

each $P(\cdot | \mathcal{B}_s)$, $s \in T$, can be interpreted as a mapping from the “past space” $(\Omega, \mathcal{B}_s, P)$ into the set of all probability measures over the future \mathcal{Y}^s . (This mapping represents the probabilistic analog to the deterministic I/O description.)

The idea of *splitting σ -field* introduced by McKean [17] is the appropriate approach to the problem.

Definition 1.1: \mathcal{S}_s is said to be a *splitting σ -field* if \mathcal{Y}^s and \mathcal{B}_s are conditionally independent given \mathcal{S}_s , i.e.,

$$P(A \cap B | \mathcal{S}_s) = P(A | \mathcal{S}_s) P(B | \mathcal{S}_s) \quad (1.4)$$

for all $A \in \mathcal{Y}^s$, $B \in \mathcal{B}_s$.

Equivalent definitions are [16]

$$P(A | \mathcal{B}_s \vee \mathcal{S}_s) = P(A | \mathcal{S}_s), \quad \forall A \in \mathcal{Y}^s \quad (1.5)$$

or

$$P(B | \mathcal{Y}^s \vee \mathcal{S}_s) = P(B | \mathcal{S}_s), \quad \forall B \in \mathcal{B}_s. \quad (1.6)$$

The definition is not vacuous, in fact both \mathcal{B}_s and \mathcal{Y}^s are splitting σ -fields. Of course, they are too large for our purposes, and to construct a meaningful theory we have to look for splitting σ -fields which are as small as possible. Before discussing this point, however, we have to decide what kind of information has to be used to build up our state variables. In other words, we have to decide whether we prefer a state

constructed over the past (i.e., measurable with respect to \mathcal{B}_s) or over the future (\mathcal{Y}^s -measurable) or even over some mixture of both. The question may seem meaningless since, in the deterministic context, we are used to thinking of the state exclusively as a function of the past input paths but, in the stochastic case, there are problems where a different philosophy might be more natural (think of some class of smoothing problems, for example).

In this paper, we shall concentrate on the first alternative only. For an example of a different possibility, we quote the paper of Akaike [1] where a Markovian realization of a Gaussian process is provided, whose state space is constructed using the future of the process.

Definition 1.2: \mathcal{S}_s is splitting over the past if $\mathcal{S}_s \subseteq \mathcal{B}_s$.

If \mathcal{S}_s is splitting over the past then (1.5) can be rewritten as

$$P(A | \mathcal{B}_s) = P(A | \mathcal{S}_s), \quad \forall A \in \mathcal{Y}^s \quad (1.7)$$

and we see that \mathcal{S}_s contains all the relevant information about the past \mathcal{B}_s which is needed to determine the probability of future events in \mathcal{Y}^s .

There exists a *minimal splitting σ -field* over the past. To see this, consider the σ -field induced by all random variables $P(A | \mathcal{B}_s)$ with A varying over \mathcal{Y}^s ,

$$\mathcal{P}_s = \sigma \{P(A | \mathcal{B}_s), \quad A \in \mathcal{Y}^s\}. \quad (1.8)$$

Since all $P(A | \mathcal{B}_s)$ are \mathcal{P}_s -measurable

$$P(A | \mathcal{B}_s) = P(A | \mathcal{P}_s), \quad \forall A \in \mathcal{Y}^s \quad (1.9)$$

and in fact \mathcal{P}_s is the smallest σ -field for which this property holds. \mathcal{P}_s is (essentially) unique modulo P -null subsets.

The following properties of \mathcal{P}_s are almost immediate.

Proposition 1.3: a) All sub σ -fields of \mathcal{B}_s which are "greater than" \mathcal{P}_s are splitting, i.e.,

$$\mathcal{B}_s \supseteq \mathcal{S}_s \supseteq \mathcal{P}_s \text{ implies } P(A | \mathcal{B}_s) = P(A | \mathcal{S}_s), \quad \forall A \in \mathcal{Y}^s \quad (1.10)$$

and vice versa. b) \mathcal{P}_s includes what past and future have in common,

$$\mathcal{P}_s \supseteq \mathcal{Y}^s \cap \mathcal{B}_s \quad (1.11)$$

and for all $B \in \mathcal{Y}^s \cap \mathcal{B}_s$, $P(B | \mathcal{P}_s)$ is either equal to 0 or 1 (zero-one law). c) The intersection of any collection of σ -fields splitting over the past is again splitting.

Property a) follows directly from the definition of \mathcal{P}_s and (1.9); c) is a trivial consequence of a), and b) is proved in McKean [17]. Q.E.D.

The notion of a splitting σ -field has deep connections with that of *sufficient σ -field* known in statistics.² It is not possible to go into details here, but the connection should already be selfevident from the fact that our search for a minimal splitting field is nothing else but a *data reduction* problem.

By (1.9), the two conditional probability measures $P(\cdot | \mathcal{B}_s)$ and $P(\cdot | \mathcal{P}_s)$ coincide over (Ω, \mathcal{Y}^s) . Thus a minimal splitting field \mathcal{P}_s represents the smallest sub σ -field of \mathcal{B}_s providing exactly the same amount of information about the future paths of the process (y_t) , as \mathcal{B}_s (in other words, \mathcal{P}_s has the same meaning as a *minimal sufficient σ -field*).

properly rephrasing the theory of sufficient yesian context, the two ideas are seen to coincide [20].

Notice also that the above is the fundamental idea at the roots of Nerode's construction [18] in the deterministic theory.

For a Gaussian process, the conditional law $P(\cdot | \mathcal{B}_s)$, on \mathcal{Y}^s , depends upon the past only through the family of conditional expectations $E(y_t | \mathcal{B}_s)$, $t \geq s$. This property underlies the following basic fact.

Proposition 1.4: In the Gaussian case, the minimal splitting σ -field is induced by the family of conditional expectations $E(y_t | \mathcal{B}_s)$, $t \geq s$, i.e.,

$$\mathcal{P}_s = \sigma \{E(y_t | \mathcal{B}_s), \quad t \geq s\}. \quad (1.12)$$

A proof can be obtained by paraphrasing McKean [17, pp. 344-345].

Let us define the Gaussian space³ $H(\mathcal{P}_s)$, as the closure in $L^2(\Omega, \mathcal{A}, P)$ of the linear span over the random variables $E(y_t | \mathcal{B}_s)$, $t \geq s$,

$$H(\mathcal{P}_s) = \overline{\text{sp}} \{E(y_t^i | \mathcal{B}_s), \quad t \geq s, \quad i = 1 \cdots m\}. \quad (1.13)$$

Of course, we may construct a whole family of such spaces, corresponding to each $s \in T$.

Let us consider $H(\mathcal{P}_\tau)$, $\tau \geq s$. Then, for any random variable $x_\tau \in H(\mathcal{P}_\tau)$ we have

$$E(x_\tau | \mathcal{B}_s) = E(x_\tau | \mathcal{P}_s). \quad (1.14)$$

In fact, (1.14) is true for the generators $E(y_t^i | \mathcal{B}_\tau)$, $t \geq \tau$, since

$$E(E(y_t^i | \mathcal{B}_\tau) | \mathcal{B}_s) = E(y_t^i | \mathcal{B}_s) = E(y_t^i | \mathcal{P}_s) \quad (1.15)$$

by (1.9), and

$$E(y_t^i | \mathcal{P}_s) = E(E(y_t^i | \mathcal{B}_\tau) | \mathcal{P}_s) \quad (1.16)$$

because of $\mathcal{B}_\tau \supseteq \mathcal{B}_s \supseteq \mathcal{P}_s$.

We come now to the fundamental result of this section.

Theorem 1.5: For each $\tau \geq s$; the minimal splitting field \mathcal{P}_τ is conditionally independent of \mathcal{B}_s , given \mathcal{P}_s . In other words

$$P(\Gamma | \mathcal{B}_s) = P(\Gamma | \mathcal{P}_s), \quad \forall \Gamma \in \mathcal{P}_\tau \quad (1.17)$$

or, what is the same $E(\varphi_\tau | \mathcal{B}_s) = E(\varphi_\tau | \mathcal{P}_s)$ for any random variable φ_τ in $L^2(\Omega, \mathcal{P}_\tau, P)$.

Proof: Actually, the theorem is an immediate consequence of (1.14). Consider the Gaussian space of (centered) \mathcal{B}_s -measurable random variables $H(\mathcal{B}_s)$. For any $x_\tau \in H(\mathcal{P}_\tau)$, $f_s \in H(\mathcal{B}_s)$ we can write

$$\begin{aligned} E(x_\tau f_s | \mathcal{P}_s) &= E[E(x_\tau f_s | \mathcal{B}_s) | \mathcal{P}_s] \\ &= E[E(x_\tau | \mathcal{P}_s) f_s | \mathcal{P}_s] \\ &= E(x_\tau | \mathcal{P}_s) E(f_s | \mathcal{P}_s) \end{aligned} \quad (1.18)$$

but [19 p. 28] this implies that \mathcal{P}_τ and \mathcal{B}_s are conditionally independent given \mathcal{P}_s . Q.E.D.

