Fujisaki, M., Kallianpur, G. and Kunita, H.
Osaka J. Math.
9 (1972), 19-40

STOCHASTIC DIFFERENTIAL EQUATIONS FOR THE NON LINEAR FILTERING PROBLEM*

MASATOSHI FUJISAKI, G. KALLIANPUR AND HIROSHI KUNITA

(Received April 12, 1971)

1. Introduction

The general nonlinear filtering or estimation problem may be described as follows. x_t, $(0 \leq t \leq T)$, called the signal or system process is a stochastic process direct observation is not possible. The data concerning x_t is provided by observation on another process z_t which is related to x_t by the model (2.1). (See Sections 2 and 4 for notation and precise definitions). In general it is assumed that x_t takes values in a complete separable metric space while z_t is an n-dimensional process. The least squares estimate of $f(x_t)$ (where f is a suitable real valued function) based on the observations $(z_\tau, 0 \leq \tau \leq t)$ is given by the conditional expectation $E[f(x_t) | z_\tau, 0 \leq \tau \leq t]$. In the general case this estimate depends non-linearly on the observations and is known as the non linear filter. A "Bayes" formula for the conditional expecation has been given in [9] but is useful in applications only when t is fixed. If the data is coming in continuously and we require an estimate which can be continuously revised to take into account the new data, this formula, while valid, is not practical since the estimate at a future time $t+\Delta$ must be computed using all the past data. The formula computed for time t is of no help in computing the estimate at $t+\Delta$. A practical as well as mathematically more interesting way of doing this is by obtaining a stochastic differential equation for the filter.

This problem has acquired a growing literature in recent years. The papers having a direct bearing on our results are the ones by Kallianpur and Striebel ([10], [11]), Shiryaev [18] and Liptzer and Shiryaev [13]. In [18] x_t is assumed to be a Markovian jump process and in [13] the system and observation processes are components of a diffusion process governed by a stochastic differential equation. The results closest in spirit to the present paper are those in [11] where x_t is a Markov process in R^n and independent of the "noise" process w_t. In that paper the Bayes formula for the filter given in [9] is used in deriving the corresponding stochastic differential equation.

Our paper differs from the above mentioned work in two essential res-

* The second author was supported in part by NSF Grant GP-1188-8.

pects. First, we develop the innovation process approach first considered by T. Kailath [7]. The method we use relies on a basic result due to I.V. Girsanov [5] and on the representation of square integrable martingales. The second difference consists in replacing the assumption of complete independence of (x_t) and (w_t) by a more natural assumption which immediately enlarges the scope of the applications. We are thus able not only to unify and extend the results of [11] and present simpler proofs but also to treat the filtering problem for controlled system processes.

Section 2 introduces the basic model (2.1) connecting the observation and system process and also the innovation Wiener process. In Section 3 is derived the stochastic integral representation for a square integrable martingale $(Y_t, \mathcal{F}_t, P)$ where $\mathcal{F}_t$ is the σ-field $\sigma\{z_\tau, 0 \leq \tau \leq t\}$. The main result is proved in Section 4. In Section 5 we study the filtering problem in a more general setting though still within the framework of the main theorem. As a corollary we obtain the stochastic differential equation for $E[f(x_t) | z_\tau, 0 \leq \tau \leq t]$ when (x_t) is a controlled process, or more precisely, when (x_t) and (z_t) are given by functional stochastic differential equations of the type considered by K. Ito-M. Nisio [6] and by W. Fleming-M. Nisio [3]. The Markov process case is discussed in detail in the last section.

We thank Professors Thomas Kailath and Charlotte Striebel for their valuable comments and suggestions.

2. Observation process and the associated innovation process

The *system or signal process* $h_t(\omega)$ and the *observation process* $z_t(\omega)(t \in [0, T])$ are assumed given on some probability space $(\Omega, \mathcal{A}, P)$ and further related as follows.

(2. 1) $$z_t = \int_0^t h_u du + w_t ,$$

where

(2. 2) w_t is an N-vector standard Wiener process

and

(2. 3) $h_t(\omega)$ is a (t, ω)-measurable N-vector process such that $\int_0^T E|h_t|^2 \mathrm{dt} < \infty$. Here $|\quad|$ denotes the norm of the N-vector.

In order to take into account applications involving stochastic control we abandon as unrealistic the condition of complete independence of (h_t) and (w_t) and substitute the more natural condition.

(2. 4) For each s, the σ-fields

$$\sigma\{h_u, w_u;\ 0<u<s\},\ \sigma\{w_v-w_u;\ s<u<v<T\}$$

are independent.

