Stochastic Differential Equations in Statistical Estimation Problems.*

by

G. Kallianpur

and

C. Striebel

University of Minnesota

TR# 109

To be presented at the Second International Symposium on Multivariate Analysis at Dayton, Ohio.

*Prepared with the support of a grant from the United States Air Force Office of Aerospace Research, Grant No. AF-AFOSR-885-65.

1. Introduction

The estimation problem considered in this paper can be described as follows: $\underline{x}(t, \eta)$, $0 \leq t \leq T$ called the "system process" is a stochastic process on a known probability space $(\Omega_X, \mathcal{B}_X, P_X)$, $(\eta \in \Omega_X)$ which takes values in R^n . It is assumed that direct observation of the system process is not possible or convenient but data concerning $\underline{x}(t)$ is provided by observations on an m-dimensional process $\underline{z}(t)$ given by

$$(1.1) \qquad \underline{z}(\tau) = \int_0^\tau \underline{h}(u, \underline{x}(u))du + \underline{y}(\tau) \ , \quad 0 \leq \tau \leq T$$

where the "noise" process $\underline{y}(t)$ is Gaussian, has independent increments and is independent of $\underline{x}(t)$. The available data, represented by $\mathcal{F}^t = \mathcal{B}[z(\tau), 0 \leq \tau \leq t]$, the σ-field generated by the family $z(\tau)$ $(0 \leq \tau \leq t)$, is to be used in estimating some given functional

$$(1.2) \qquad g(\eta) \equiv G[\underline{x}(\tau, \eta), \quad 0 \leq \tau \leq T]$$

of the system process. The precise conditions to be satisfied by $\underline{h}$ will be stated later. The space Ω_X on which the system process is defined corresponds to the parameter space in the usual Bayes approach to the theory of estimation. Thus the process $\underline{x}(t)$ may be regarded as the unknown parameter and P_X the à-priori distribution. If, as we shall always assume, g is an integrable random variable its least squares estimate based on $\mathcal{F}^t$ is the conditional expectation $E[g|\mathcal{F}^t]$ (which for brevity, we write as $E^t(g)$) . By suitably choosing g it can be seen that this problem includes smoothing, prediction and filtering. A "Bayes" formula for $E^t(g)$ is obtained in Theorem 2.1. This is a central result from a theoretical point of view since the results of Sections 5 and 6 follow from it. However, while such a formula might be considered satisfactory for fixed t , if the data is coming in continuously we require an estimate which can be continuously revised to take into account the new data. A practical method of describing the estimate or filter which depends continuously on time is furnished by a stochastic differential equation. To achive this we have to specialize to Markov system processes.

Both these problems (discussed in Sections 2 and 5) have been investigated by us in two earlier papers. In [6] is derived the formula for $E^t(g)$ when $\underline{x}(t)$ and $\underline{z}(t)$ are one dimensional, $\underline{h}(t, x) \equiv x$ and $\underline{y}(t)$ is a one dimensional standard Wiener process. Here in Section 2 we give the extension to vector valued processes and with nonlinear h as in (1-1). In Section 5 we consider the model (1.1) with $\underline{x}(t)$ Markov but assume both system and observation processes to be one dimensonal. The main theorems of this section give conditions under which the conditional expectation $E^t[f(x(t))]$ satisfies a stochastic differential equation of Ito type whenever f belongs to the domain of the extended infinitesimal generator of $x(t)$. The proofs of these results being lengthy and rather formidable have been omitted since they would greatly overburden the paper. For the special case $h(t, x) = x$ these details are to be found in our second paper [7] which has been submitted for publication. However, in order to give the reader some idea of how we proceed we have collected in Sections 3 and 4 the concepts and auxiliary results which are found essential in establishing these results.

As our only application of the theorems of Section 5 we give a rigorous derivation of the important results of R.E. Kalman and R.S. Bucy on linear filtering [8]. Although the Kalman-Bucy theory is by now familiar to engineers working in problems of stochastic estimation and control and has even found its way into textbooks in this field, we do not know of a published proof that is completely satisfactory. Another reason for including it is the hope that it will attract the interest of probabilists and statisticians.

Of the literature on the problems considered in this paper we mention only the few that have a direct bearing: the paper of W.M. Wonham [14] which treats special cases; the short note by R.S. Bucy [1], and the recent paper by

H. J. Kushner [9]. The last mentioned paper derives the stochastic differential equations for $E^t[f(\underline{x}(t)]$ when $\underline{x}(t)$ is the solution of a diffusion equation under conditions which seem more stringent than ours. To our knowledge the work that comes closest to ours is a recent paper, in two parts, by R.Sh. Liptzer and A.N. Shiryaev [10]. Without studying it carefully it is difficult to compare their results with ours. They consider a model in which the system process is the (unobservable) component of a two dimensional diffusion process. Their paper also contains a discussion of the Kalman-Bucy results and discusses several examples.

2. A Bayes formula for the conditional expectation. $E[g|\underline{z}(\tau) \quad 0 \leq \tau \leq t]$.

In [6] we have obtained a formula for the conditional expectation $E[g|z(\tau), \quad 0 \leq \tau \leq t]$ when both the system and the observation process are one dimensional. In this section we present the vector-valued version of that resul Only a sketch of the proof is given and the auxiliary results that point the way to the proof are stated. We have elaborated only on those aspects of the argument where the vector-valued situation presents features absent from the scalar case while details, mostly of a tedious nature, have been omitted. The reader interested in a meticulous construction of the proof of Theorem 2.1 will find these complementary details in our paper cited above.

We begin with a brief explanation of notation. $\underline{a}'$ denotes the transpose of the row vector $\underline{a}$ in R^m. The co-ordinates of $\underline{a}$ with respect to some fixed basis will be denoted by $a_j (j=1,\ldots,m)$ and the inner product in R^m by $(\ ,\ )$. We shall write $\|\underline{a}\|^2 = \sum_{j=1}^{m} a_j^2$.

Let $\underline{\xi}(t)$ be a process of independent increments with $E\underline{\xi}(t) = 0$ and variance operator

$$E[\underline{\xi}(t)]'[\underline{\xi}(t)] = \underline{A}(t) \qquad (0 \leq t \leq T), \text{ where} \tag{2.1}$$

$$\underline{A}(t) = [A_{ij}(t)] \quad (i,\ j = 1, \ldots, m).$$

We shall adopt the following conveniant notation due to Skorokhod [12]. If $\underline{a}(t)$ is a measurable function such that

$$\int_0^T \sum_{i,j} a_i(t)\ a_j(t)\ dA_{ij}(t) < \infty, \tag{2.2}$$

the integral

$$\int_0^T (\underline{a}(t),\ d\underline{\xi}(t)) = \sum_i \int_0^T a_i(t)\ d\xi_i(t) \tag{2.3}$$

is defined as in the scalar case. We then have

(2.4) $$E \int_0^T (\underline{a}(t), \; d\underline{\xi}(t)) = 0$$

and

(2.5) $$E \left[\int_0^T (\underline{a}(t) \; , \; d\underline{\xi}(t))\right]^2 = \int_0^T \sum_{i,j} a_i(t) \; a_j(t) \; dA_{ij}(t) \quad .$$

The notation for the integral given on the left hand side of (2.3) is Skorokhod's [12] and enables us to write the formulae in a more compact form.