Theorem 1.5 says that the family of minimal splitting σ -fields $(\mathcal{P}_t)_{t \in T}$ enjoys a *Markovian property*. Again, it is interesting to notice that a similar result was proved in the theory of sufficient σ -fields by assuming a special structure for the conditional law (belonging to the so-called "expo-

³For the fundamental facts about Gaussian spaces, we refer to Neveu's monograph [19].

nential family" which includes Gaussian distributions) but independent of observations [22].

The splitting field \mathcal{P}_s is a "coordinate free" representation of the state of the process at time s . Since the idea of state in system theory is customarily associated with real numbers, we should discuss how one could give a *numerical parameterization* of \mathcal{P}_s . As there may be many such parameterizations, our first concern will be to select the more efficient ones.

Definition 1.6: A family of \mathcal{B}_s -measurable real random variables (x_s^1, \dots, x_s^N) , $N \leq \infty$, $x_s^i: (\Omega, \mathcal{B}_s, P) \rightarrow R$ constitutes a *parameterization* of \mathcal{P}_s if the family $(x_s^i, i = 1 \dots N)$ induces \mathcal{P}_s , i.e.,

$$\sigma \{ (x_s^i, i = 1, \dots, N) \} = \mathcal{P}_s. \quad (1.19)$$

Condition (1.19) says that the $x_s^i, i = 1 \dots N$, have to be constant exactly over the atoms of \mathcal{P}_s (the smallest subsets of Ω in \mathcal{P}_s which are different from \emptyset). This suggests the possibility of defining an equivalence relation over Ω whose equivalence classes coincide with the atoms of \mathcal{P}_s . Then the (x_s^i) , would be a complete system of invariants for this equivalence [20].

As for the choice of a coordinate system, the process of parametrizing \mathcal{P}_s is inherently nonunivocal. Indeed, for a given parametrization, we can construct infinitely many others. The way to do this is described by the following proposition (which is a standard statement of measure theory).

Proposition 1.7: Let $X \subseteq R^N$ be the image set of the family (x_s^1, \dots, x_s^N) and \mathcal{X} the trace over X of the Borel σ -field \mathcal{B}^N over R^N ; then a family of \mathcal{B}_s -measurable random variables (z_s^1, \dots, z_s^M) , $z_s^i: (\Omega, \mathcal{B}_s, P) \rightarrow R$ with image set $X' \subset R^M$ and σ -field $\mathcal{X}' = \mathcal{B}^M \cap X'$ is another parametrization of \mathcal{P}_s if and only if there exists a measurable bijectivity (a σ -isomorphism) φ mapping (X, \mathcal{X}) onto (X', \mathcal{X}') such that

$$z_s^i = \varphi^i(x_s^1, \dots, x_s^N), \quad i = 1, \dots, M. \quad (1.20)$$

Definition 1.8: A parameterization is said to be *canonical* if it can be reduced (by σ -isomorphism) to one with *pairwise independent* random variables.

Let (x_s^1, \dots, x_s^N) be a canonical parameterization with independent random variables. Then no one of the x_s^i can be expressed as a function of the others and hence eliminated. In fact, if we could express, say, x_s^1 as a function of x_s^2, \dots, x_s^N ,

$$x_s^1 = \psi(x_s^2, \dots, x_s^N) \quad (1.21)$$

with ψ a Borel function, then, obviously, x_s^1 would be dependent on $x_s^2 \dots x_s^N$ (in the probabilistic sense) and moreover

$$\sigma \{ (x_s^1) \} \subseteq \sigma \{ (x_s^i, i = 2, \dots, N) \} \quad (1.22)$$

which means that \mathcal{P}_s can be induced by a smaller family than $(x_s^1, x_s^2, \dots, x_s^N)$. This property is preserved under σ -isomorphism.

Proposition 1.9: We can find canonical parameterizations in $H(\mathcal{P}_s)$ (i.e., constituted by Gaussian random variables).

Proof: Take a complete set of linearly independent elements in $H(\mathcal{P}_s)$, $x_s^i, i = 1, 2, \dots, N$, ($N \leq \infty$). Then, by (1.13)

$$E(y_t^i | \mathcal{B}_s) = \lim_h \sum_{j=1}^h c_{ij} x_s^j, \quad i = 1, \dots, m \quad (1.23)$$

for all $t \geq s$. By (1.23)

$$\mathcal{P}_s = \sigma \{ E(y_t^i | \mathcal{B}_s), \quad i = 1 \dots m, t \geq s \} \\ \subseteq \sigma \{ x_s^j, \quad j = 1 \dots N \} \quad (1.24)$$

and since each x_s^j is \mathcal{P}_s -measurable

$$\mathcal{P}_s \supseteq \sigma \{ x_s^j, \quad j = 1 \dots N \}. \quad (1.25)$$

These two relationships show that $(x_s^1 \dots x_s^N)$ is actually a parameterization of \mathcal{P}_s .

On the other hand, a complete system of linearly independent elements of a Hilbert space can be orthogonalized. The orthogonalization is expressed by a measurable mapping which is easily seen to define a bijective transformation.

Notice now that orthogonal elements in $H(\mathcal{P}_s)$ are independent by Gaussianity, and so we get the conclusion. Q.E.D.

We might wonder whether a more efficient parametrization exists if we allow random variables other than Gaussian. For example, we can find canonical parametrizations in $L^2(\Omega, \mathcal{P}_s, P)$ (this notation is abbreviated to $L^2(\mathcal{P}_s)$ in the following), i.e., with "nonlinear" transformations of Gaussian variables. Do these parameterizations have fewer elements than the "linear" one? The answer is NO.

In fact, consider $A(\mathcal{P}_s)$, the algebra generated by the constant mapping 1 and the elements of $H(\mathcal{P}_s)$. $A(\mathcal{P}_s)$ is dense in $L^2(\mathcal{P}_s)$ [19, p. 143] and can be constructed starting from generators $(x_s^1 \dots x_s^N)$ forming a complete orthogonal system in $H(\mathcal{P}_s)$. $A(\mathcal{P}_s)$ cannot be generated by "less than N " generators since this would imply that at least one element in $(x_s^1 \dots x_s^N)$ can be expressed as a polynomial in the others. But this contradicts independence. Thus Proposition 1.9 defines the "natural" class of state variables for our realization problem.

Since the scope of this paper is to examine realizations with a finite-dimensional state space, we state this here as an assumption, with the intention to discuss it in Section III, where we will give the corresponding conditions on the covariance kernel of the process.

Assumption 1.10: $H(\mathcal{P}_s)$ is finite dimensional for each $s \in T$.

We can now describe the time evolution of the family of minimal splitting fields by means of the "stochastic process" $x_t = (x_t^1 \dots x_t^N)_{t \in T}$, whose random variables are constructed by taking, at each instant t , a canonical parametrization of \mathcal{P}_t (i.e., a basis in the finite-dimensional Hilbert space $H(\mathcal{P}_t)$). We stress that, by Proposition 1.7, these parameterizations are merely defined up to an arbitrary nonsingular transformation of the state space X_t .

An obvious consequence of Theorem 1.5 is the following.

Proposition 1.11: If we define the stochastic process $x_t = (x_t^1 \dots x_t^N)_{t \in T}$ by taking a basis in $H(\mathcal{P}_t)$ at each instant of time $t \in T$, then x_t is a Markov process over $\{\Omega, (\mathcal{B}_t), P\}$ with instantaneous state space (X_t, \mathcal{X}_t)

$$X_t = R^N, \quad \mathcal{X}_t = \mathcal{B}^N, \quad N = \dim H(\mathcal{P}_t) \quad (1.26)$$

and positive definite variance matrix $P(t)$

$$P(t) = [E x_t^i x_t^j]_{i,j=1 \dots N}. \quad (1.27)$$

Given an arbitrary choice of basis $(x_s^1 \dots x_s^N)$ in $H(\mathcal{P}_s)$, we can also express the generating family $(E(y_t^i | \mathcal{B}_s), i =$

$1 \cdots m, t \geq s$) as

$$E(y_t^i | \mathcal{B}_s) = \sum_{j=1}^N h_{ij}(t, s) x_s^j, \quad i = 1, \dots, m \quad (1.28)$$

for all $t \geq s$. If we collect the coefficients $h_{ij}(t, s)$ into a $m \times N$ matrix $H(t, s)$, then (1.28) assumes a more compact form

$$E(y_t | \mathcal{B}_s) = H(t, s) x_s, \quad t \geq s. \quad (1.29)$$

This factorization of $E(y_t | \mathcal{B}_s)$ could also have been derived directly from the splitting property $E(y_t | \mathcal{B}_s) = E(y_t | \mathcal{P}_s) = E(y_t | x_s)$, of which it represents a "coordinate dependent" version.

Proposition 1.12: Every vector $\xi \in X_s$ for which

$$H(t, s)\xi = 0 \quad \forall t \geq s \quad (1.30)$$

is necessarily the zero vector.