Let

$$\mathcal{F}_t = \sigma\{z_s;\ 0<s<t\}\ . \tag{2. 5}$$

The monotone family $(\mathcal{F}_t)$ represents all the data available concerning the process (h_t). We observe before proceeding further that if $(\mathcal{M}_t)$ is a monotone family of sub σ-fields of $\mathcal{A}$, then in view of condition (2.3) it can be shown that there exists a mondification of $E(h_t \mid \mathcal{M}_t)=(E(h_t^1 \mid \mathcal{M}_t), \cdots, E(h_t^N \mid \mathcal{M}_t))$ which is jointly measurable and adapted to $(\mathcal{M}_t)$. In other words there exists an N-vector function $H(t, \omega)$ with the latter properties such that for a.e. t in $[0, T]$, $E(h_t \mid \mathcal{M}_t)(\omega)=H(t, \omega)$ a.s. P. In what follows we shall always work with such modifications. The same remark will apply to conditional expectations of other processes to be encountered later. (Sections 4 and 5). Writing $\hat{h}_t=E(h_t \mid \mathcal{F}_t)$ let us define the so-called "innovation" process

$$\nu_t = z_t - \int_0^t \hat{h}_s ds\ . \tag{2. 6}$$

The starting point of our work is the fact that $(\nu_t, \mathcal{F}_t, P)$ is a Wiener process. Proofs of this result under varying hypotheses have appeared in earlier works ([7], [8], [13]) and the importance of the innovation approach to filtering problems has been particularly stressed by Kailath ([7]). Let us state Doob's theorem concering the standard Wiener process in a slightly sharper form, since our proof of the innovation process is based on this.

Lemma 2.1. (Doob [2], Chapter 7, Theorem 11.9, Kunita-Watanabe [12], Theorem 2.3). *Let $\beta_t=(\beta_t^1, \cdots \beta_t^N)$ be an N-vector process with continuous sample paths such that $(\beta_t^i, \mathcal{M}_t, P)$ is a square integrable martingale. If*

$$E[(\beta_t^i-\beta_s^i)(\beta_t^j-\beta_s^j) \mid \mathcal{M}_s] = (t-s)\delta_{ij} \qquad {}^\forall t > {}^\forall s > 0$$

is satisfied, then β_t is a standard Wiener process. Moreover, $\sigma\{\beta_v-\beta_u;\ s\le u<v\le T\}$ is independent of $\mathcal{M}_s$

Proof. The result except the last assertion is found in the above cited reference. Also, the independence of $\beta_t-\beta_s$ and $\mathcal{M}_s$ is found in [12]. Now let $s=t_1<t_2<\cdots<t_n<t_{n+1}=T$ and $\{\alpha_k\}_{k=1}^n$ be a sequence of constant N-vectors. Then $E[e^{i(\alpha_{t_k}, \beta_{t_{k+1}}-\beta_{t_k})} \mid \mathcal{M}_{t_k}]=E(e^{i(\alpha_{t_k}, \beta_{t_{k+1}}-\beta_{t_k})})$ holds. Hence

$$E[\prod_{k=1}^n e^{i(\alpha_{t_k}, \beta_{t_{k+1}}-\beta_{t_k})} \mid \mathcal{M}_s]$$

$$= E[E(e^{i(\alpha_{t_n}, \beta_{t_{n+1}}-\beta_{t_n})} \mid \mathcal{M}_{t_n}) \prod_{k=1}^{n-1} e^{i(\alpha_{t_k}, \beta_{t_{k+1}}-\beta_{t_k})} \mid \mathcal{M}_s]$$

$$= E[e^{i(\alpha_{t_n}, \beta_{t_{n+1}} - \beta_{t_n})}] E[\prod_{k=1}^{n-1} e^{i(\alpha_{t_k}, \beta_{t_{k+1}} - \beta_{t_k})} | \mathcal{M}_s] = \cdots$$

$$= \prod_{k=1}^{n} E[e^{i(\alpha_{t_k}, \beta_{t_{k+1}} - \beta_{t_k})}] .$$

This proves that $\sigma\{\beta_{t_{k+1}} - \beta_{t_k};\ k=1, \cdots, n\}$ is independent of $\mathcal{M}_s$. The proof is complete.

Lemma 2.2. *Under assumptions* (2.1)$\sim$(2.4), $(\nu_t, \mathcal{F}_t, P)$ *is an N-vector standard Wiener process. Furthermore* $\mathcal{F}_s$ *and* $\sigma\{\nu_v - \nu_u;\ s \leq u < v \leq T\}$ *are independent.*

Proof. From (2.6) we have for $s<t$,

$$E[\nu_t | \mathcal{F}_s] = \nu_s + E[\int_s^t \bar{h}_u du + w_t - w_s | \mathcal{F}_s] \tag{2.7}$$

where $\bar{h}_u = h_u - \hat{h}_u$. The second term on the right hand side of (2.7) is clearly 0 since $E[\bar{h}_u | \mathcal{F}_u] = 0$ and $E[w_t - w_s] = 0$. The last statement is true because $w_t - w_s$ is indepent of $\mathcal{F}_s$. Hence $(\nu_t, \mathcal{F}_t, P)$ is a martingale, that is $(\nu_t^i, \mathcal{F}_t, P)$ is a martingale for each i, $1 \leq i \leq N$. Now let $\prod_n = \{t_k^n\}$ be a finite partition of $[s, t]$ and let $\sigma_n^{ij} = \sum_k (\nu^i{}_{t_{k+1}^n} - \nu^i{}_{t_k^n}) (\nu^j{}_{t_{k+1}^n} - \nu^j{}_{t_k^n})$.
Then making use of (2.3) it is easy to see

$$E[\sum_k | \int_{t_k^n}^{t_{k+1}^n} \bar{h}_u \, \mathrm{du} |^2]$$

tends to zero as $n \to \infty$ and $\max_k (t_{k+1}^n - t_k^n) \to 0$. A direct calculation shows

$$E[| \sum (w^i{}_{t_{k+1}^n} - w^i{}_{t_k^n}) (w^j{}_{t_{k+1}^n} - w^j{}_{t_k^n}) - (t-s)\delta_{ij} |^2] \to 0 . \tag{2.8}$$