Let us now describe precisely the model for the observation process introduced in the last section. First, we make the following assumptions:

(2.6) $\underline{y}(t)$, $t \in [0, T]$ $(\underline{y}(0) = 0)$ is a separable, m-vector valued Gaussian process of independent increments such that

(2.7) $$E \; \underline{y}(t) \equiv 0 \quad , \quad \text{and}$$

(2.8) $$E \; [\underline{y}(t)]' [\underline{y}(t)] = \underline{F}(t) \quad ,$$

where $\underline{F}(t)$ is a continuous function of t and $\underline{F}(t) \neq 0$ for each $t > 0$.

(2.9) $\underline{x}(t, \eta)$ is an n-vector valued, jointly measurable stochastic process defined on a probability space $(\Omega_X, \mathcal{B}_X, P_X)$ $(0 \leq t \leq T \; , \; \eta \in \Omega_X)$.

(2.10) The $\underline{x}(t)$ and $\underline{y}(t)$ processes are independent.

From (2.8) it follows that $\mu(t) = \text{Trace of } \underline{F}(t)$ is also continuous in t . Since $\underline{F}(t)$ is a positive semi-definite operator in R^m the following statements are verified easily.

(2.11) $\mu(t)$ is a non decreasing function of t ,

and

(2.12) $$\mu(t) > 0 \quad \text{for} \quad t > 0 \; .$$

(2.13) $F_{ij}(t)$ is absolutely continuous with respect to $\mu(t)$ $(i, j = 1, \ldots, m)$ and the derivative of the matrix function

(2.14) $\frac{dF}{d\mu}(t) = \hat{\underline{F}}(t)$

where

(2.15) $\hat{\underline{F}}(t)$ is a symmetric, positive semi definite operator in R^m .

With respect to some fixed basis in R^m let $\lambda_j(t)$ be the eigenvalues with corresponding orthonormal eigenvectors $\underline{e}_j(t)$ $(j = 1, \ldots, m)$ of $\hat{\underline{F}}(t)$. Clearly $\lambda_j(t) \geq 0$ and we may assume $\lambda_1(t) \geq \ldots \geq \lambda_m(t) \geq 0$. Since the λ_j's are the roots of a determinantal equatinn whose coefficients are measurable functions of t , it follows that the eigenvalues λ_j , and consequently, the eigenvectors $\underline{e}_j(t)$ are Borel measurable functions of t .

Let $\xi_j(t)$ $(j = 1, \ldots, m)$ be one dimensional, mutually independnet Gaussian processes with independent increments with

$$E\,\xi_j(t) \equiv 0 \quad \text{and} \quad E\,\xi_j^2(t) = \mu(t) .$$

Although the main result of this section can be proved without the following additional assumption on $\mu(t)$ we make it nevertheless because it is easier to verify the various steps in the proof by comparing with the proof given in detail in [6] for the one dimensional case.

(2.16) $\mu(t)$ is an absolutely continuous function of t with $\mu'(t) > 0$ for $(t > 0)$

Let

(2.17) $\underline{h}(t, \underline{x}): [0, T] \times R^n \to R^m$

be a Borel measurable function of $(t, \underline{x})$.

Abusing the notation somewhat we shall write $h(u, \eta)$ for $\underline{h}(u, \underline{x}(u, \eta))$. The conditions to be satisfied by the process $\underline{h}(u, \eta)$ are as follows.

There exists a jointly measurable m-vector valued process $\underline{p}(u, \eta)$ such that for $0 \leq u \leq T$

(2.18) $$\underline{h}(u, \eta) = \sum_{j=1}^{m} p_j(u, \eta) \sqrt{\lambda_j(u)} \; \underline{e}_j(u) \; \mu'(u) \quad \text{a.s.} P_X .$$

(2.19) $$\int_0^T \|\underline{p}(u, \eta)\|^2 \mu'(u) du < \infty \quad \text{a.s.} \; P_X .$$

A.S. P_X

(2.20) $$p_j(u, \eta) = 0 \quad \text{for all } u \text{ in } [0, T] \text{ for which } \lambda_j(u) = 0 .$$

When this happens we shall set $\frac{p_j(u, \eta)}{\sqrt{\lambda_j(u)}} = 0$.

The observation process $\underline{z}(t)$ takes values in R^m and is defined by

(2.21) $$\underline{z}(t) = \int_0^t \underline{h}(u, \underline{x}(u, \eta)) du + \underline{y}(t) \; , \; 0 \le t \le T$$

or

(2.21') $$d\underline{z}(t) = \underline{h}(t, \underline{x}(t, \eta)) dt + d\underline{y}(t) .$$

Observe that from (2.6) and (2.21), $\underline{z}(0) = 0$.

It is easy to see that (2.18) is equivalent to the relation

(2.22) $$(\underline{h}(u, \eta), \underline{e}_j(u)) = \sqrt{\lambda_j(u)} \; p_j(u, \eta) \; \mu'(u) \quad j = 1, \ldots, m .$$

Let us now set

(2.23) $$w_j(\tau) = \int_0^\tau \frac{1}{\sqrt{\mu'(u)}} \; d\xi_j(u)$$

and

(2.24) $$\hat{z}_j(\tau) = \int_0^\tau \left(\frac{\underline{e}_j(u)}{\sqrt{\lambda_j(u)\mu'(u)}} \; , \; d\underline{z}(u) \right) .$$

Then $w_j(\tau)$ $(0 \le \tau \le T)$ $(j = 1, \ldots, m)$ are independent standard Wiener processes and we have from (2.21) and

$$\underline{y}(t) = \sum_{j=1}^{m} \int_0^t \sqrt{\lambda_j(u)} \; \underline{e}_j(u) \; d\xi_j(u)$$

that

(2.25) $$d\hat{z}_j(u) = \sqrt{\mu'(u)} \; p_j(u, \eta) du + dw_j(u) \; , \quad 0 \le u \le T \; , \; j = 1, \ldots, m .$$

The arguments used in [6] can now be applied to vector valued system and observation processes related by the model (2.21) or (2.25). Theorem 2.1 and its corollaries are based on two results stated below. The first of these, given in [6], constitutes a general version of Bayes' theorem. The second is an extension due to Skorokhod [12] of a theorem of R. Cameron and R. Graves (see [6] for reference).

Lemma 2.1

(i) On the probability space $(\Omega, \mathcal{A}, P)$ let $g(\omega)$ be an integrable random variable measurable with respect to a sub σ-field $\mathcal{A}_X$ and let $Q(A, \omega)$ be a version of the conditional probability

(2.26) $$Q(A, \omega) = E(I_A | \mathcal{A}_X) \quad \text{a.s.}$$

for $A \in \mathcal{A}_Z$, a sub σ-field of $\mathcal{A}$. Then φ_g defined by

(2.27) $$\varphi_g(A) = \int g(\omega)\ Q(A, \omega)\ P(d\omega)$$

is a finite signed measure on $(\Omega, \mathcal{A}_Z)$. It is absolutely continuous with respect to P_Z $(\varphi_g \ll P_Z)$, the restriction of P to $\mathcal{A}_Z$ with Radon-Nikodym derivative given by

(2.28) $$E(g|\mathcal{A}_Z) = \frac{d\varphi_g}{dP_Z} \quad \text{a.s. } P_Z \ .$$

(ii) Suppose that the following conditions are fulfilled.