Proof: Since $H(\mathcal{P}_s)$ is the smallest Gaussian space containing all random variables $E(y_t^i | \mathcal{B}_s)$, $i = 1 \cdots m$, $t \geq s$, no vector v in $H(\mathcal{P}_s)$ different from zero can be orthogonal to every $E(y_t^i | \mathcal{B}_s)$. On the other hand, if some $\xi \in X_s$ has the property (1.30), the vector

$$v = \sum_{k=1}^N \eta_k x_s^k \quad (1.31)$$

whose coordinates are given by $\eta = P^{-1}(s)\xi$ is orthogonal to $E(y_t^i | \mathcal{B}_s)$, $i = 1 \cdots m$, $t \geq s$. In fact

$$\begin{aligned} E(E(y_t^i | \mathcal{B}_s)v) &= \sum_j h_{ij}(t, s) \sum_k E(x_s^j x_s^k) \eta_k \\ &= \sum_j h_{ij}(t, s) \xi_j \end{aligned} \quad (1.32)$$

which is zero for all i and $t \geq s$ by assumption. Hence v has all coordinates equal to zero and consequently $\xi = 0$. Q.E.D.

We remark that the proposition just proved asserts that the N columns, $h_1(\cdot, s), \dots, h_N(\cdot, s)$, of $H(\cdot, s)$; $h_i(\cdot, s): (s, +\infty) \rightarrow R^m$ are linearly independent (R^m -valued) functions.

For $t \in T$ we may express y_t as a sum of orthogonal components

$$y_t = E(y_t | \mathcal{B}_t) + w_t \quad (1.33)$$

where w_t is a Gaussian random variable orthogonal to (independent of) the past $H(\mathcal{B}_t)$. It represents the "part of y_t " which is not coupled with \mathcal{B}_t . Moreover, since $E(y_t | \mathcal{B}_t) \in H(\mathcal{P}_t)$, there exists some $m \times N$ matrix $\hat{H}(t)$ for which $E(y_t | \mathcal{B}_t) = \hat{H}(t)x_t$ with respect to some basis x_t in $H(\mathcal{P}_t)$. Thus (1.33) can be rewritten as

$$y_t = \hat{H}(t)x_t + w_t. \quad (1.34)$$

Putting $s = t$ in (1.29) and comparing with the preceding formula, we see that $\hat{H}(t) = H(t, t)$, if the basis is the same, or

$$\hat{H}(t) = H(t, t) T(t) \quad (1.35)$$

for some nonsingular ($N \times N$) matrix $T(t)$ in the general case.

Definition 1.13: If there exists a Markov process x_t in $\{\Omega, (\mathcal{B}_t), P\}$ with instantaneous state space (X_t, \mathcal{X}_t) , (X_t a finite-dimensional Euclidean space) such that the family of conditional laws $P(\cdot | \mathcal{B}_s)$, $s \in T$, relative to a Gaussian process

$(y_t)_{t \in T}$ and to an increasing family of σ -fields $(\mathcal{B}_t)_{t \in T}$, can be factored according to the diagram

$$\begin{array}{ccc} (\Omega, \mathcal{B}_s, P) & \xrightarrow{P(\cdot | \mathcal{B}_s)} & \mathcal{P}(\mathcal{Y}^s)^4 \\ x_s & \searrow & \nearrow \hat{P}_s(\cdot, \cdot) \\ & (X_s, \mathcal{X}_s) & \end{array} \quad (1.36)$$

that is, if

$$P(A | \mathcal{B}_s) = \hat{P}_s(A, x_s), \quad \forall A \in \mathcal{Y}^s \quad (1.37)$$

for some (Gaussian) transition probability $\hat{P}_s(A, \xi)$ on $(\mathcal{Y}^s \times X_s)$, then we say that the triple

$$\{x_s, (X_s, \mathcal{X}_s), \hat{P}_s(A, \xi)\}, \quad s \in T \quad (1.38)$$

is a *realization* of y relative to the family (\mathcal{B}_t) .

Then:

- i) the σ -field induced by x_s, \mathcal{B}_s , is necessarily splitting over the past;
- ii) the conditional mean $E(y_t | \mathcal{B}_s)$ has a representation of the form

$$E(y_t | \mathcal{B}_s) = H(t, s) x_s, \quad \forall t \geq s \quad (1.39)$$

with $(H(t, s))$, $t \geq s$, ($m \times N$) matrices;

- iii) we can write y_t as a sum of an instantaneous function of the process x_t plus "noise," i.e.,

$$y_t = H(t) x_t + w_t \quad (1.40)$$

where $w_t \perp \mathcal{B}_t$.

The justification of the above statements is conceptually rather simple and will be omitted. We also remark that a $\hat{P}_s(\cdot, \xi)$ satisfying (1.37) is uniquely defined as soon as $E(y_t | \mathcal{B}_s)$ can be factored in the form (1.39). This means that, the possibility of assigning a factorization of the conditional means $E(y_t | \mathcal{B}_s)$ of the type (1.39) is equivalent to the existence of (Gaussian) probability laws $\hat{P}_s(\cdot, \xi)$ on \mathcal{Y}^s , $\xi \in X_s$, satisfying (1.37) $\forall \xi = x_s(\omega)$.

Indeed, by taking $\hat{P}_s(\cdot, \xi)$, defined by the covariance kernel

$$\Lambda(t, \tau) - H(t, s)P(s)H'(\tau, s), \quad P(s) = E x_s x_s' \quad (1.41)$$

for $t, \tau \geq s$ (here Λ is the covariance of y), and mean

$$\int_{\Omega} y_t(\omega) \hat{P}_s(d\omega, \xi) = H(t, s)\xi, \quad t \geq s \quad (1.42)$$

we may easily check that (1.41) and (1.42) coincide with the (conditional) covariance and mean of $P(\cdot | \mathcal{B}_s)$.

Therefore, any realization (1.38) uniquely defines a factorization of the type (1.39) and vice versa.

We may hence legitimately call also the triple

$$\{x_s, (X_s, \mathcal{X}_s), (H(t, s))_{t \geq s}\}, \quad s \in T \quad (1.43)$$

a *realization* of (y_t) relative to (\mathcal{B}_t) .

Definition 1.14: A realization of (y_t) relative to (\mathcal{B}_t) is said to be

- i) *reachable at time s* , if x_s is a canonical parameterization of the induced splitting field \mathcal{B}_s ;

⁴ $\mathcal{P}(\mathcal{Y}^s)$ is the set of all probability measures over (Ω, \mathcal{Y}^s) .

- ii) *Observable at time s* , if $H(t, s)\xi = 0$ for some $\xi \in X_s$ and all $t \geq s$ implies that $\xi = 0$.

Remarks: Since there is a positive probability for the values of a Gaussian random variable to fall into any given Borel subset of R with positive Lebesgue measure, the image of the mapping $x_s: (\Omega, \mathcal{B}_s, P) \rightarrow X_s$ with $x_s^i \in H(\mathcal{S}_s)$ is a vector subspace of X_s . Its dimension equals the number of independent components of x_s . Thus if a realization is reachable at time s , x_s is onto as in the deterministic theory.

Observability at time s means that the mapping $\xi \mapsto \hat{P}_s(\cdot, \xi)$ is *one to one*; i.e., for each law determining the future evolution of the process, there corresponds exactly one value of the parameter ξ . This intuitively agrees with the following property, consequence of (1.12).

Proposition 1.15: For any realization $\{x_s, (X_s, \mathcal{X}_s), (H(t, s))_{t \geq s}\}$, which is observable at time s , x_s is a parameterization of the minimal splitting field \mathcal{P}_s .

The proof can be obtained from ii) and (1.39), by reversing the arguments used in the proof of Proposition 1.12.

Summing up Propositions 1.11 and 1.12, we have the final result of this section.

Theorem 1.16: Every Gaussian process satisfying assumption (1.10) has a realization which is both reachable and observable at each instant of time $s \in T$. For such a *minimal* realization

$$\{\hat{x}_s, (\hat{X}_s, \mathcal{X}_s), (\hat{H}(t, s))_{t \geq s}\}, \quad s \in T \quad (1.44)$$

\hat{x}_s is a canonical parameterization of the minimal splitting field \mathcal{P}_s . Any two minimal realizations can be obtained one from the other by means of a (nonsingular) transformation of basis in $H(\mathcal{P}_s)$.

II. GENERALIZED HIDA-CRAMER REPRESENTATION

Let us consider any increasing family of σ -fields $(\mathcal{B}_t)_{t \in T}$. As we already noticed, the process $(y_t)_{t \in T}$ can be split up additively

$$y_t = z_t + w_t, \quad t \in T \quad (2.1)$$

into the sum of a component $z_t = E(y_t | \mathcal{B}_t)$ which is \mathcal{B}_t -measurable and a component $w_t = y_t - z_t$ independent of \mathcal{B}_t . In this section, we shall deal only with the part of (y_t) coupled with (\mathcal{B}_t) . Notice that

$$\sigma\{z_s, s \leq t\} \subseteq \mathcal{B}_t, \quad \forall t \in T \quad (2.2)$$

so we may, without loss of generality, assume that the σ -field, included by the original process (y_t) up to the time t , is a sub σ -field of \mathcal{B}_t for all t and forget about the process (w_t) .

Our main concern will be to derive a kind of representation of the "input" family (\mathcal{B}_t) by means of a process with orthogonal increments (the *innovation process* of (\mathcal{B}_t)). We shall see that this can be done for any family (\mathcal{B}_t) in an essentially unique way. Moreover we shall exploit this representation for expressing (y_t) as the output of a causal linear dynamical system driven by the above innovation process representative of (\mathcal{B}_t) .

This representation generalizes the so called Hida-Cramer representation [13], where a process (y_t) is considered evolving in time as a function of its own past.