It follows that

$$\lim_{n\to\infty} E | \sigma_n^{ij} - (t-s)\delta_{ij} | = 0 . \tag{2.9}$$

Finally since

$$E[(\nu_t^i - \nu_s^i)(\nu_t^j - \nu_s^j) | \mathcal{F}_s] = E[\sigma_n^{ij} | \mathcal{F}_s]$$

for every $\prod_n$ and since

$$E[| E(\sigma_n^{ij} - (t-s)\delta_{ij} | \mathcal{F}_s) |] \leq E | \sigma_n^{ij} - (t-s)\delta_{ij} | \to 0 \tag{2.10}$$

from (2.9) we obtain

$$E[(\nu_t^i - \nu_s^i)(\nu_t^j - \nu_s^j) | \mathcal{F}_s] = (t-s)\delta_{ij} . \tag{2.11}$$

The conclusion of the lemma follows from the previous lemma.

REMARK 2.1. It is worth pointing out that Lemma 2.2 can be established under weaker conditions on (h_t). First of all, in proving that $(\nu_t, \mathcal{F}_t)$ is a martingale we need the existence of a jointly measurable modification of $E(h_u|\mathcal{F}_u)$. We need it and also the finiteness of $\int_0^T E|h_u|du$ to validate the steps which show that the second term on the right hand side of (2.7) vanishes. Given this, one sees that for $(\nu_t, \mathcal{F}_t)$ to be an L^2-martingale it is necessary and sufficient that for each $s<t$,

$$E|\int_s^t \bar{h}_u du|^2 < \infty . \tag{2.12}$$

Assuming these conditions the proof of the lemma is modified as follows; Defining σ_n^{ij} as before we have

$$\sum_k |\int_{t_k^n}^{t_{k+1}^n} \bar{h}_u du|^2 \leq \max_k |\int_{t_k^n}^{t_{k+1}^n} \bar{h}_u du| (\sum_k |\int_{t_k^n}^{t_{k+1}^n} (h_u - \hat{h}_u) du|)$$

$$\leq [\max_k |\int_{t_k^n}^{t_k^n+1} |h_u| du + \max_k \int_{t_k^n}^{t_{k+1}^n} |\hat{h}_u| du] [\int_0^T |\bar{h}_u| du]$$

which tends to 0 a.s. as $\max_k(t_{k+1}^n - t_k^n) \to 0$. Secondly (2.8) holds. Hence it follows that there is a sequence Π_n of partitions such that $\sigma_n^{ij} \to (t-s)\delta_{ij}$ a.s. Now the sequence σ_n^{ij} of random variables is known to be uniformly integrable (see [16], Appendix), so that (2.10) holds. The rest of the proof is unchanged. It has been shown that with the measurability condition mentioned and the assumption

$$\int_0^T E|h_u| du < \infty \tag{2.13}$$

$(\nu_t, \mathcal{F}_t, P)$ is a martingale which is square integrable if and only if it is a Wiener process

REMARK 2.2. It should be noted that the innovation process ν_t can be calculated from the observation data z_s, $s \leq t$. In fact, since ν_t is an $(\mathcal{F}_t, P)$-martingale, we have

$$\lim_{h\to 0} \frac{1}{h} E[z_{s+h} - z_s | \mathcal{F}_s] = \lim_{h\to 0} \frac{1}{h} \int_s^{s+h} E(h_u|\mathcal{F}_s) \mathrm{du} = \hat{h}_s \qquad \text{a.s.} \quad (s, \omega) .$$

This shows that the two processes $\int_0^t \hat{h}_s ds$ and ν_t are determined directly by the observation data z_s, $s \leq t$.

By the definition of the innovation process, ν_t is $(\mathcal{F}_t)$-measurable. It is conjectured that the converse would be also true, that is $\mathcal{F}_t = \sigma\{\nu_s;\ s \leq t\}$. It seems to us that no satisfactory result is known to this problem. We shall prove in the next section a slightly weaker assertion: Every separable $\mathcal{F}_t$-martingale is represented as a stochastic integral by the innovation process. This will play a fundamental role in deriving stochastic differential equation for the filtering process. (Sections 4 and 5).

3. Stochastic integral representation of a separable martingale on $(\Omega, \mathcal{F}_t, P)$

Theorem 3.1. *Under conditions* (2.1), (2.2), (2.3) *and* (2.4) *every separable square integrable martingale* $(Y_t, \mathcal{F}_t, P)$ *is sample continuous and has the representation*

$$Y_t - E[Y_0] = \int_0^t (\Phi_s, d\nu_s) \equiv \sum_{i=1}^N \int_0^t \Phi_s^i \, d\nu_s^i \tag{3. 1}$$

where

$$\int_0^T E|\Phi_s|^2 ds < +\infty \tag{3. 2}$$

and $\Phi_s = (\Phi_s^1, \cdots, \Phi_s^N)$ *is jointly measurable and adapted to* $(\mathcal{F}_s)$.