(2.29) The conditional probabilities $Q(A, \omega)$ are regular (see [11], p. 137).

(2.30) $\mathcal{A}_Z$ is generated by a countable family of sets.

(2.31) There exists a probability Λ on $(\Omega, \mathcal{A}_Z)$ such that $Q(\cdot, \omega) \ll \Lambda$ for $\omega \in \Omega'$ where $P(\Omega') = 1$.

Then

(2.32) $P_Z \ll \Lambda$,

(2.33) there exists a function $q(\omega', \omega)$ which is measurable on $(\Omega \times \Omega, \mathcal{a}_Z \times \mathcal{a}_X)$ and satisfies

(2.34) $$q(\omega', \omega) = \frac{dQ}{d\Lambda}(\cdot, \omega)(\omega') \quad \text{a.e } \Lambda \times P ,$$

(2.35) $$0 < \int q(\omega', \omega)\ P(d\omega) < \infty \quad \text{a.s. } P_Z$$

and

(2.36) $$E(g|\mathcal{a}_z) = \frac{\int g(\omega)\ q(\omega', \omega)\ P(d\omega)}{\int q(\omega', \omega)\ P(d\omega)} \quad \text{a.s. } P_Z$$

for g $\mathcal{a}_X$-measurable and integrable.

Lemma 2.2 [12]. Let $\underline{y}(t)$ $(0 \leq t \leq T)$ be as in (2.6) - (2.8) and $\underline{y}^{(1)}(t)$ be such that

(2.37) $$\underline{y}^{(1)}(t) = \underline{y}(t) + \underline{a}(t)$$

where the non random mean function $\underline{a}$ satisfies

(2.38) $$\underline{a}(t) = \int_0^t \sum_{j=1}^m p_j(u) \sqrt{\lambda_j(u)}\ \underline{e}_j(u)\ \mu'(u) du ,$$

(2.39) $$\int_0^T \|\underline{p}(u)\|^2\ \mu'(u) du < \infty$$

where $p_j(u)$ are Borel measurable functions and $\mu(u)$, $\lambda_j(u)$ $\underline{e}_j(u)$ and $\underline{y}(t)$ have been defined earlier.

If $P_{\underline{y}}$ and $P_{\underline{y}^{(1)}}$ denote the respective probability measures then,

$$P_{\underline{y}^{(1)}} \ll P_{\underline{y}}$$

with Radon-Nikodym derivative given by

(2.40) $$\exp\{ \sum_{j=1}^{m} \int_0^t (\frac{p_j(u)}{\sqrt{\lambda_j(u)}} \underline{e}_j(u) , d\underline{y}(u)) - \frac{1}{2} \sum_{j=1}^{m} \int_0^t p_j^2(u) \mu'(u) du\} .$$

In applying the Lemmas 2.1 and 2.2 to deduce Theorem 2.1 the following probability structure is assumed.

Let $(R^m)^{[0,\, t]}$ be the space of all m-vector valued functions $\underline{z}(\tau)$ $(0 \le \tau \le t)$, $\mathcal{B}_m^{[0,\, t]}$ the product σ-field in $(R^m)^{[0,\, t]}$, defined in the usual manner, and let $C_m[0, T]$ be the space of all continuous functions on $[0, T]$ taking values in R^m . Define

(2.41) $$W_m = C_m[0, T] , \mathcal{B}_{W_m} = W_m \cap \mathcal{B}_m^{[0,\, t]} , Z_m = C_m[0, t] ,$$

$\mathcal{B}_{Z_m} = Z_m \cap \mathcal{B}_m^{[0,\, t]}$ $(0 < t \le T)$. Let P_{W_m} be a standard (m-dimensional) Wiener measure on $(W_m, \mathcal{B}_{W_m})$ amd $(\Omega_X, \mathcal{B}_X, P_X)$ is the probability space in (2.9). Elements of Ω_X and W_m will be denoted by η and $\underline{w}$ respectively. We then take

(2.42) $$(\Omega, \mathcal{A}, P) = (\Omega_X \times W_m , \mathcal{B}_X \times \mathcal{B}_{W_m} , P_X \times P_{W_m}) .$$

Consider the transformations

(2.43) $$\Phi: (\Omega, \mathcal{A}) \to (\Omega_X, \mathcal{B}_X)$$

defined by

(2.44) $$\Phi(\eta, \underline{w}) = \eta$$

and

(2.45) $$\underline{\theta}: (\Omega, \mathcal{A}) \to (Z_m, \mathcal{B}_{Z_m})$$

defined by

(2.46) $$\theta_j(\eta, \underline{w})(\tau) = \theta_j^*(\eta) \quad (\tau) + w_j(\tau) \quad , 0 \le \tau \le t$$ where $\underline{\theta}^*$ is give η by

$$(2.47) \qquad \theta^*_j(\eta)(\tau) = \int_0^\tau \sqrt{\mu'(u)}\, p_j(u, \eta)\,du \quad \text{for} \quad 0 \le \tau \le t \quad \text{and} \quad j = 1, \ldots, m$$

if $\sum_{j=1}^{m} \int_0^t p_j^2(u, \eta)\mu'(u)\,du < \infty$, and $\theta^*_j(\eta)(\tau) = 0$ for $0 \le \tau \le t$ and $j = 1, \ldots, m$

if $\sum_{j=1}^{m} \int_0^t p_j^2(u, \eta)\,\mu'(u)\,du = \infty$.

The σ-fields $\mathcal{A}_X$ and $\mathcal{A}_Z$ are then defined by

$$(2.48) \qquad \mathcal{A}_X = \Phi^{-1}(\mathcal{B}_X), \quad \mathcal{A}_Z = \underline{\theta}^{-1}(\mathcal{B}_{Z_m}).$$

Lastly we identify

$$(2.49) \qquad Q(A, \omega) = P_{\underline{\theta}^*[\Phi(\omega)]}(\underline{\theta}A)$$

and

$$(2.50) \qquad \Lambda(A) = P_{W_m}\, \psi^{-1}(\underline{\theta}A) \qquad (A \in \mathcal{A}_Z)$$

where ψ defined on W_m by

$$(2.51) \qquad [\psi(\underline{w})](\tau) = \underline{w}(\tau) \qquad 0 \le \tau \le t.$$

The methods of [6] now yield

<u>Theorem</u> 2.1 Let the assumptions (2·6) - (2·10) and (2·16) - (2·20) hold. Then for any integrable random variable $g(\eta)$ defined on $(\Omega_X, \mathcal{B}_X, P_X)$ we have the formula

$E[g|z(\tau), 0 \le \tau \le t]$

$$(2.52) \qquad = \frac{\int g(\eta)\left\{\exp\left[\int_0^t \sum_{j=1}^{m} \left(\frac{\hat{h}_j(u, \eta)\, \underline{e}_j(u)}{\mu'(\mu)\, \lambda_j(u)}, d\underline{z}(u)\right) - \frac{1}{2}\int_0^t \sum_{j=1}^{m} \frac{\hat{h}_j^2(u, \eta)}{\mu'(u)\, \lambda_j(u)}\,du\right]\right\} P_X(d\eta)}{\int \exp\left[\int_0^t \sum_{j=1}^{m} \left(\frac{\hat{h}_j(u, \eta)\, \underline{e}_j(u)}{\mu'(u)\, \lambda_j(u)}, d\underline{z}(u)\right) - \frac{1}{2}\int_0^t \sum_{j=1}^{m} \frac{\hat{h}_j^2(u, \eta)}{\mu'(u)\, \lambda_j(u)}\,du\right] P_X(d\eta)}$$

where the denominator is positive a.s., and $\hat{h}_j(u, \eta) = (\underline{h}(u, \eta), \underline{e}_j(u))$. It is also understood that the summation on the right side of (2.52) is taken over those indices j for which $\lambda_j(u) > 0$.