The behavior of the family (\mathcal{B}_t) as a function of time is important in what follows. At a point $t \in T$, (\mathcal{B}_t) may have both

right and left discontinuities, i.e., the σ -fields

$$\mathcal{B}_{t-} = \bigvee_{s < t} \mathcal{B}_s, \quad \mathcal{B}_t, \quad \mathcal{B}_{t+} = \bigcap_{s > t} \mathcal{B}_s \quad (2.3)$$

may all be different. In order to work with a more manageable object, we will modify (\mathcal{B}_t) in such a way as to obtain a right continuous family. We define

$$\hat{\mathcal{B}}_t = \bigcap_{s > t} \mathcal{B}_s, \quad t \in T \quad (2.4)$$

in this way $\hat{\mathcal{B}}_t \supseteq \mathcal{B}_t$, and $(\hat{\mathcal{B}}_t)$ has right limit equal to $\hat{\mathcal{B}}_t$ everywhere. We shall always work with the family $(\hat{\mathcal{B}}_t)$ hereafter, and this will dispense us from the use of the hatted symbol.

Let us take any bounded interval $(t_0, t_1] \subseteq T$ and define the Hilbert spaces⁵

$$H = H(\mathcal{B}_{t_1}) \ominus H(\mathcal{B}_{t_0}) \quad (2.5)$$

$$H_t = H(\mathcal{B}_t) \ominus H(\mathcal{B}_{t_0}), \quad t_0 < t \leq t_1. \quad (2.6)$$

We recall that H_t is the Gaussian space of all \mathcal{B}_t -measurable random variables independent of \mathcal{B}_{t_0} . Then $H_t \subseteq H$ for all $t \leq t_1$ and we may define a family of projection operators from H onto H_t

$$P_t: H \longrightarrow H_t, \quad t_0 < t \leq t_1. \quad (2.7)$$

The family (P_t) is constructed in such a way that

- i) P_t is right continuous on $t_0 \leq t < t_1$;
- ii) $\lim_{t \downarrow t_0} P_t = 0$ (the projection over $\{0\}$);
- iii) $P_{t_1} = I$ (the identity).

These properties are standard consequences of the right continuity of (\mathcal{B}_t) (actually iii) is true by definition).

$(P_t)_{t_0 \leq t \leq t_1}$ is a so-called *resolution of the identity* over H . Any resolution of the identity defines, in a unique way, a *spectral measure* P defined on the Borel field over $[t_0, t_1]$, with values that are projections over subspaces of H . The defining relationship for P is the natural one, i.e.,

$$P((a, b]) = P_b - P_a, \quad t_0 \leq a \leq b \leq t_1. \quad (2.8)$$

Let us consider the bounded self-adjoint operator

$$A = \int_{t_0}^{t_1} s dP_s. \quad (2.9)$$

As we know, the spectrum of A , $\sigma(A)$ coincides with the support of the spectral measure P (the complementary in $[t_0, t_1]$ of the union of all open sets having 0 measure, see e.g., Halmos [12]). Moreover, each jump point for P_t in $[t_0, t_1]$ is an eigenvalue of A . These properties can be interpreted in terms of σ -fields (\mathcal{B}_t) .

a) Any point $t \in [t_0, t_1]$ for which the information flow is strictly increasing, i.e.,

$$\mathcal{B}_{t-h} \subset \mathcal{B}_{t+h} \quad (\text{strict inclusion}) \quad (2.10)$$

for all $h > 0$, belongs to the spectrum.

b) Every point of jump for (\mathcal{B}_t) is an eigenvalue of A . (A point of jump t is such that $\mathcal{B}_{t-} \neq \mathcal{B}_t$).

⁵ If H_1 is a Hilbert subspace of H then $H_2 \triangleq H \ominus H_1$ is the orthogonal complement of H_1 in H , i.e., $H = H_1 \oplus H_2$.

Suppose A is a *cyclic operator* in H . Then there exists some $z \in H$ for which

$$\overline{\text{sp}} \{A^n z, \quad n \geq 0\} = H \quad (2.11)$$

z is called a *cyclic vector* for A . For any cyclic z , condition (2.11) implies that

$$\overline{\text{sp}} \{P_t z, \quad t_0 \leq t \leq t_1\} = H. \quad (2.12)$$

Formula (2.12) has a very important probabilistic meaning. Let us define a stochastic process (z_t) on $[t_0, t_1]$ by putting

$$z_t = P_t z, \quad t_0 \leq t \leq t_1. \quad (2.13)$$

Then (z_t) is a *right continuous orthogonal increments process spanning the Gaussian space H* .

In fact,

$$E(z_t - z_s)z_s = E(P(s, t)z \cdot P_s z) = 0 \quad (2.14)$$

shows that (z_t) has orthogonal increments and formula (2.12) that (z_t) generates H .

The variance function of (z_t) is

$$F(t) = E z_t^2 = \langle P_t z, P_t z \rangle = \langle P_t z, z \rangle \quad (2.15)$$

(we use probabilistic and vector space symbols jointly in order to stress the probabilistic meaning of the latter) and coincides with the distribution function of the measure

$$\mu_z(\Delta) = \langle P(\Delta)z, z \rangle, \quad \Delta: \text{Borel set in } [t_0, t_1]. \quad (2.16)$$

The support of μ_z is precisely the support of the spectral measure P (this is a classical argument: for any open Δ , $\mu_z(\Delta) = 0 \iff P(\Delta)z = 0$, but $P(\Delta)z = 0$ implies $P(\Delta) = 0$ since $P(\Delta)A^n z = A^n P(\Delta)z$ and $(A^n z)$ span the space). Thus since (z_t) jumps at t if and only if $F(t)$ jumps there [7 p. 425] and has zero increment in the same intervals as $F(t)$, we conclude that

- i) (z_t) jumps exactly at the jumping points of (\mathfrak{B}_t) ;
- ii) (z_t) has a strictly increasing variance (i.e., it is nonconstant) for all $t \in \sigma(A)$, i.e., iff the family (\mathfrak{B}_t) is strictly increasing.

Since H is spanned by the vectors $(A^n z, n = 0, 1, 2, \dots)$ we can show that any $x \in H$ can be represented in an essentially *unique* way as

$$x = \varphi(A)z \quad (2.17)$$

where $\varphi \in L^2(\mu_z)$. By definition of $\varphi(A)$ as $\int \varphi(s) dP_s$ and (2.17), we can also get the representation of x as a "stochastic integral"

$$x = \int_{t_0}^{t_1} \varphi(s) dP_s \cdot z = \int_{t_0}^{t_1} \varphi(s) dz_s. \quad (2.18)$$

(Of course the last member in (2.18) can also be defined directly by means of a classical procedure [7, p. 426].)

If the vector x in (2.18) belongs to the subspace H_t then $P_t x = x$ and

$$P_t x = \int_{t_0}^t \varphi(s) dz_s = \int_{t_0}^{t_1} \varphi(s) dz_s = x \quad (2.19)$$

so that $\int_t^{t_1} \varphi^2(s) d\mu_z = 0$ which implies $\varphi(s) = 0$, μ_z -a.s. for all $s > t$. Actually, it is more accurate to denote explicitly the kernel function φ corresponding to $x \in H_t$ as $\varphi(t, s)$. By

the last remark, $\varphi(t, s) = 0$, $s > t$. As we have seen, the representation in (2.18), i.e., φ , is unique (in $L^2(\mu_z)$) for a given cyclic vector z . But there may be many cyclic z 's.

Let us take two of such vectors z_1, z_2 and denote by $(z_1(t))$ and $(z_2(t))$ the corresponding orthogonal increments processes. Then, for any $t_0 < t \leq t_1$

$$z_1(t) = \int_{t_0}^t \varphi(t, s) dz_2(s) \quad (2.20)$$

and since $P_\tau z_1(t) = z_1(\tau) \forall t \geq \tau$, we have

$$\int_{t_0}^\tau \varphi(t, s) dz_2(s) = \int_{t_0}^\tau \varphi(\tau, s) dz_2(s) \quad (2.21)$$

which implies $\varphi(t, s) = \varphi(\tau, s)$ a.s. $\forall t \geq \tau \geq s \geq t_0$.

In conclusion, φ does not depend on t . Since for any cyclic z $\text{supp } \mu_z = \text{supp } P$, all measures corresponding to cyclic vectors are equivalent (i.e., mutually absolutely continuous). This implies

$$\varphi(s) > 0, \quad \mu_z - \text{a.s.} \quad (2.22)$$

By repeating the above argument with z_2 in place of z_1 , we can easily show that $1/\varphi(s) > 0$ μ_{z_1} -a.s. and square integrable (with respect to μ_{z_1}).

We summarize all the above discussion in the following proposition.

Proposition 2.1: Let (\mathfrak{B}_t) be any increasing right continuous family of σ -fields on $t_0 \leq t \leq t_1$ and let A in (2.9) be a cyclic operator. Then there exists an *orthogonal increments process* $(z_t)_{t_0 \leq t \leq t_1}$ with the following properties.