Since our proof is based on a theorem of Girsanov [5], we shall state it here in a modified form for convenience of later reference.

Lemma 3.1. *Let* (β_t) $(0<t<T)$ *be an N-vector standard Wiener process on* $(\Omega, \mathcal{M}_t, P)$ *and* $\varphi(t, \omega)$, *a jointly measurable N-vector process adapted to* $(\mathcal{M}_t)$ *and further satisfies*

$$\int_0^T E|\varphi_t|^2 dt < \infty. \tag{3. 3}$$

Define

$$\alpha_s^t(\varphi) = exp\,[\int_s^t (\varphi_u, d\beta_u) - \frac{1}{2}\int_s^t |\varphi_u|^2 du]\,, \tag{3. 4}$$

$$\begin{aligned} T_n &= inf\{t;\, 0 \le t \le T\,,\ \alpha_0^t(\varphi) > n \ or \int_0^t |\varphi_s|^2 ds > n\} \\ &= T \qquad \textit{if the above set } \{\cdots\} \textit{ is empty.} \end{aligned} \tag{3. 5}$$

Then for each n the measure

$$\tilde{P}_n = \alpha_0^{T_n} dP \tag{3. 6}$$

is a probability measure and the process

$$\tilde{\beta}_t^n = \beta_t - \int_0^{t \wedge T_n} \varphi_u du \tag{3. 7}$$

is an N-vector standard Wiener process in $(\mathcal{M}_t, \tilde{P}_n)$.

Proof. Set $\varphi_t^n = \varphi_t I_{[0,T_n]}(t)$, where I is the indicator function of the set $[0, T_n]$. Then $\alpha_0^{t\wedge T_n}(\varphi) = \alpha_0^t(\varphi^n)$ and $\tilde{\beta}_t^n = \beta_t - \int_0^t \varphi_s^n ds$ holds. We shall drop the index n for notational convenience in the following discussion. Ito's formula [14] implies

$$\alpha_0^t = 1 + \int_0^t (\alpha_0^s \varphi_s, d\beta_s)\,. \tag{3.8}$$

The last member is a square-integrable martingale because

$$\int_0^T E((\alpha_0^s)^2 |\varphi_s|^2) ds \leq n^2 E\Big(\int_0^T |\varphi_s|^2 ds\Big) \leq n^4\,.$$

Therefore we have $E(\alpha_0^T) = 1$ by (3.8), proving that $\tilde{P}$ is a probability measure.

We shall next prove that $\tilde{\beta}_t$ is a square-integrable $(\mathcal{M}_t, \tilde{P})$-martingale. The square integrability is clear since α_0^T is dominated by n. By the definition of the conditional expectation,

$$\tilde{E}(\tilde{\beta}_t | \mathcal{M}_s) = E(\alpha_0^t \tilde{\beta}_t | \mathcal{M}_s)(\alpha_0^s)^{-1}, \qquad t > s$$

so that $\tilde{\beta}_t$ is an $(\mathcal{M}_t, \tilde{P})$-martingale if and only if $\alpha_0^t \tilde{\beta}_t$ is an $(\mathcal{M}_t, P)$-martingale. Now Ito's formula implies

$$\int_0^t (\alpha_0^s \varphi_s, d\beta_s) \times \beta_t^i = P\text{-martingale} + \int_0^t \alpha_0^s \varphi_s^i ds \tag{3.9}$$

and

$$\Big(\int_0^t (\alpha_0^s \varphi_s, d\beta_s)\Big) \times \int_0^t \varphi_s^i ds = P\text{-martingale} + \int_s^t \Big(\int_0^s (\alpha_0^u \varphi_u, d\beta_u)\Big) \varphi_s^i ds \tag{3.10}$$

$$= P\text{-martingale} + \int_0^t \alpha_0^s \varphi_s^i ds - \int_0^t \varphi_s^i ds\,,$$

the last equality following from (3.8). Thus we get

$$\alpha_0^t \tilde{\beta}_t^i = \Big(1 + \int_0^t (\alpha_0^s \varphi_s, d\beta_s)\Big)\Big(\beta_t^i - \int_0^t \varphi_s^i ds\Big) = P\text{-martingale}$$

by (3.9) and (3.10), proving that $(\tilde{\beta}_t, \mathcal{M}_t, \tilde{P})$ is a square integrable martingale with continuous sample paths. Now the argument in the proof of Lemma 2.2 can be applied to the present case, replacing $(\nu_t, \bar{h}_t, w_t, \mathcal{F}_t, P)$ by $(\tilde{\beta}_t, -\varphi_t, \beta_t, \mathcal{M}_t, \tilde{P})$ there. We then have

$$\tilde{E}[(\tilde{\beta}_t^i - \tilde{\beta}_s^i)(\tilde{\beta}_t^j - \tilde{\beta}_s^j) | \mathcal{M}_s] = (t-s)\delta_{ij}\,.$$

The proof is complete.