Corollary to Theorem 2.1 Suppose there is an integer r , $1 \leq r \leq m$ such that for $j = 1, \ldots, r$

(2.53) $$\lambda_j(u) > 0 \text{ for } 0 \leq u \leq T \text{ and for } j = r+1, \ldots, m ,$$

$$\lambda_j(u) = 0 \text{ for } 0 \leq u \leq T .$$

Then formula (2.52) holds with the sum on the right hand side expression extending from $j = 1$ to r .

It is possible to cast the formula in a more compact form in the case of full rank i.e., when $r = m$.

Theorem 2.2 Let the assumptions of Theorem 2.1 hold. Further suppose that the matrix function $\hat{F}(t)$ is invertible for $0 \leq t \leq T$. Then

(2.53) $$E[g | z(\tau), 0 \leq \tau \leq t] = \frac{\int g(\eta) \{ \exp [\int_0^t ([\mu'(u) \hat{F}(u)]^{-1} \underline{h}(u, \eta) , dz(u)) - \frac{1}{2} \int_0^t ([\mu'(u) \hat{F}(u)]^{-1} \underline{h}(u, \eta), \underline{h}(u, \eta)) du] \} P_X(d\eta)}{\int \exp [\int_0^t ([\mu'(u) \hat{F}(u)]^{-1} \underline{h}(u, \eta), d\underline{z}(u)) - \frac{1}{2} \int_0^t ([\mu'(u) F(u)]^{-1} \underline{h}(u, \eta), \underline{h}(u, \eta)) du] P_X(d\eta)} .$$

If the "noise" in the model (2.21) is an m-dimensional, standard Wiener process, i.e. if $\underline{v}(t) = \underline{w}(t) = (w_1(t), \ldots, w_m(t))$ where the w_j's are independent, standard one dimensional Wiener processes, then the right hand side of (2.53) take on an even simpler form, since $[\mu'(u) \hat{F}(u)]^{-1} = I$, the identity matrix.

3. Stochastic Integrals and Differentials of Ito Type.

Since the stochastic integral of K. Ito [5] is a basic tool used in the solution of the problems considered in this paper, it seems desirable to list some of its important properties. We state a very important result on stochastic differentials also due to Ito which might be looked upon as a "change of variables" formula for an Ito integral. The reader is referred to the book by I. I. Gikhman and A. V. Skorokhod for proofs and details [4].

We shall be concerned with stochastic integrals defined with respect to the system

(3.1) $$(\Omega, \mathcal{A}, P), \quad \mathcal{F}_t, \quad w(t,\omega), \quad 0 \le t \le T, \quad \omega \in \Omega$$

where $(\Omega, \mathcal{A}, P)$ is an arbitrary probability space. It will be assumed that the $\mathcal{F}_t$ are complete with respect to P, that

(3.2) $$\mathcal{F}_{t_1} \subset \mathcal{F}_{t_2} \subset \mathcal{A} \qquad \text{for } 0 \le t_1 \le t_2 \le T,$$

and that $w(t,\omega)$ is a Wiener process such that

(3.3) $$w(t,\cdot) \text{ is } \mathcal{F}_t\text{-measurable for } 0 \le t \le T,$$

and for $0 \le t \le T$

(3.4) $$\mathcal{F}_t \text{ is independent of } w(v) - w(t) \text{ for } t \le v \le T.$$

The notation $\overline{\mathcal{A}}$ indicates the completion of the σ-field with respect to P. We shall consider two classes of processes defined on $(\Omega, \mathcal{A}, P)$. The class $\mathcal{M}_1$ consists of those processes $a(t,\omega)$, $0 \le t \le T$, $\omega \in \Omega$ which satisfy the following conditions:

(3.5) $$a(t,\omega) \text{ is measurable on } \left([0,T] \times \Omega,\ \overline{\mathcal{B}_{[0,T]} \times \mathcal{A}},\ \mu \times P\right)$$

where $\mathcal{B}_{[0,T]}$ is the family of Borel subsets of the interval $[0,T]$ and μ is Lebesgue measure on $[0,T]$, (3.6) $a(t,\cdot)$ is $\mathcal{F}_t$-measurable a.e. μ,

and

(3.7) $$\int_0^T |a(t,\omega)|dt < \infty \qquad \text{a.s. } P.$$

The class $\mathcal{M}_2$ is defined to consist of those processes $b(t,\omega)$ $(0 \le t \le T, \ \omega \in \Omega)$ which satisfy (3.5), (3.6) and

(3.8) $$\int_0^T [b(t,\omega)]^2 dt < \infty \qquad \text{a.s. } P.$$

The stochastic integral $\int_0^T b(t,\omega)dw(t,\omega)$ is defined for $b \in \mathcal{M}_2$ as follows:

(i) If $b \in \mathcal{M}_2$ is a step function

(3.9) $$b(t,\omega) = b_i(\omega) \qquad \text{for } t_i \le t < t_{i+1}$$

where

(3.10) $$0 = t_0 < t_1 < \cdots < t_m = T,$$

then define

(3.11) $$\int_0^T b(t,\omega)dw(t,\omega) = \sum_{i=0}^{m-1} b_i(\omega)\{w(t_{i+1},\omega) - w(t_i,\omega)\} \qquad \text{a.s. } P.$$

(ii) If $b \in \mathcal{M}_2$, there exist step functions $b_n \in \mathcal{M}_2$, $n = 1, 2, \ldots$ such that

(3.12) $$\int_0^T [b_n(t) - b(t)]^2 dt \to 0 \quad \text{in probability}$$

and the integrals

(3.13) $$\int_0^T b_n(t)dw(t)$$

converge in probability to a random variable. We then define

(3.14) $$\int_0^T b(t)dw(t) = \operatorname*{prob\,lim}_{n\to\infty} \int_0^T b_n(t)dw(t).$$

The following lemma states some of the familiar properties of the Ito integral.