- a) (z_t) is a process continuous in mean square from the right, generating the family $\mathfrak{B}_t \ominus \mathfrak{B}_{t_0}^6, t \geq t_0$.
- b) (z_t) has a strictly increasing variance at the point s iff (\mathfrak{B}_t) is strictly increasing at s . In particular (z_t) has a jump at time s iff the family (\mathfrak{B}_t) jumps at s .
- c) Every random variable $x \in H_t, t_1 \geq t > t_0$ has a unique representation as

$$x = \int_{t_0}^t \varphi(t, s) dz_s \quad (2.23)$$

with $\varphi(t, \cdot) \in L^2(\mu_z), \varphi(t, s) = 0, s > t$.

d) Every other orthogonal increments process $(w_t)_{t_0 \leq t \leq t_1}$ with property a) (and thus with the properties b) and c)) can be obtained from (z_t) by means of an invertible "scale factor" q

$$w_t = \int_{t_0}^t q(s) dz_s \quad (2.24)$$

with $q \in L^2(\mu_z), q(s) > 0, \mu_z$ -a.s., $1/q \in L^2(\mu_w)$. The (equivalence class of) orthogonal increments process(es) (z_t) ((w_t)) is the *innovation process* of (\mathfrak{B}_t) over the interval $(t_0 \leq t \leq t_1)$.

A cyclic operator has a spectrum of *multiplicity one* [12]. In general, the self-adjoint operator A will have multiplicity

⁶ If $\mathfrak{B}_1 \subseteq \mathfrak{B}$, then $\mathfrak{B}_2 \triangleq \mathfrak{B} \ominus \mathfrak{B}_1$ is such that, i) \mathfrak{B}_2 is independent of \mathfrak{B}_1 with respect to the Gaussian measure P , and ii) $\mathfrak{B}_1 \vee \mathfrak{B}_2 = \mathfrak{B}$.

greater than one. The intuitive meaning of this fact is that the family (\mathcal{B}_t) has a "rate of growth" with time, which is faster than that corresponding to the family of σ -fields induced by a single (Gaussian) orthogonal increments process. In other words, to carry the information flow corresponding to (\mathcal{B}_t) , one innovation process is not enough and we have to use many. The minimal number M of orthogonal increments processes, possessing property a) of the preceding Proposition, is said to be the *multiplicity* of the family (\mathcal{B}_t) . It can be shown that M coincides with the spectral multiplicity of A . There is no place to give details here; the fundamental mathematical facts can be found in [8, vol. II, p. 909]. The stochastic interpretation is more or less classical if for (\mathcal{B}_t) , we take the past of the process (y_t) , i.e., the family defined by (1.1) (see [6], [13], [15], and [21]).

Theorem 2.2: For any right continuous family of σ -fields (\mathcal{B}_t) on $t_0 \leq t \leq t_1$, there exists $M(\leq \infty)$ orthogonal increments (Gaussian) processes $(z_t^1, \dots, z_t^M)_{t_0 \leq t \leq t_1}$, such that

A₁) Each (z_t^n) is right continuous in mean square and

$$\sigma \{z_s^n; n = 1 \dots M, s \leq t\} = \mathcal{B}_t \ominus \mathcal{B}_{t_0} \quad (2.25)$$

for all $t_0 \leq t \leq t_1$.

A₂) (z_t^1, \dots, z_t^M) are mutually orthogonal, i.e., $E z_t^i z_t^j = 0$ if $i \neq j$ for all $t, s \in [t_0, t_1]$, and

$$\begin{aligned} H_t &= \overline{\text{sp}} \{z_s^n; n = 1 \dots M, s \leq t\} = \bigoplus_{n=1}^M \overline{\text{sp}} \{z_s^n; s \leq t\} \\ &= \bigoplus_{n=1}^M H_t(z^n). \end{aligned} \quad (2.26)$$

B) The Borel measures $\mu_n, n = 1 \dots M$, defined by

$$\mu_n((t_0, t]) = E(z_t^n)^2 = \langle P_t z^n, z^n \rangle \quad (2.27)$$

can be ordered by absolute continuity, in the sense that $\mu_1 \gg \mu_2 \dots \gg \mu_M$. More precisely, if we denote by S_n the support of μ_n , we have the inclusion property

$$\text{supp } P = S_1 \supseteq S_2 \dots \supseteq S_M. \quad (2.28)$$

This means that $z_t^i, i = 1, \dots, M$ have increasing variance at time s iff $s \in S_i, i = 1 \dots M$.

C) Every random variable $x \in H_t$ has a unique representation as the orthogonal sum in H_t

$$x = \sum_{n=1}^M \int_{t_0}^t \varphi_n(t, s) dz_s^n, \quad t_0 \leq t \leq t_1 \quad (2.29)$$

where $\varphi_n(t, \cdot) \in L^2(\mu_n)$ $t_0 \leq t \leq t_1$, $\varphi_n(t, s) = 0, s > t \forall n = 1 \dots M$.

D) Any other M' -ple of orthogonal increments processes $(w_t^1, \dots, w_t^{M'})_{t_0 \leq t \leq t_1}$, satisfying A)-B) is such that $M' = M$ and moreover,

$$w_t^n = \int_{t_0}^t q_n(s) dz_s^n, \quad n = 1 \dots M, \quad t_0 \leq t \leq t_1 \quad (2.30)$$

with $q_n \in L^2(\mu_n)$, $q_n > 0$ μ_n -a.s. and $1/q_n \in L^2(\mu_n')$, μ_n' being the Borel measure (equivalent to μ_n) induced by the variance of w_t^n .

Let $(y_t)_{t \in T}$ be any Gaussian process adapted to (\mathcal{B}_t) (i.e., y_t is \mathcal{B}_t -measurable for each $t \in T$). We take the projection of y_t over $H(\mathcal{B}_{t_0})$ for all $t > t_0$. Let us denote this projection by

the symbol $P_{t_0} y_t$; then the difference $y_t - P_{t_0} y_t$ belongs to H_t for all $t > t_0$, and in virtue of Theorem 2.2, we can write

$$y_t = P_{t_0} y_t + \sum_1^M \int_{t_0}^t \varphi_n(t, s) dz_s^n \quad (2.31)$$

where all elements in the sum are pairwise orthogonal.

If y_t is vector valued, say $y_t = (y_t^1, \dots, y_t^m)$ then the preceding formula is valid for each component separately. We may introduce vector notations

$$z_t = [z_t^1, \dots, z_t^M]'$$

$$W(t, s) = [\varphi_n^{(i)}(t, s)], \quad i = 1 \dots m, \quad n = 1 \dots M \quad (2.32)$$

and rewrite (2.31) as

$$y_t = P_{t_0} y_t + \int_{t_0}^t W(t, s) dz_s, \quad t \geq t_0. \quad (2.33)$$

This formula is the fundamental result of this section. It says that every Gaussian process can be represented as the output of a causal linear dynamical system whose input is the (generalized) innovation process of any increasing family of σ -fields with respect to which the process is adapted.

The random variable y_t appears as the sum of two (orthogonal) terms.

i) A "free evolution" term $P_{t_0} y_t = E(y_t | \mathcal{B}_{t_0}), t \geq t_0$, which depends only upon the past history (i.e., the \mathcal{B}_{t_0} -measurable part) of the process.

Let $a = \inf \{t, t \in T\}$ be the left endpoint of the interval T . Then \mathcal{B}_a may or may not be the trivial σ -field $\{\Omega, \phi\}$. If it is, then $\lim_{t_0 \downarrow a} P_{t_0} y_t = 0$ for all t , and the effect of the "remote

past" over the present is zero. Taking limits in (2.33) as $t_0 \downarrow a$, we obtain y_t expressed as a stochastic integral (a "forced response") over the entire past history of (\mathcal{B}_t) . If \mathcal{B}_a is not trivial then, two possible situations may occur:

1) $a > -\infty$. In this case we extend (\mathcal{B}_t) to all $t \in (-\infty, a)$ by putting it equal to the trivial σ -field. The extended family turns out to introduce a jump at a , which implies that z_t has a jump at the same time. By means of this transfer of singularities, the problem is reduced to the previous one and the "free evolution" term disappears as soon as we take $t_0 < a$ in (2.33).

2) $a = -\infty$. The preceding trick fails to work and we have to take into account the part of y_t which depends upon the infinitely remote past. This contribution, i.e., the projection $P_{-\infty} y_t = E(y_t | \mathcal{B}_{-\infty})$ is commonly referred to as the "deterministic part" of y (in case $\mathcal{B}_t = \mathcal{B}_t$).

ii) A "forced response" individuated by the impulse response matrix $W(t, s)$ (which is zero for $s > t$). There is a slight indeterminacy in $W(t, s)$ since for an equivalent (in the sense explained in D) innovation process w_t , we have to change $W(t, s)$ to $W(t, s)Q^{-1}(s)$, where $Q(s) = \text{diag } q_1(s) \dots q_M(s)$ is such that

$$dw_t = Q(t) dz_t. \quad (2.34)$$

This indeterminacy disappears if we agree to "normalize" (z_t) in some way.

III. STATE EQUATIONS FOR A STOCHASTIC PROCESS

The representation theory we have just developed is a very natural starting point for discussing the structure of input-state and state-output equations which were obtained at a rather abstract level in Section I.

In the following, we shall always make the assumption that we can choose \mathcal{B}_a as the trivial σ -field (i.e., (y_t) is a "purely nondeterministic" process). The assumption is not an essential one and can be removed at the price of additional complexity in some of the formulas. As a rule, N will always denote a (finite) natural number. Let us first recall the following fundamental property of the representation (2.33).