Proof of Theorem 3.1: Set $\beta_t = \nu_t$, $\varphi_t = -\hat{h}_t$ and $\mathcal{M}_t = \mathcal{F}_t$ and apply the preceding lemma. Then we see that

$$z_t^n = \nu^t + \int_0^{t\wedge T_n} \hat{h}_s ds \tag{3.11}$$

is an $(\mathcal{F}_t, \tilde{P}_n)$-standard Wiener process. Set $\mathcal{F}_t^n = \sigma\{z_s^n ; s<t\}$. It is well known that every separable square integrable $(\mathcal{F}_t^n, \tilde{P}_n)$-martingale $\tilde{Y}_t$ is sample continuous and is represented as $\tilde{Y}_t = \int_0^t (\tilde{\Phi}_s^n, dz_s^n)$[1]. Note that $z_s = z_s^n$ holds for $s<T_n$ and apply Doob's optional sampling theorem. Then we see that $\tilde{Y}_{t\wedge T_n} = \int_0^{t\wedge T_n} (\tilde{\Phi}_s^n, dz_s)$ holds and it is an $(\mathcal{F}_{t\wedge T_n}^n, \tilde{P}_n)$-martingale, where

$$\mathcal{F}_{t\wedge T_n}^n = \{B\in\mathcal{F}_t^n ; B\cap\{T_n\le s\}\in\mathcal{F}_s^n \quad \text{for all } 0\le s\le T\}\ . \tag{3.12}$$

On the other hand, it is easily seen that

$$\mathcal{F}_{t\wedge T_n}^n = \sigma\{z_{s\wedge T_n}^n ; 0\le s\le t\} = \sigma\{z_{s\wedge T_n} ; 0\le s\le t\} = \mathcal{F}_{t\wedge T_n}\ ,$$

where the σ-field $\mathcal{F}_{t\wedge T_n}$ is defined by (3.12) replacing $\mathcal{F}_t^n$ by $\mathcal{F}_t$. We have thus seen that if $\tilde{Y}_t$ is a separable square-integrable $(\mathcal{F}_{t\wedge T_n}, \tilde{P}_n)$-martingale, it is represented as

$$\tilde{Y}_t = \tilde{Y}_{t\wedge T_n} = \int_0^{t\wedge T_n} (\tilde{\Phi}_s^n, dz_s)\ . \tag{3.13}$$

Suppose now that Y_t is a separable square-integrable $(\mathcal{F}_t, P)$-martingale and let $\tilde{Y}_t = (\alpha_0^t)^{-1} Y_t$. Then $\tilde{Y}_{t\wedge T_n}$ is a separable square-integrable $(\mathcal{F}_{t\wedge T_n}, \tilde{P}_n)$-martingale as we have noted in the proof of the preceding lemma so that it has the representation (3.13). Consequently,

$$\tilde{Y}_{t\wedge T_n} = \int_0^{t\wedge T_n} (\tilde{\Phi}_s^n, d\nu_s) + \int_0^{t\wedge T_n} (\tilde{\Phi}_s^n, \hat{h}_s) ds\ . \tag{3.14}$$

Ito's formula applied to $Y_{t\wedge T_n} = \alpha_0^{t\wedge T_n} \tilde{Y}_{t\wedge T_n}$ enables us to write

$$Y_{t\wedge T_n} = \int_0^{t\wedge T_n} (\Phi_s^n, d\nu_s) + \int_0^{t\wedge T_n} \Psi_s^n ds\ .$$

The second term of the right hand vanishes a.s. since it is an $(\mathcal{F}_{t\wedge T_n}, P)$-martingale with bounded variation ([12], Corollary to Theorem 1.3). We have thus obtained

$$Y_{t\wedge T_n} = \int_0^{t\wedge T_n} (\Phi_s^n, d\nu_s)\ , \tag{3.15}$$

where $\int_0^T E|\Phi_s^n|^2 ds \le E[Y_T^2] < \infty$. The uniqueness of the representation (3.15)

1) We may and do assume that $\tilde{Y}_0 = 0$. The same remark is applied to other martingales.

yields $\Phi_s^n = \Phi_s^m$ if $s < T_n$ and $n < m$, i.e., there exists Φ_s adapted to $(\mathcal{F}_t)$ such that $\Phi_s = \Phi_s^n$ for $s < T_n$ and $\int_0^T E|\Phi_s|^2 ds < \infty$. The proof is complete.

Corollary to Theorem 3.1. *Every separable martingale* $(Y_t, \mathcal{F}_t, P)$ *is sample continuous and has the representation* (3.1), *where* $\int_0^T |\Phi_s|^2 ds < \infty$ *and* Φ_s *is jointly measurable and adapted to* $(\mathcal{F}_s)$.

The proof is based on Theorem 3.1 and follows, almost without change, the arguments in Theorem 3 of J.M.C. Clark's recent paper ([1], p. 1291). To avoid misunderstanding it should be pointed out that Clark has proved the theorem in case $\mathcal{F}_t = \sigma\{\nu_s;\ s \le t\}$. We are not sure in the present context if the above two σ-fields coincide or not.

In many applications, it may happen that the signal and the observation processes are related in more complicated forms. We shall introduce in the next two sections two types of observation processes and derive stochastic differential equations satisfied by the filtering processes.

4. A stochastic differential equation for the non-linear filtering problem

The signal process that we consider in this section is denoted by $x_t(\omega)$, $t \in [0, T]$ and is assumed to take values in a complete metric space S. We assume as in Section 2 that

(4 .1) For each s, the σ-fields

$$\sigma\{x_u, w_u;\ u \le s\}, \quad \sigma\{w_v - w_u;\ s < u < v \le T\}$$

are independent. The observation process is again denoted by z_t and is defined by (2.1), where $h_t(\omega)$ is an N-vector process satisfying (2.3) and

(4. 2) For each $s < T$, $h_s(\omega)$ is $\sigma\{x_u, w_u;\ u \le s\}$-measurable.