Lemma 3.1:

(i) If $b_n \in \mathcal{M}_2$ $n = 0, 1, 2, \ldots$

and

(3.15) $$\int_0^T [b_n(t) - b_0(t)]^2 dt \to 0 \quad \text{in probability}$$

then

(3.16) $$\int_0^T b_n(t)dw(t) \to \int_0^T b(t)dw(t) \quad \text{in probability.}$$

(ii) If $b_1, b_2 \in \mathcal{M}_2$ and α_1, α_2 are real numbers then

(3.17) $$\int_0^T [\alpha_1 b_1(t) + \alpha_2 b_2(t)]\, dw(t) = \alpha_1 \int_0^T b_1(t)dw(t) + \alpha_2 \int_0^T b_2(t)dw(t) \quad \text{a.s. P.}$$

(iii) If $b \in \mathcal{M}_2$ and $\int_s^t E^{\mathcal{F}_s}[b^2(u)]du < \infty$ a.s. for $0 \leq s \leq u \leq t \leq T$, then

(3.18) $$E^{\mathcal{F}_s}\{\int_s^t b(u)dw(u)\} = 0$$

and

(3.19) $$E^{\mathcal{F}_s}\{[\int_s^t b(u)dw(u)]^2\} = \int_s^t E^{\mathcal{F}_s}[b^2(u)]du \quad \text{a.s. P.}$$

Integrals on the restricted range $[s,t]$ as in (3.19) are obtained by considering functions $b \in \mathcal{M}_2$ for which

(3.20) $$b(u,\omega) = 0 \quad \text{for } 0 \leq u < s \text{ and } t < u \leq T.$$

In addition to the properties mentioned above the proofs of the principal theorems of Section 5 require results of a rather specialized nature concerning Ito integrals in which the integrand process depends on two probability parameters. More precisely, the situation is as follows:

Let $(\Omega_X, \mathcal{B}_X, P_X)$ be a probability space and define

(3.21) $$(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P}) = (\Omega_X \times \Omega,\ \mathcal{B}_X \times \mathcal{A},\ P_X \times P),$$

(3.22) $$\tilde{\mathcal{F}}_t = \overline{\mathcal{B}_X \times \mathcal{F}_t} \quad (0 \leq t \leq T).$$

The product system

(3.23) $$(\tilde{\Omega},\tilde{\mathcal{A}},\tilde{P}),\quad \tilde{\mathcal{F}}_t,\quad w(t,\tilde{\omega}) \qquad 0 \le t \le T,\ \tilde{\omega} \in \tilde{\Omega}$$

satisfies the conditions (3.2), (3.3) and (3.4) where $w(t)$ on $\tilde{\Omega}$ is defined by

(3.24) $$w(t,\tilde{\omega}) = w(t,\eta,\omega) = w(t,\omega).$$

Let $\tilde{\mathcal{M}}_1$ and $\tilde{\mathcal{M}}_2$ be the classes of processes for the product system (3.21) defined by (3.5), (3.6), (3.7) and (3.5), (3.6) and (3.8), respectively. The following results (of which Lemma 3.3(ii) appears to be new) are Fubini-type theorems about the Ito stochastic integral. [See [7]).

<u>Lemma 3.2</u>: If $a \in \tilde{\mathcal{M}}_i$

(3.25) $$\int_0^T \left|\int_{\Omega_X} a(u,\eta,\omega)\, P_X(d\eta)\right|^i du < \infty \qquad \text{a.s. } P$$

and either

(3.26) $$a(u,\eta,\omega) \ge 0 \qquad \text{a.e. } \mu \times P$$

or

(3.27) $$\int_{\Omega_X} |a(u,\eta,\omega)|\, P_X(d\eta) < \infty \qquad \text{a.e. } \mu \times P,$$

then

(3.28) $$\int_{\Omega_X} a(u,\eta,\omega)\, P_X(d\eta) \in \mathcal{M}_i \qquad \text{for } i = 1, 2.$$

<u>Lemma 3.3</u>

(i) Lf $a \in \tilde{\mathcal{M}}_1$ and

(3.29) $$\int_0^T \int_{\Omega_X} |a(t,\eta,\omega)|\, P_X(d\eta)dt < \infty \qquad \text{a.s. } P,$$

then

(3.30) $$\int_{\Omega_X} a(t,\eta,\omega)\, P_X(d\eta) \in \mathcal{M}_1$$

(3.31) $$\int_{\Omega_X}\int_0^T a(t,\eta,\omega)dt\ P_X(d\eta) = \int_0^T\int_{\Omega_X} a(t,\eta,\omega)\ P_X(d\eta)dt \qquad \text{a.s. } P.$$

(ii) Let $b \in \widetilde{\mathcal{M}}_2$ satisfy

(3.32) $$\int_s^t [\int_{\Omega_X} |b(u,\eta,\omega)| P_X(d\eta)]^2 du < \infty \qquad \text{a.s. } P$$

and

(3.33) $$\int_{\Omega_X} \{\int_s^t E^{\mathcal{F}_s}[b^2(u,\eta,\omega)]du\}^{\frac{1}{2}} P_X(d\eta) < \infty \qquad \text{a.s. } P.$$

Then

(3.34) $$\int_{\Omega_X} b(u,\eta,\omega)\ P_X(d\eta) \in \mathcal{M}_2$$

and

(3.35) $$\int_s^t [\int_{\Omega_X} b(u,\eta,\omega)\ P_X(d\eta)]\ dw(u,\omega) = \int_{\Omega_X} [\int_s^t b(u,\eta,\omega)dw(u,\omega)]\ P_X(d\eta) \qquad \text{a.s. } P$$

and the integrals in (3.35) are finite a.s. P.

It is with the help of Lemma 3.3 that we are able to obtain the stochastic differentials for the process referred to in Section 2.

We now turn to Ito's definition of a stochastic differential ([5], p. 187). A process $\zeta(t)$ defined on $(\Omega,\mathcal{A},P)$ $[\text{or}(\tilde{\Omega},\tilde{\mathcal{A}},\tilde{P})]$ has a stochastic differential of Ito type

(3.36) $$d\zeta(t) = a(t)dt + b(t)dw(t) \qquad 0 \le t \le T$$

provided

(3.37) $$a \in \mathcal{M}_1 \quad (\text{or } \widetilde{\mathcal{M}}_1)$$

(3.38) $$b \in \mathcal{M}_2 \quad (\text{or } \widetilde{\mathcal{M}}_2),$$

and

(3.39) $$\zeta(t) - \zeta(s) = \int_s^t a(u)du + \int_s^t b(u)dw(u) \quad \text{a.s. } P \ (\text{or } \tilde{P})$$

for all $0 \le s < t \le T$.

The following result of Ito's ([3], p. 222) is used extensively in the proofs of the theorems of Section 5 and 6.

Lemma 3.4: Assume that the processes $\zeta_i(t)$ $(i = 1, 2, \ldots, n)$ have differentials

$$d\zeta_i(t) = a_i(t)dt + b_i(t)dw(t) \qquad 0 \le t \le T. \tag{3.40}$$

Let $\Gamma(x)$ be a real-valued function of the n-vector $x=(x_1,\ldots,x_n)$ defined on an open subset G of R^n which contains almost surely all points $(\zeta_1(t),\ldots,\zeta_n(t))$ $(0 \le t \le T)$. Further suppose that

$$\frac{\partial^2\Gamma(x)}{\partial x_i \partial x_j} \tag{3.41}$$

is continuous on G for $i, j = 1, 2, \ldots, n$. Let

$$\xi(t) = \Gamma(\zeta(t)) \qquad \text{where } \zeta(t) = [\zeta_1(t),\ldots,\zeta_n(t)]. \tag{3.42}$$