Proposition 3.1: For all $t \geq s \geq t_0$, the following formula holds:

$$E(y_t | \mathcal{B}_s) = P_s y_t = \int_{t_0}^s W(t, \sigma) dz_\sigma + P_{t_0} y_{t_0}. \quad (3.1)$$

By (1.29), the first member of (3.1) can be written as $H(t, s)x_s$, with $x_s = (x_s^1 \cdots x_s^N)$ random variables spanning $H(\mathcal{P}_s) \subseteq H(\mathcal{B}_s)$. This last inclusion relationship implies that x_s has a representation

$$x_s = \int_{t_0}^s \Gamma(s, \sigma) dz_\sigma + P_{t_0} x_{t_0} \quad (3.2)$$

for some matrix Γ of dimension $(N \times M)$ whose ij th entry belongs to $L^2(\mu_j)$, $j = 1 \cdots M$ for all $i = 1 \cdots N$.

By comparing (1.29), (3.1), and (3.2) we get immediately the factorization property

$$W(t, \sigma) = H(t, s)\Gamma(s, \sigma) \quad (3.3)$$

which holds (a.s. with respect to σ) for all $t \geq s \geq \sigma > a$.

If for x_s we take a canonical parametrization of \mathcal{P}_s then it is easy to see that the linear operator $(N \times N$ matrix)

$$P(s) = E x_s x_s' = \int_a^s \Gamma(s, \sigma) dF(\sigma) \Gamma'(s, \sigma) \quad (3.4)$$

is positive definite. The last symbol on the right side represents a $(N \times N)$ matrix with entries

$$P_{ij}(s) = \sum_{k=1}^M \int_a^s \Gamma_{ik}(s, \sigma) \Gamma_{jk}(s, \sigma) d\mu_k. \quad (3.5)$$

Positive definiteness of $P(s)$ entails that the rows of $\Gamma(s, \cdot)$ have to be linearly independent as elements of the function space $L^2((a, s], \mu_1) \times \cdots \times L^2((a, s], \mu_M)$.

Theorem 3.2: The Gaussian process (y_t) admits a minimal realization with respect to (\mathcal{B}_t) which at time s has dimension N , if and only if the kernel $W(t, \sigma)$ can be factorized as $W(t, \sigma) = H(t, s)\Gamma(s, \sigma)$ for all $t \geq s \geq \sigma > a$, where

$H(\cdot, s) : [s, \infty) \rightarrow R^{m \times N}$ is a matrix whose column functions

$$\begin{bmatrix} h_{11}(\cdot, s) \\ \vdots \\ h_{m1}(\cdot, s) \end{bmatrix}, \dots, \begin{bmatrix} h_{1N}(\cdot, s) \\ \vdots \\ h_{mN}(\cdot, s) \end{bmatrix} \quad (3.6)$$

are linearly independent,

$\Gamma(s, \cdot) : (a, s] \rightarrow R^{N \times M}$ is a matrix whose rows

$$[\Gamma_{11}(s, \cdot) \cdots \Gamma_{1M}(s, \cdot)], \dots, [\Gamma_{N1}(s, \cdot), \dots, \Gamma_{NM}(s, \cdot)] \quad (3.7)$$

belong to $L^2((a, s], \mu_1) \times \cdots \times L^2((a, s], \mu_M)$ and are linearly independent as members of this space.

Proof: The necessity was discussed before the statement of Theorem 3.2. Sufficiency follows straightforwardly by defining x_s through (3.2) (with $t_0 = a$) and by checking that the resulting factorization of $E(y_t | \mathcal{B}_s)$ which is of the form (1.29) satisfies the observability and reachability conditions. Q.E.D.

The above result is the stochastic analog of Kalman's factorization criterion [23].

Clearly, our result is not stated in such form so as to allow for a direct check of the condition. It is, in fact, plausible to think that the original data describing the interaction of (y_t) and (\mathcal{B}_t) will, in general, be in the form of a cross covariance kernel rather than a Hida-Cramer representation (which has to be obtained via "spectral factorization" procedures).

Let (u_t) be any process inducing the increasing family (\mathcal{B}_t) (there may be infinitely many such processes, e.g., the innovation of (\mathcal{B}_t) is one) and taking values in R^p . We assume that the $(m \times p)$ cross covariance kernel

$$\Lambda(t, \tau) = E(y_t u_\tau'), \quad t, \tau > a \quad (3.8)$$

constitutes the original description of our problem.

Proposition 3.3: The process (y_t) admits a realization with respect to (\mathcal{B}_t) which at time s has minimal dimension equal to N if and only if $\Lambda(t, \tau)$ factorizes according to

$$\Lambda(t, \tau) = H(t, s)Q'(s, \tau), \quad t \geq s \geq \tau \quad (3.9)$$

where

the matrix $H(\cdot, s)$ has N columns which are linearly independent (R^m -valued) functions over $[s, \infty)$;
the matrix $Q(s, \cdot)$ has N linearly independent (R^p -valued) column functions over $[a, s]$.

Proof: Let

$$u_t = \int_a^t U(t, \sigma) dz_\sigma, \quad t > a \quad (3.10)$$

be the Hida-Cramer representation of (u_t) with respect to (\mathcal{B}_t) . Since (u_t) induces (\mathcal{B}_t) the kernel $U(t, \sigma) = [u_{ij}(t, \sigma)]$ has the following canonical property [13]: any vector f in $L^2((a, s], \mu_1) \times \cdots \times L^2((a, s], \mu_M)$ for which

$$\int_a^t U(t, \sigma) f(\sigma) dz_\sigma = 0, \quad \forall t \in (a, s] \quad (3.11)$$

is necessarily the zero vector, i.e., $f_i(\sigma) = 0$ a.s. μ_i , on $(a, s]$, $i = 1 \cdots M$ (this means that the rows of $U(s, \cdot)$ span the space; an immediate consequence of the spectral representation theorem).

Necessity is proved as follows. Let us take a minimal factorization of (y_t) of the form (1.29). Then

$$\begin{aligned} \Lambda(t, \tau) &= E(P_s y_t, u_\tau') = E\left(\int_a^s \int_a^\tau W(t, \sigma) dz_\sigma dz_\eta' U'(\tau, \eta)\right) \\ &= H(t, s) \int_a^\tau \Gamma(s, \sigma) dF(\sigma) U'(\tau, \sigma), \quad t \geq s \geq \tau \end{aligned} \quad (3.12)$$

which is valid for all τ in $(a, s]$. The integral in the last member of (3.12) is defined in a similar way as in (3.4); it is an $N \times p$ matrix which we shall denote by $Q'(s, \tau)$. All that we have to show is that the columns of $Q(s, \cdot)$ are linearly independent (R^p -valued) functions on $(a, s]$.

Suppose they were not. Then we could find N nonzero real numbers, say $(\alpha_j, j = 1 \cdots N)$ such that

$$\sum_{j=1}^N \alpha_j q_j(s, \cdot) = 0 \text{ on } (a, s] \quad (3.13)$$

where $q_j(s, \cdot)$ denotes the j th column of $Q(s, \cdot)$.

Taking into account the definition of $Q(s, \cdot)$, we see that (3.13) can be explicitly rewritten as

$$\sum_{k=1}^M \int_a^\tau u_{ik}(\tau, \sigma) \left[\sum_{j=1}^N \alpha_j \Gamma_{jk}(s, \sigma) \right] d\mu_k = 0 \quad (3.14)$$

for all $i = 1 \cdots p$ and for all $\tau \in (a, s]$. In short, there is a linear combination of the columns of $\Gamma'(s, \cdot)$, let us call it f , for which (3.11) is true. Then f has to be the zero element in $L^2((a, s], \mu_1) \times \cdots \times L^2((a, s], \mu_M)$ which is contradicted by Theorem 3.2.

Sufficiency: In Section I, we proved that $E(y_t | \mathcal{B}_s)$ has a factorization of the form (1.29) which is both reachable and observable. Let $P_s y_t = M(t, s) x_s$ be such a factorization with $M(t, s)$ of dimension $(m \times N')$.

Repeating the preceding calculations, we end up with a factorization of $\Lambda(t, \tau)$ of the form

$$\Lambda(t, \tau) = M(t, s) R'(s, \tau), \quad t \geq s \geq \tau \quad (3.15)$$

where $R'(s, \cdot)$ is $(N' \times p)$ and, by reachability, has N' linearly independent columns on $(a, s]$.

If we compare (3.9) and (3.15), by standard linear algebra, we see that the columns of $H(\cdot, s)$ and $M(\cdot, s)$ must span isomorphic vector spaces and hence be related through a nonsingular transformation.

Thus $N = N'$ and

$$\begin{aligned} M(t, s) &= H(t, s) T(s), & \forall t \geq s \\ R'(s, \tau) &= T^{-1}(s) Q'(s, \tau), & \forall \tau \leq s \end{aligned} \quad (3.16)$$

for some nonsingular $T(s)$. Q.E.D.

We believe that the above result is important in that it shows that a "time-varying" state space is possible and provides some insight into it.