Then conditions (4.1) and (4.2) imply that the new observation process z_t satisfies (2.4). Thus the three processes (w_t, h_t, z_t) satisfies conditions (2.1)~(2.4) of Section 2.

Let us introduce the following notation for the family of σ-fields.

$$\mathcal{F}_t = \sigma\{z_s;\ s \le t\}\ , \quad \mathcal{G}_t = \sigma\{x_s, w_s;\ s \le t\}\ . \tag{4. 3}$$

Let f be a real measurable function on S such that

$$E|f(x_t)|^2 < \infty \qquad \text{for all } 0 \le t \le T\ . \tag{4. 4}$$

The function f is said to belong to space $\boldsymbol{D}(\hat{A})$ if there exists a jointly (t, ω)-

measurable real function $\hat{A}_t f(\omega)$ adapted to $\sigma(x_s, z_s; s \leq t)$ such that

$$\int_0^T E|\hat{A}_t f|^2 dt < \infty \tag{4.5}$$

and

$$M_t(f) \equiv f(x_t) - E[f(x_0)|\mathcal{F}_0] - \int_0^t \hat{A}_s f ds \tag{4.6}$$

is a $(\mathcal{G}_t, P)$-martingale. Such a function $\hat{A}_t f(\omega)$ is at most unique up to (t, ω)-measure 0. We assume in the following that $\boldsymbol{D}(\hat{A})$ is non-empty. Then $\boldsymbol{D}(\hat{A})$ is a linear space and $\hat{A}_t$ is a linear transformation from $\boldsymbol{D}(\hat{A})$ into $L^2(\Omega)$ for a.e. t.

REMARK. Let $x^t = (x_s, 0 \leq s \leq t)$ and $z^t = (z_s, 0 \leq s \leq t)$. Then in case that $((x^t, z^t), \mathcal{F}_t, P)$ is a Markov process, the above operator $\hat{A}_t$ is a stochastic analogue of the generator defined for the function $f(x_t)$ depending only on the component x_t. We shall discuss this problem in Section 6.

Lemma 4.1. *Let $f \in \boldsymbol{D}(\hat{A})$ and let*

$$\bar{M}_t(f) = E[f(x_t)|\mathcal{F}_t] - E[f(x_0)|\mathcal{F}_0] - \int_0^t E[\hat{A}_s f|\mathcal{F}_s] ds \tag{4.7}$$

Then $(\bar{M}_t(f), \mathcal{F}_t, P)$ is a square integrable martingale.

Proof. The square integrability of (4.7) is obvious from (4.4) and (4.5). Observe that the σ-fields $(\mathcal{F}_s)$ are monotone and $\mathcal{F}_s \subset \mathcal{G}_s$. Then for $s < t$, $E[\bar{M}_t(f) - \bar{M}_s(f)|\mathcal{F}_s]$ coincides with

$$\begin{aligned} & E[f(x_t)|\mathcal{F}_s] - E[f(x_s)|\mathcal{F}_s] - E\left[\int_s^t E[\hat{A}_u f|\mathcal{F}_u|]du|\mathcal{F}_s\right] \\ &= E\left[f(x_t) - f(x_s) - \int_s^t \hat{A}_u f\, du|\mathcal{F}_s\right] \\ &= E[M_t(f) - M_s(f)|\mathcal{F}_s] \\ &= E[E[M_t(f) - M_s(f)|\mathcal{G}_s]|\mathcal{F}_s], \end{aligned}$$

the last term being zero since $M_t(f)$ is a $(\mathcal{G}_t, P)$-martingale. The proof is complete.

In the proof of the principal theorem of this section we shall need a fact which it seems convenient to separate out as a lemma although it is hardly more than an observation and follows directly from H. Kunita-S. Watanabe [12] and Meyer [16].

Lemma 4.2. *Let $(M_t(f), \mathcal{G}_t, P)$ be the square integrable martingale of (4.6). Then there exist unique sample continuous processes $\langle M(f), w^i \rangle$ $(i = 1, \cdots, N)$ adapted to $(\mathcal{G}_t)$ such that almost all sample functions are of bounded variation and*

$M_t(f)w_t^i - \langle M(f), w^i \rangle_t$ are $\mathcal{G}_t$-martingales. Furthermore each $\langle M(f), w^i \rangle_t$ has the following properties: It is absolutely continuous with respect to Lebesgue measure in $[0, T]$. There exists a modification of the Radon-Nikodym derivative which is (t, ω)-measurable and adapted to $(\mathcal{G}_t)$ and which we shall denote by $\tilde{D}_t^i f(\omega)$. Then using the vector notation $\tilde{D}_t f = (\tilde{D}_t^1 f, \cdots, \tilde{D}_t^N f)$,

$$\langle M(f), w \rangle_t = \int_0^t \tilde{D}_s f ds \qquad a.s, \tag{4.8}$$

where

$$\int_0^T E|\tilde{D}_s f|^2 ds < \infty . \tag{4.9}$$

If the process (x_t) and (w_t) are completely independent then a.s.