Then $\xi(t)$ has a differential

$$d\xi(t) = A(t)dt + B(t)dw(t) \qquad 0 \le t \le T \tag{3.43}$$

where

$$A(t) = \sum_i a_i(t) \frac{\partial\Gamma}{\partial x_i}(\zeta(t)) + \frac{1}{2}\sum_{i,j} \frac{\partial^2\Gamma(\zeta(t))}{\partial x_i \partial x_j} b_i(t)b_j(t) \tag{3.44}$$

and

$$B(t) = \sum_i \frac{\partial\Gamma(\zeta(t))}{\partial x_i} b_i(t). \tag{3.45}$$

4. Markov Processes With Extended Infinitesimal Generator.

In this section it will be assumed that $x(t,\eta)$, $0 \le t \le T$, $\eta \in \Omega_X$, is a Markov process. We will say that $x(t,\eta)$ has jointly measurable transition probabilities provided it has regular transition probabilities

$$P[x(t) \in B | x(s) = x] = P(s,x;t,B) \tag{4.1}$$

(x real and $0 \le s \le t \le T$) which are jointly measurable in (s,x,t) for all Borel sets of the real line $B \in \mathcal{B}_R$. We will be concerned with the generalized semi-group (in the sense of Loeve [11], p. 568) P_s^t $(0 \le s \le t \le T)$ defined by

$$(P_s^t f)(x) = \int_{-\infty}^{\infty} f(y) \, P(s,x;t,dy) \tag{4.2}$$

on the Banach space with sup norm of bounded measurable functions of a real variable; i.e., for $f \in B(R, \mathcal{B}_R)$. The generalized semi-group property which corresponds to the Chapman-Kolmogorov equation for Markov processes, is given by

$$P_s^t = P_s^u P_u^t \qquad (0 \le s \le u \le t \le T). \tag{4.3}$$

We suppose that $\{P_s^t\}$ has a generalized infinitesimal generator G_t $(0 \le t \le T)$ defined on a domain $\mathcal{D} \subset B(R, \mathcal{B}_R)$. Specifically, it is assumed that for $0 \le t \le T$, G_t is a linear operator with domain $\mathcal{D}$ and range $B(R, \mathcal{B}_R)$ which satisfies

$$\sup_{\substack{-\infty<x<\infty \\ 0\le t\le T}} |(G_t f)(x)| < \infty \tag{4.4}$$

and

$$\sup_{\substack{-\infty<x<\infty \\ 0\le t\le T}} \left| G_t f(x) - \frac{(P_t^{[t+h]} f - f)(x)}{h} \right| \to 0 \tag{4.5}$$

as $h \to 0$ for all $f \in \mathcal{D}$, where

(4.6) $$[t+h] = t+h \quad \text{if} \quad t+h \le T$$
$$= T \quad \text{if} \quad t+h > T.$$

It can be shown that if $x(t)$ is a measurable Markov process with jointly measurable transition probabilities (4.1), then for $f \in \mathcal{D}$

(4.7) $$(G_t f)(x)$$

is jointly measurable in (t,x) and

(4.8) $$(P_s^t f - f)(x) = \int_s^t (P_s^u G_u f)(x)du \qquad (0 \le s \le t \le T) \quad (-\infty < x < \infty).$$

For purposes of application we need to obtain stochastic differential equations for $E^t[f(x(t))]$ for certain unbounded functions f. The most natural class, say $\mathcal{D}^*$ of functions for which we can derive the basic stochastic differential equation is defined as follows. $\mathcal{D}^*$ is the class of Borel measurable functions f which satisfy

(4.9) $$E|f(x(t))| < \infty \quad \text{for each } t;$$

there exists a (t,x) Borel measurable function $(G_t^* f)(x)$ such that

(4.10) $$\int_0^T E|(G_t^* f)(x(t))|dt < \infty,$$

and for $0 \le s < t \le T$

(4.11) $$(P_s^t f)(x(s)) - f(x(s)) = \int_s^t (P_s^u G_u^* f)(x(u))du \qquad \text{a.s. } P_X.$$

We shall call G_t^* the extended infinitesimal generator of the process $x(t)$. It is easy to see that $\mathcal{D}^*$ contains $\mathcal{D}$. The definition of a generalized or extended infinitesimal generator of a vector-valued Markov process presents no difficulties.

5. Ito Stochastic Differential Equation For $E^t[f(x(t))]$.

In this section we make the following assumptions concerning the process $x(t,\eta)$, $0 \le t \le T$, $\eta \in \Omega_X$.

(5.1) $x(t,\eta)$ is a jointly measurable Markov process.

(5.2) $x(t,\eta)$ has regular transition probabilities given by (4.1) which are jointly measurable.

(5.3) $x(t,\eta)$ has an extended generator G_t^* with domain $\mathcal{D}^*$.

Further let $h(x,t)$ be a real valued, Borel measurable function of (x,t) satisfying the following conditions:

(5.4) $$\int_0^T E[h(x(t),t)]^4 dt < \infty;$$

there exists a $\Delta > 0$ such that

(5.5) $$E\left| e^{16\int_t^{t+\Delta} h^2(x(u),u)du} \right| < \infty$$

for all t for which $0 \le t < t+\Delta \le T$.

Theorem 5.1: Let $x(t,\eta)$ satisfy (5.1 - (5.5) and let $f \in \mathcal{D}^*$ satisfy

(5.6) $$\int_0^T E[f^4(x(t))]dt < \infty,$$

(5.7) $$\int_0^T E[G_t^* f(x(t))]^4 dt < \infty,$$

and

(5.8) $$\int_0^T E\{h(x(t),t)[f(x(T)) - \int_t^T G_u^* f(x(u))du]\}^4 du < \infty.$$

Then the process $E^t[f(x(t))]$ on $(\Omega,\mathcal{Q},P)$ has a stochastic differential

(5.9) $$dE^t[f(x(t))] = E^t[(G_t^* f)(x(t))]dt + \{E^t[h(x(t),t) \cdot f(x(t))] - E^t[h(x(t),t)] \cdot E^t[f(x(t))]\} \cdot \{dz(t) - E^t[h(x(t),t)]dt\}.$$

Condition (5.5) can be replaced by a stronger but more easily verifiable condition. Then, for functions f in $\mathcal{D}$ the above result assumes the following particularly simple form.

Theorem 5.2: Suppose $x(t,\eta)$ satisfies (5.1), (5.2) and the following condition:

(5.10) There exists a positive number c such that for all t in $[0,T]$

$$E[e^{ch^2(x(t),t)}] \leq M < \infty,$$

where M does not depend on t.

Then for $f \in \mathcal{D}$, $E^t[f(x(t))]$ satisfies the stochastic equation (5.9).

Proof: By a use of Jensen's convexity inequality ([10], p. 159) it is easy to see that (5.10) implies that condition (5.5) holds with Δ such that $0 < \Delta < c/16$. Moreover, it follows that for every positive integer n, $E|h(x(t),t)|^n$ is finite for all t in $[0,T]$ and that

(5.11) $$\int_0^T E|h(x(t),t)|^n dt < \infty.$$

In particular, (5.4) holds and since $f \in \mathcal{D}$, f and $G_t f$ are bounded functions. Hence conditions (5.6) - (5.8) are all satisfied and equation (5.9) follows.