Corollary 3.4: Suppose $\Lambda(t, \tau)$ factorizes "uniformly", i.e.,

$$\Lambda(t, \tau) = A(t) B'(\tau), \quad t \geq \tau \quad (3.17)$$

where $A(\cdot)$ and $B(\cdot)$ are matrices with N linearly independent columns on every subinterval of (a, ∞) (this happens, for example, if the entries of A and B are analytic functions of t); then the dimension of a minimal realization is N and is independent of time.

In fact, for all s , we may put $H(t, s) = A(t)$ for $t \geq s$ and $Q(s, \tau) = B(\tau)$ for $\tau \leq s$.

In Section I, we proved that the process (x_t) defined by taking a basis in $H(\mathcal{P}_t)$ at each instant of time is Markovian with respect to the family (\mathcal{B}_t) . This means that the conditional expectation $E(x_t | \mathcal{B}_s)$ has to depend upon x_s only, and moreover, by Gaussianity, this dependence has to be linear, i.e.,

$$E(x_t | \mathcal{B}_s) = \Phi(t, s) x_s \quad (3.18)$$

for some (in general, nonsquare) matrix $\Phi(t, s)$.

Proposition 3.5: $\Phi(\tau, s)$, defined for all $\tau \geq s$ by formula (3.18), is a homomorphism of X_s into X_τ satisfying the fol-

lowing conditions:

$$\Phi(t, \tau) \Phi(\tau, s) = \Phi(t, s), \quad t \geq \tau \geq s \quad (3.19)$$

$$\Phi(t, t) = I, \quad t \in T \quad (3.20)$$

$$H(t, s) = H(t, \tau) \Phi(\tau, s), \quad t \geq \tau \geq s \quad (3.21)$$

$$\Gamma(\tau, \sigma) = \Phi(\tau, s) \Gamma(s, \sigma), \quad \tau \geq s \geq \sigma \quad (3.22)$$

the last relationship holding a.s. on $(a, s]$.

Proof: The first assertion follows from the orthogonal decomposition of x_τ

$$x_\tau = \int_s^\tau \Gamma(\tau, \sigma) dz_\sigma + \Phi(\tau, s) x_s \quad (3.23)$$

and by taking scalar products with vectors of $H(\mathcal{P}_s)$.

The Properties (3.19)–(3.22) are fairly obvious, (3.19) and (3.20) follow from

$$E(x_t | \mathcal{B}_s) = E(E(x_t | \mathcal{B}_\tau) | \mathcal{B}_s) = E(\Phi(t, \tau) x_\tau | \mathcal{B}_s) \quad (3.24)$$

(3.21) is a consequence of

$$P_s y_t = E(E(y_t | \mathcal{B}_\tau) | \mathcal{B}_s) = H(t, \tau) \Phi(\tau, s) x_s \quad (3.25)$$

and the usual relationship $P_s y_t = H(t, s) x_s$ (where we took the same basis as in (3.25)).

Finally, (3.22) follows by comparing

$$E(x_t | \mathcal{B}_s) = \int_a^s \Gamma(t, \sigma) dz_\sigma, \quad t \geq s \quad (3.26)$$

with

$$E(x_t | \mathcal{B}_s) = \Phi(t, s) x_s = \Phi(t, s) \int_a^s \Gamma(s, \sigma) dz_\sigma. \quad (3.27)$$

Q.E.D.

One consequence of the preceding relationships is the popular factorization of the "impulse-response" kernel as

$$W(t, \sigma) = H(t) \Phi(t, \sigma) G(\sigma) \quad (3.28)$$

where $G(\sigma)$ is a $(N \times M)$ matrix with the property that $\Gamma(s, \sigma) = \Phi(s, \sigma) G(\sigma)$ for all $s \geq \sigma$.

Up to now we have not been worrying too much about a precise characterization of the Markov process (x_t) . Now, to go any further in the derivation of state equations, we have to assure ourselves that the "Markov semigroup" $\Phi(t, s)$ is "strongly continuous." As we will recognize in a moment this is equivalent to the right continuity (in mean square) of (x_t) .

Thus we are compelled to find conditions ensuring the possibility of choosing the bases x_t in $H(\mathcal{P}_t)$ for each time t , in such a way so as to obtain a right continuous dependence on time.

Proposition 3.6: There are (minimal) factorizations (3.9) where $H(\cdot, s)$, $s \in T$, intended as a family of matrix functions on $t \geq s$, is right continuous $\forall s \in T$, i.e.,

$$\lim_{h \downarrow 0} H(t, s+h) = H(t, s), \quad \forall t > s \quad (3.29)$$

if and only if for each $t \in T$ there is a choice of basis x_t in $H(\mathcal{P}_t)$ which defines a right continuous Markov process. Moreover, if (3.29) holds, then the two-parameter semigroup $\Phi(t,$

s), defined by (3.18), is right continuous at the diagonal $t = s$ of $T \times T$ and hence nonsingular $\forall t$ in a right neighborhood of each point $s \in T$.

Proof: Let us fix $t > s$, then $E(y_t | \mathcal{B}_s)$ considered as a function of s on (a, t) is a right continuous stochastic process since \mathcal{B}_s is.

From the factorization

$$E(y_t | \mathcal{B}_{s+\epsilon}) = H(t, s + \epsilon) x_{s+\epsilon} \quad (3.30)$$

valid for all $t > s + \epsilon$ and from (3.29) passing to the limit (in mean square) as $\epsilon \downarrow 0$, we conclude that $x_{s+\epsilon} \rightarrow x_s$.

Conversely if (x_t) is right continuous, by a similar argument, we easily conclude that (3.29) holds.

In order to prove the second part of the theorem, we reason as follows. By right continuity of (x_t) , there has to exist some $h > 0$ such that $\dim H(\mathcal{P}_t)$ is constant and equal to $\dim H(\mathcal{P}_s)$ for all t in $[s, s + h)$. Let t vary over the interval $[s, s + h)$, where $\dim H(\mathcal{P}_t)$ is constant. It is trivial that if (x_t) is right continuous at s , then the covariance $P(t, s) = E(x_t x_s')$ is right continuous at $t = s$ and thus

$$P(t, s) = \Phi(t, s)P(s) \longrightarrow P(s) \text{ as } t \downarrow s. \quad (3.31)$$

Since $P(s)$ is nonsingular, (3.31) implies

$$\lim_{t \downarrow s} \Phi(t, s) = I \quad (3.32)$$

that is, $\Phi(t, s)$ is (right continuous and) nonsingular in some right neighborhood of s . By (3.19), we can show that $\Phi(t, s)$ is actually nonsingular $\forall t$ in $[s, s + h)$. Q.E.D.

Let us agree to call a realization of (y_t) , with respect to a right continuous family (\mathcal{B}_t) , right continuous, if $H(t, s)$ can be chosen (in its equivalence class defined modulo right multiplication by a nonsingular matrix $T(s)$) at each time s , in such a way as to satisfy (3.29), and (consequently) x_s can be selected for each $s \in T$ in such a way as to form a right continuous Markov process. We have the following rather obvious proposition.

Proposition 3.7: Right continuous minimal realizations are equivalent under right continuous changes of bases in $H(\mathcal{P}_t)$, i.e., if $\hat{H}(t, s)$, \hat{x}_s and $H(t, s)$, x_s both define minimal realizations, then

$$\hat{H}(t, s)\hat{x}_s = H(t, s)x_s, \quad t \geq s > a \quad (3.33)$$

$\forall t$ and s if and only if

$$\begin{aligned} \hat{H}(t, s) &= H(t, s)T(s) \\ \hat{x}_s &= T^{-1}(s)x_s \end{aligned} \quad (3.34)$$

with $T(\cdot)$ a right continuous (nonsingular) matrix function.

Hereafter, we shall assume that the cross covariance $\Lambda(t, \tau)$ has a minimal (i.e., of the smallest possible dimension) factorization of the form (3.9) which

- i) has a finite dimension for all $s \in T$,
- ii) is right continuous (i.e., satisfies (3.29)) for all $s \in T$.

Then we may show that T can be partitioned in at most a countable number of subintervals $[t_k, t_{k+1})$, $k = 0, 1, 2, \dots$ with $a = t_0 < t_1 < t_2 < \dots$ where the dimension of the minimal state space, $\dim H(\mathcal{P}_t)$, is constant and has a finite nonzero jump through each point t_k .

Over each subinterval $[t_k, t_{k+1})$, $\Phi(t, s)$ is square, and nonsingular for all values of t and s , ($t \geq s$). Of course Φ depends

on the choice of basis on $H(\mathcal{P}_t)$ and $H(\mathcal{P}_s)$. Let us examine what kind of dependence we have.

Let $\hat{x}_t = T(t)x_t$ and $\hat{x}_s = T(s)x_s$ be different bases on $H(\mathcal{P}_t)$ and $H(\mathcal{P}_s)$. By the Markov property

$$E(\hat{x}_t | \mathcal{P}_s) = \hat{\Phi}(t, s)\hat{x}_s \quad (3.35)$$

as in (3.18), and, by some simple algebra, we see that the transformation rule for $\Phi(t, s)$ is expressed by

$$\hat{\Phi}(t, s) = T^{-1}(t)\Phi(t, s)T(s) \quad (3.36)$$

where, of course, $t_{k+1} > t \geq s \geq t_k$.