$$\langle M(f), w \rangle_t = 0 . \tag{4.10}$$

Proof. The first part of the lemma, (4.8) and (4.9) follow from [12, 16] upon noting that $(M_t(f), \mathcal{G}_t)$ and $(\nu_t, \mathcal{G}_t)$ are both square integrable martingales and that $\langle w^i, w^j \rangle_t = t\delta_{ij}$. To show (4.10) it suffices to prove the equivalent assertion that $(M_t(f)w_t, \mathcal{G}_t)$ is a martingale. For $s<t$, a direct calculation shows

$$E[(M_t(f)-M_s(f))(w_t-w_s)|\mathcal{G}_s] = E[M_t(f)w_t|\mathcal{G}_s]-M_s(f)w_s .$$

On the other hand, the assumption of the complete idnependence of (x_t) and (w_t) processes implies that $w_t - w_s$ and $\mathcal{G}_s \vee \sigma\{M_t(f)-M_s(f)\}$ are indpendent. Hence

$$E[(M_t(f)-M_s(f))(w_t-w_s)|\mathcal{G}_s] = E[M_t(f)-M_s(f)|\mathcal{G}_s]E[w_t-w_s] = 0$$

a.s., thus proving (4.10).

The lemma just proved indicates the possibility that in the situation where in place of the independence of (x_t) and (w_t) we only have (4.1) the stochastic differential equation we seek might have a more general form the one derived in [10]. As we shall see below this is indeed the case.

In our next theorem it will be understood that we are always considering separable versions of the martingales of $M_t(f)$ and $\bar{M}_t(f)$. We shall also use the shorter $E^t(\cdot)$ for $E(\cdot|\mathcal{F}_t)$.

Theorem 4.1. *Assume* (4.1), (4.2) (*and* (2.1), (2.3)). *If $f \in \boldsymbol{D}(\hat{A})$ satisfies*

$$\int_0^T E|f(x_t)h_t|^2 dt < \infty \tag{4.11}$$

then $E^t[f(x_t)]$ satisfies the following stochastic differential equation

$$E^t[f(x_t)] = E[f(x_0)] + \int_0^t E^s[\hat{A}_s f] ds \tag{4.12}$$

$$+\int_0^t ([E^s(f(x_s)h_s)-E^s(f(x_s))E^s(h_s)+E^s(\tilde{D}_s f)]\,,\,d\nu_s)\,.$$

Proof. The equation (4.12) is equivalent to that $\bar{M}_t(f)$ equals the last term in (4.12) involving the stochastic integral, which we shall denote as $M_t^*(f)$. The proof is then reduced to proving $E[\bar{M}_t(f)Y_t]=E[M_t^*(f)Y_t]$ for all Y_t such that

$$Y_t=\int_0^t (\Phi_s,\,d\nu_s) \qquad (\Phi_s\text{: bounded}), \tag{4.13}$$

since such Y_t is dense in $L^2(\mathcal{F}_t, P)$ (up to constants) by virtue of Theorem 3.1. We shall show this calculating $E[(M_t(f)-\bar{M}_t(f))Y_t]$ and $E[M_t(f)Y_t]$ separately. In what follows we write M_t, $\bar{M}_t$ and M_t^* suppressing f as it is fixed througout the argument. Using (4.6) and (4.7) a simple calculation yields

$$E[(\bar{M}_t-M_t)Y_t]=E\Big[\int_0^t (Y_t-Y_s)\hat{A}_s f\,ds\Big]\,. \tag{4.14}$$

From (4.13) writing $Y_t=\int_0^t (\Phi_s,\,dw_s)+\int_0^t (\Phi_s,\,h_s-\hat{h}_s)ds$ the right hand side of (4.14) is reduced to the form

$$\int_0^t E\Big[\hat{A}_s f\int_s^t (\Phi_u,\,dw_u)\Big]ds+E\Big[\int_0^t \hat{A}_s f\Big(\int_s^t (\Phi_u,\,h_u-\hat{h}_u)\,du\Big)ds\Big]\,. \tag{4.15}$$

The integrand in the first term of (4.15) is zero because $\hat{A}_s f$ is $\mathcal{G}_s$-measurable and $E\Big(\int_s^t (\Phi_u,\,dw_u)\,|\,\mathcal{G}_s\Big)=0$. The latter fact follows since $\int_0^t (\Phi_s,\,dw_s)$ is a $\mathcal{G}_t$-martingale. The quantity inside the brackets in the second term of (4.15) becomes (after an integration by parts) $\int_0^t \Big[\int_0^s \hat{A}_u f\,du\Big](\Phi_s,\,h_s-\hat{h}_s)ds$. Hence the right hand side of (4.14) equals

$$E\Big[\int_0^t \Big[\int_0^s \hat{A}_u f\,du\Big]\,(\Phi_s,\,h_s-\hat{h}_s)ds\Big]\,. \tag{4.16}$$

On the other hand it is easy to verify that

$$\begin{aligned} E(M_t Y_t) &= E\Big[M_t\int_0^t (\Phi_s,\,dw_s)\Big]+E\Big[\int_0^t f(x_s)(\Phi_s,\,h_s-\hat{h}_s)ds\Big] \\ &\quad -E\Big[\int_0^t\Big(\int_0^s \hat{A}_u f\,du\Big)\,(\Phi_s,\,h_s-\hat{h}_s)ds\Big]\,. \end{aligned} \tag{4.17}$$