In many important applications the process $x(t,\eta)$ is supposed to be a dynamical system. In probabilistic terms this means that the temporal development of the process is described by a stochastic differential equation. Accordingly we shall assume that $x(t)$ $(0 \leq t \leq T)$ is the solution of the Ito stochastic differential equation

$$(5.12)\qquad dx(t) = m[t,x(t)]dt + \sigma[t,x(t)]dB(t), \quad 0 \le t \le T,$$

or equivalently of the equation

$$(5.13)\qquad x(t) = x(s) + \int_s^t m[u,x(u)]du + \int_s^t \sigma[u,x(u)]dB(u) \quad (0 \le s < t \le T).$$

Here $B(t)$ is a standard Wiener process and $x(0) = x_0$, a given initial random variable. The coefficients m and σ are Baire functions of (u,ξ) which will be assumed to satisfy the following conditions

$$(5.14)\qquad |m(u,\xi)| \le K(1 + \xi^2)^{1/2} \qquad (-\infty < \xi < \infty, \quad 0 \le u \le T)$$

$$(5.15)\qquad |m(u,\xi_2) - m(u,\xi_1)| \le K|\xi_1 - \xi_2|$$

$$(5.16)\qquad 0 < \sigma(u,\xi) \le K$$

and

$$(5.17)\qquad |\sigma(u,\xi_2) - \sigma(u,\xi_1)| \le K|\xi_1 - \xi_2|.$$

Under these conditions it is well known ([2], Chapter VI) that $x(t)$ is a Markov process and is the unique solution of (5.12) almost all of whose sample functions are continuous.

The probability space on which equations (5.12) and (5.13) hold is, of course, that of the system process $(\Omega_X, \mathcal{B}_X, P_X)$.

<u>Lemma 5.1</u>: Let $x(t)$ be the solution of (5.12) where $m(t,x)$ and $\sigma(t,x)$ satisfy (5.14) - (5.17). Suppose further, that the initial value $x(0)$ satisfies

$$(5.18)\qquad E[e^{c_0 x^2(0)}] < \infty \quad \text{for some} \quad c_0 > 0.$$

Then the functions of the form

$$(5.19)\qquad f(x) = x^n p(x)$$

where n is any nonnegative integer and p, p' and p'' are bounded continuous functions, belong to $\mathcal{D}^*$. The extended infinitesimal generator G_t^* is given by

$$(5.20) \qquad (G_t^* f)(x) = m(t,x)\, f'(x) + 1/2\, \sigma^2(t,x)\, f''(x).$$

With the aid of the above lemma and Theorem 5.1 we can derive stochastic differential equations for the conditional moments of $x(t)$ and, if so desired, a stochastic differential equation for the conditional characteristic function of $x(t)$.

<u>Theorem 5.3</u>: Let $x(t)$ be the solution of the equation (5.12) and let the conditions (5.4), (5.5), (5.14) - (5.18) be satisfied. Then if p, p' and p'' are bounded and continuous and n is a nonnegative integer $E^t[x^n(t)\, p(x(t))]$ has a stochastic differential of Ito type

$$(5.21) \qquad dE^t[x^n(t)\, p(x(t))] = E^t[(G_t^* x^n p)(x(t))]dt$$
$$+ \{E^t[x^n(t) h(x(t),t)\, p(x(t))] - E^t[x^n(t)\, p(x(t))]\, E^t[h(x(t),t)]\}$$
$$\cdot \{dz(t) - E^t[h(x(t),t)]dt\} \qquad 0 \le t \le T$$

where G_t^* is given by (5.20).

6. The Linear Filter: The Kalman-Bucy Theory.

We shall consider now the important special case of the linear filter, i.e., when the observation process is given by

$$(6.1) \qquad dz(t) = k(t)x(t)dt + dw(t) \qquad (0 \le t \le T), \quad z(0) = 0,$$

where

(6.2) $k(t)$ is a continuous function of t.

Theorem 6.1: Suppose the system process $x(t)$ satisfies the stochastic differential equation (5.12) whose coefficients $m(t,x)$ and $\sigma(t,x)$ satisfy the conditions (5.14) - (5.17). Let the initial value $x(0)$ satisfy (5.18). If the observation process is defined by (6.1) then

$$(6.3) \qquad dE^t[x^n(t)p(x(t))] = E^t[(G_t^* x^n p)(x(t))]dt$$
$$+ k(t)\{E^t[x^{n+1}(t)p(x(t))] - E^t[x^n(t)p(x(t))]\cdot E^t[x(t)]\}$$
$$\cdot \{dz(t) - k(t)\, E^t[x(t)]dt\}.$$

Let us now specialize to the case when $x(t)$ is a Gaussian process which is the solution of the equation

$$(6.4) \qquad dx(t) = m(t)x(t)dt + \sigma(t)dB(t) \qquad (0 \le t \le T)$$

where

(6.5) $x(0) = x_0$, a Gaussian random variable,

and

(6.6) $m(t)$ and $\sigma(t)$ are Baire functions and $\sigma(t)$ is bounded and nonnegative.

For the process $x(t)$ satisfying (6.4) - (6.6) the following corollary is easily deduced from Theorem 6.1.

Corollary to Theorem 6.1: Suppose that the assumptions (6.1), (6.2), (6.4) - (6.6) hold for the processes $z(t)$ and $x(t)$. Then

$$dE^t[x(t)] = m(t)\, E^t[x(t)]dt + k(t)\{E^t[x^2(t)] - (E^t[x(t)])^2\}\{dz(t) - k(t)E^t[x(t)]dt\} \tag{6.7}$$

and, for $n \geq 2$

$$dE^t[x^n(t)] = \{nm(t)E^t[x^n(t)] + \frac{1}{2}\sigma^2(t)n(n-1)E^t[(x(t))^{n-2}]\}dt + k(t)\{E^t[x^{n+1}(t)] - E^t[x^n(t)]E^t[x(t)]\}\{dz(t) - k(t)E^t[x(t)]dt\} \tag{6.8}$$

The above corollary, together with Ito's lemma given in Section 3 enables us to obtain a rigorous derivation of the basic result first obtained by R. Kalman and R. S. Bucy [8].

For convenience let us first set

$$\hat{x}(t) = E^t[x(t)]$$

(6.9) and

$$R(t) = E^t[x^2(t)] - [\hat{x}(t)]^2.$$

Theorem 6.2: $\hat{x}(t)$ $(0 \leq t \leq T)$ is a Gaussian process satisfying the stochastic differential equation

$$d\hat{x}(t) = m(t)\,\hat{x}(t)\,dt + k(t)R(t)\{dz(t) - k(t)\hat{x}(t)dt\} \quad (0 \leq t \leq T),\ \hat{x}(0)=Ex_0 \tag{6.10}$$

where $R(t)$ satisfies the Riccati equation

$$\frac{dR(t)}{dt} = \sigma^2(t) + 2m(t)R(t) - k^2(t)R^2(t). \tag{6.11}$$

Before proving the result we need the following lemmas:

Lemma 6.1: Let t be fixed $(0 < t \leq T)$ and let $0 \leq t_1 < \cdots < t_n \leq t$. Then the conditional distribution of $x(t)$ given $z(t_1),\ldots,z(t_n)$ is Gaussian.