Note that, if we take

$$T(t) = \Phi(t, t_k), \quad T(s) = \Phi(s, t_k) \quad (3.37)$$

where t and s are as before, we define a right continuous change of basis by means of which $\Phi(t, s)$ transforms into the identity matrix. Thus we have the following.

Proposition 3.8: By a right continuous change of basis we can obtain

$$\hat{\Phi}(t, s) = I \quad (3.38)$$

over all subintervals of T of constant dimension. Correspondingly, we may express y_t as a function of the Markov process (x_t) as

$$x_t = \int_a^t G_k(\sigma) dz_\sigma \quad (3.39)$$

$$y_t = H_k(t)x_t \quad (3.40)$$

where $G_k(\cdot)$ and $H_k(\cdot)$ are suitable matrices.

Proof: The proof relies on the special form taken by (3.21) and (3.22) for a $\Phi(\tau, s)$ reducing to the identity.

By taking $t_{k+1} > \tau \geq s \geq t_k$ we observe that

$$H(t, s) = H(t, \tau), \quad \forall \tau, s \text{ in } [t_k, t_{k+1}) \quad (3.41)$$

$$\Gamma(\tau, \sigma) = \Gamma(s, \sigma), \quad \forall \tau, s \text{ in } [t_k, t_{k+1}) \quad (3.42)$$

for all t and σ .

These formulas tell us that $H(t, s)$ does not depend on s and likewise $\Gamma(s, \sigma)$. $H(\cdot, s)$ and $\Gamma(s, \cdot)$ propagate with the variable s by simple restriction or extension of the respective domains. If we let

$$G_k(\sigma) = \Gamma(s, \sigma), \quad t_{k+1} > s \geq t_k \quad (3.43)$$

and

$$H_k(t) = H(t, s), \quad t_{k+1} > s \geq t_k \quad (3.44)$$

then (3.43) defines $G_k(\sigma)$ on (a, t_{k+1}) with the same analytical properties possessed by $\Gamma(s, \sigma)$. The function $H_k(\cdot)$ is also defined by (3.44) over $[t_k, +\infty)$, in particular, if $t \in [t_k, t_{k+1})$ we have $H_k(t) = H(t, t)$. Hence (3.39) and (3.40) follow.

Q.E.D.

Remarks: If we assume $N = \dim H(\mathcal{P}_t)$ constant everywhere over T , then the above representation holds for all times, and we can compute the cross covariance in (3.12) without worrying about the proper choice of time domains. It is immediate that we can get a "uniform" factorization for $\Lambda(t, \tau)$ of the form (3.17). In particular, the restriction of $A(\cdot)$ to $[s, \infty)$ can be identified with $H(\cdot, s)$ which now has constant rank as s varies for the obvious reason that it is independent of

s. Thus Corollary 3.4 actually gives a necessary and sufficient condition for the state space to be time invariant.

Notice that (3.39) can be given the differential form

$$dx_t = G_k(t)dz_t, \quad t \in [t_k, t_{k+1}) \quad (3.45)$$

which has to be associated with an "initial condition" x_{t_k} resulting from the evolution over the $(k-1)$ th interval. Here the temporal evolution of the state (x_t) is that of an orthogonal increments process obtained from the innovation (z_t) by a very simple kind of transformation.

This very particular kind of Markov process can be transformed through changes of basis on $H(\mathcal{P}_t)$.

What we have to keep in mind, however, is that if we desire a differential equation for x_t we have to take "smooth" transformations $T(t)$.

Theorem 3.9: All state-space descriptions of (y_t) which are of the type

$$dx_t = F(t)x_t dt + K(t)dz_t \quad (3.46)$$

$$y_t = J(t)x_t, \quad t \in [t_k, t_{k+1}) \quad (3.47)$$

with $F(t)$ square and measurable and $K(t), J(t)$ matrices of the same dimension as $G_k(t)$ and $H_k(t)$ in (3.39), (3.40), can be obtained from this representation by means of an absolutely continuous change of basis $T(t)$ on $H(\mathcal{P}_t)$, $t \in [t_k, t_{k+1})$, and moreover,

$$\begin{aligned} F(t) &= \dot{T}(t)T^{-1}(t) \\ K(t) &= T(t)G_k(t) \\ J(t) &= H_k(t)T^{-1}(t). \end{aligned} \quad (3.48)$$

Proof (sketch): Indeed, if we perform an absolutely continuous transformation of basis of the form $\hat{x}_t = T(t)x_t$, then we end up with a differential equation for \hat{x}_t of the type (3.46) plus a state-output equation like (3.47) where the new matrices are precisely given by (3.48).

Conversely, we may solve (3.46) on $[t_k, t_{k+1})$ and obtain an explicit expression for x_t involving the "fundamental matrix" solution of

$$\dot{\Phi}(t, s) = F(t)\Phi(t, s) \quad \Phi(s, s) = I. \quad (3.49)$$

Then we use $T(t)$ given by (3.37). Q.E.D.

IV. CONCLUSION

We feel that our approach to the stochastic realization problem is the natural one, since it enables us to develop our formalism in a quite general time-varying setting.

It should also be pointed out that the realization of a process with respect to its own past σ -field (\mathcal{Y}_t) (starting from an autocovariance kernel) is just a special case of the general problem considered here. This special realization leads to the so-called innovation representation of y_t [11].

Actually, the *raison d'être* for the work reported here is to develop the probabilistic framework for studying the problem of equivalence between realizations constructed with respect to different σ -fields. This should provide further insight into various "recursive estimation" problems.

ACKNOWLEDGMENT

The author wishes to thank S. K. Mitter for many enlightening discussions during the preparation of this paper.

REFERENCES

- [1] H. Akaike, "Markovian representation of stochastic processes by canonical variables," *SIAM J. Contr.*, vol. 13, pp. 162-173, 1975.
- [2] —, "Stochastic theory of minimal realization," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 667-674, Dec. 1974.
- [3] B. D. O. Anderson, "The inverse problem of stationary covariance generation," *J. Stat. Phys.*, vol. 1, pp. 133-147, 1969.
- [4] B. D. O. Anderson, J. B. Moore, and S. G. Loo, "Spectral factorization of time-varying covariance functions," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 550-557, Sept. 1969.
- [5] B. D. O. Anderson and P. J. Moylan, "Spectral factorization of finite dimensional nonstationary matrix covariance," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 680-692, Dec. 1974.
- [6] H. Cramer, "On the structure of purely nondeterministic stochastic processes," *Ark. Mat.*, vol. 4, pp. 2-3 and 249-266, 1961.
- [7] J. L. Doob, *Stochastic Processes*. New York: Wiley, 1953.
- [8] N. Dunford and J. T. Schwartz, *Linear Operators*, vol. I-II. New York: Interscience, 1963.
- [9] P. Faurre, "Realizations Markoviennes de processus stationnaires," IRIA Rep. no. 13, 1973.
- [10] T. Kailath and R. A. Geesey, "An innovations approach to least squares estimation—Part IV: Recursive estimation given lumped covariance function," *IEEE Trans. Automat. Contr.*, vol. AC-16, pp. 720-727, Dec. 1971.
- [11] —, "An innovations approach to least-squares estimation—Part V: Innovations representations and recursive estimation in colored noise," *IEEE Trans. Automat. Contr.*, vol. AC-18, pp. 435-453, Oct. 1973.
- [12] P. R. Halmos, *Introduction to Hilbert Spaces*. New York: Chelsea, 1957.
- [13] T. Hida, "Canonical representations of Gaussian processes and their applications," *Mem. Coll. Sci. Kyoto Univ.*, vol. A-XXXIII, pp. 109-155, 1960.
- [14] R. E. Kalman, "Linear stochastic filtering—Reappraisal and outlook," in *Proc. Symp. System Theory Polytech. Inst. Brooklyn*, pp. 197-205, 1965.
- [15] G. Kallianpur and V. Mandrekar, "Multiplicity and representation theory for purely nondeterministic stochastic processes," *Theory Prob. Appl. (USSR)*, vol. X, pp. 553-580, 1965.
- [16] M. Loeve, *Probability Theory*. Princeton, N.J.: Van Nostrand, 1963.
- [17] H. P. McKean, "Brownian motion with a several dimensional time," *Theory Prob. Appl. (USSR)*, vol. VIII, pp. 335-354, 1963.
- [18] A. Nerode, "Linear Automata Transformations," in *Proc. Amer. Math. Soc.*, vol. 9, pp. 541-544, 1958.
- [19] J. Neveu, *Processus Aleatoires Gaussiennes*. Montreal, Canada: La Presse de l'Universite de Montreal, 1968.
- [20] G. Picci, "Theory of sufficient statistics and structure analysis of transition probabilities with applications to identifiability," LADSEB Rep. 74-06, 1974.
- [21] Y. A. Rozanov, "Innovations and nonanticipative processes," in *Multivariate Analysis*, vol. 3, P. R. Krishnaiah, Ed. New York: Academic Press.
- [22] R. A. Wijsman, "On the theory of B.A.N. estimates," *Ann. Math. Stat.*, vol. 30, pp. 185-191, 1959.
- [23] L. Weiss and R. E. Kalman, "Contributions to linear system theory," *Int. J. Eng. Sci.*, vol. 3, pp. 141-171, 1965.