Proof: For real constants $c_0, c_1,\ldots,c_n$ define the random variable

$$\xi = c_0 x(t) + c_1 z(t_1) + \cdots + c_n z(t_n).$$

Setting $y_j = \int_0^{t_j} k(u)x(u)du$, $w_j = w(t_j)$, $\xi' = c_0 x(t) + \sum_{j=1}^{n} c_j y_j$ and $\xi'' = \sum_{j=1}^{n} c_j w_j$ we have

$$\xi = \xi' + \xi'',$$

where ξ' and ξ'' are clearly independent. The random variable ξ'' is Gaussian since w is a Gaussian process. From (6.4), a separable version of the $x(t)$ process (we consider only separable versions here) is a.s. sample continuous, i.e.; almost all of its sample functions are continuous. Hence by taking an appropriate sequence of subdivisions of the interval $[0,t]$ it is seen that ξ' is the almost sure limit of a sequence of random variables each of which is a finite linear combination of the random variables of the family $\{x(u), 0 \leq u \leq t\}$. Since x is Gaussian it follows that ξ' is Gaussian. Hence ξ is Gaussian. Since the constants $c_0,\ldots,c_n$ are arbitrary we have the joint normality of $\{x(t), z(t_1),\ldots,z(t_n)\}$. The conclusion of the lemma follows immediately.

Lemma 6.2: The conditional distribution of $x(t)$ given $\mathcal{F}_t$ $(= \mathcal{B}(z(\tau), 0 \leq \tau \leq t)$ is Gaussian with mean $\hat{x}(t)$ and variance $R(t)$ given by (6.9).

Proof: It follows from (6.1) that a separable version of $z(t)$ is almost surely sample continuous. Let $\{D_n\}$ $(n = 1, 2,\ldots)$ be a sequence of finite partitions of $[0,t]$ such that $D_n \subset D_{n+1}$ for each n and $\bigcup_{n=1}^{\infty} D_n$ is dense in $[0,t]$. Now let $\mathcal{F}_t^n = \mathcal{B}[z(\tau), \tau \in D_n]$. Then $\mathcal{F}_t^n \subset \mathcal{F}_t^{n+1}$. From the martingale convergence theorem and the a.s. sample continuity of the z process it follows that

$$\lim_{n \to \infty} P\{x(t) \in B | \mathcal{F}_t^n\} = P\{x(t) \in B | \mathcal{F}_t\} \quad \text{a.s.,} \tag{6.12}$$

for every Borel set B. By Lemma 6.1 the left hand side probabilities in (6.10) are Gaussian. Lemma 6.2 follows.

Lemma 6.3: The process $\hat{x}(t)$ $(0 \le t \le T)$ is Gaussian.

Proof: The joint normality of the random variables $\hat{x}(v_1),\ldots,\hat{x}(v_p)$ $(0 \le v_1 < \cdots < v_p \le T)$ is shown by proceeding essentially as in the preceding two lemmas.

Proof of Theorem 6.2: Equation (6.10) is nothing but equation (6.7) of the Corollary to Theorem 6.1. What remains to be established is that $R(t)$ is the nonrandom function of t which is the solution of (6.11). From (6.9) $R(t)$ is the difference of the two stochastic processes $E^t[x^2(t)]$ and $[\hat{x}(t)]^2$. Taking $n = 2$ in (6.8) we obtain

$$dE^t[x^2(t)] = \{2m(t)E^t[x^2(t)] + \sigma^2(t)\}dt + k(t)\{E^t[x^3(t)] - E^t[x^2(t)]\hat{x}(t)\}\{dz(t) - k(t)\hat{x}(t)dt\}. \tag{6.13}$$

Since the conditional distribution of $x(t)$ given $\mathcal{F}_t$ is Gaussian with mean $\hat{x}(t)$ and variance $R(t)$ (Lemma 6.2) we have

$$E^t[x^3(t)] = [\hat{x}(t)]^3 + 3[\hat{x}(t)]R(t).$$

Hence

$$\begin{aligned} E^t[x^3(t)] - E^t[x^2(t)]\hat{x}(t) &= [\hat{x}(t)]^3 + 3[\hat{x}(t)]R(t) - \hat{x}(t)[R(t) + \{\hat{x}(t)\}^2] \\ &= 2\hat{x}(t)R(t). \end{aligned}$$

From equation (6.13) we then obtain the following stochastic differential for $E^t[x^2(t)]$:

$$dE^t[x^2(t)] = \{2m(t)E^t[x^2(t)] + \sigma^2(t)\}dt + 2k(t)R(t)\hat{x}(t)\,\{dz(t) - k(t)\hat{x}(t)dt\}. \tag{6.14}$$

From (6.10) and applying Ito's lemma to $[\hat{x}(t)]^2$ we have

$$(6.15)\qquad d[\hat{x}(t)]^2 = \{2m(t)[\hat{x}(t)]^2 + k^2(t)R^2(t)\}dt + 2k(t)R(t)\hat{x}(t)\{dz(t) - k(t)\hat{x}(t)dt\}.$$

Finally from Ito's lemma applied to

$R(t) = E^t[x^2(t)] - [\hat{x}(t)]^2$ and from (6.14) and (6.15)

it follows that $R(t)$ has the following stochastic differential from which, however, the random term is absent.

$$(6.16)\qquad dR(t) = \{2m(t)R(t) + \sigma^2(t) - k^2(t)R^2(t)\}dt \qquad (0 \le t \le T).$$

Hence $R(t)$ is, almost surely, a nonrandom function of t $(0 \le t \le T)$ satisfying the ordinary differential equation (6.11). The proof of Theorem 6.2 is complete.

In view of (6.10) it follows that $\hat{x}(t)$ is a Gaussian process which is almost surely sample continuous.

References

[1] Bucy; R.S., "Nonlinear filtering theory", IEEE Trans. Automatic Control, 10 (1965).

[2] Doob, J.L. Stochastic Processes, J. Wiley, New York (1952).

[3] Dynkin, E.B. Markov Processes-I, (English translation), Springer-Verlag (1965).

[4] Gikhman, I.I. and Skorokhod, A.V. Vvedeniye v teoriyu sluchainikh protsessov, Moscow (1965).

[5] Ito, K., Lectures on Stochastic processes, Tata Institute of Fundamental Research, Bombay (1961).

[6] Kallianpur G. and Striebel, C., "Estimation of stochastic processes: Arbitrary system process with additive white noise observation errors", to appear in the Ann. Math. Stat.

[7] Kallianpur G. and Striebel, C., "Stochastic differential equations occuring in the estimation of continuous parameter stochastic processes", (submitted for publication).

[8] Kalman R.E. and Bucy, R.S., "New results in linear filtering and prediction theory", J. Basic Eng. 83, Series D (1961).

[9] Kushner, H.J., "Dynamical equations for optimal nonlinear filtering," Journal of Differential Equations 3 (1967).

[10] Liptzer R. Sh., and Shiyaev A.N., Nonlinear filtering of Markov diffusion processes , I and II, to appear in the Trudy M.I.A.N. Vol 104.

[11] Loeve, M., Probability Theory, Third edition, Van Nostrand, New York (1963).

[12] Skorokod A.V., Sluchainie protsessi s nezavisimimi prirasheniyami, Moscow (1964)

[13] Skorokhod A.V., Studies in the Theory of Random Processes, (English translation), Addison-Wesley (1965)

[14] Wnham, W.M. "Some applications of stochastic differential equations to optimal nonlinear filtering.", J. SIAM Control, Series A, 2 (1965).