Consider the right hand side of (4.17). From Lemma 4.2 and the properties of stochastic integrals the first term is equal to

$$\begin{aligned} E\Big[\int_0^t (\Phi_s,\,\tilde{D}_s f)ds\Big] &= E\Big[\int_0^t (\Phi_s,\,E(\tilde{D}_s f\,|\,\mathcal{F}_s))ds\Big] \\ &= E\Big[\,Y_t\int_0^t (E^s(\tilde{D}_s f),\,d\nu_s)\Big]\,. \end{aligned} \tag{4.18}$$

The second term equals

$$(4.19)\qquad E\left[\int_0^t (\Phi_s, E^s[f(x_s)(h_s-\hat{h}_s)])ds\right] = E\left[Y_t\int_0^t (E^s[f(x_s)(h_s-\hat{h}_s)], d\nu_s)\right].$$

From (4.14), (4.16), (4.17), (4.18) and (4.19) we immediately obtain $E[\bar{M}_t Y_t]= E[M_t^* Y_t]$. The proof is complete.

5. A filtering problem connected with the observation process determined by a stochastic differential equation

In this section we shall discuss the filtering problem for the case when the observation process denoted by y_t is determined by the following form of a stochastic differential equation

$$(5.1)\qquad dy_t = a(t, x_s, s\leq t, y_s, s\leq t)dt+b(t, y_s, s\leq t)dw_t .$$

As we shall see below this type of the filtering problem is reduced to the former one with a suitable modification.

The sample paths x_t of the system process are assumed to take values in a separable, complete metric space S and to be right continuous with left hand limits. We donote by C the space of all continuous mappings from $[-T, 0]$ to R^N with the usual uniform topology, and by D the space of all right continuous mappings with left hand limits from $[-T, 0]$ to S, the topology of which is that of Skorokhod [18]. Let $a(t, g, f)$ be an N-vector valued functional in $[0, T]\times D\times C$ and $b(t, g, f)$, an $N\times M$-matrix valued functional in $[0, T]\times D\times C$ satisfying the following conditions.

(5. 2) $a(t, g, f)$ and $b(t, g, f)$ are Borel measurable in $[0, T]\times D\times C$

There exists a bounded measure Γ on $[-T, 0]$ and a positive contant K such that

$$(5.3)\qquad |a(t,g,f)-a(t,g,\tilde{f})|^2+|b(t,g,f)-b(t,g,\tilde{f})|^2 \leqq K\int_{-T}^0 |f(s)-\tilde{f}(s)|^2 d\Gamma(s)$$

$$(5.4)\qquad |a(t,g,f)|^2+|b(t,g,f)|^2\leqq K(1+\int_{-T}^0 |f(s)|^2 d\Gamma(s)+|L(t,g)|^2),$$

where $L(t, g)$ is a Borel measurable real valued functional in $[0, T]\times D$ such that

$$\int_0^T E|L(t, \pi_t x)|^2 dt<\infty .$$

Here $\pi_t x\in D$ is defined for (x_t) by

$$\pi_t x(u) = x_{t+u} \qquad \text{if } -t\leq u\leq 0$$
$$= x_0 \qquad \text{if } -T\leq u\leq -t .$$

(Note that $L(t, \pi_t x)$ is jointly (t, ω)-measurable. The norm of the $N\times M$-matrix $b=(b_{ij})$ is defined as $|b| = \sqrt{\sum_{i,j}^{N,M} b_{ij}^2})$.

Let w_t be an M-vector standard Wiener process satisfying (4.1). The dimention M is assumed to be larger than N. A stochastic process y_t with continuous paths is called a solution of the stochastic differential equation

$$y_t = \eta+\int_0^t a(s, \pi_s x, \pi_s y)ds+\int_0^t b(s, \pi_s x, \pi_s y)dw_s \tag{5. 5}$$

if $\sigma\{y_s;\ 0\leq s\leq t\}$ is independent of $\sigma\{w_v-w_u;\ t\leq u\leq v\leq T\}$ and satisfies the above formula. Here η is a constant.

Lemma 5.1. *Under conditions* (5.2) ~ (5.4), *there exists an unique solution* y_t *for* (5.5) *such that it is* $\sigma\{x_u, w_u;\ 0\leq u\leq t\}$*-measurable and square integrable.*

Since the lemma can be established by the standard method of successive approximations, we shall state the outline of the proof.
Set $y_0(t)=\eta$ and

$$y_t^n = \eta+\int_0^t a(s, \pi_s x, \pi_s y^{n-1})ds+\int_0^t b(s, \pi_s x, \pi_s y^{n-1})dw_s, \quad n\geqq 1 .$$

Then a direct calculation show that $\rho_n(t)\equiv \sup_{s\leq t} E|y_s^n-y_s^{n-1}|^2$ satisfies

$$\rho_{n+1}(t)\leq 2K(T+1)||\Gamma||\int_0^t \rho_n(s)ds \leq \cdots \leq (2K(T+1)||\Gamma||)^n \frac{\rho_1(t)}{n!}$$

making use of (5.3) and

$$\rho_1(t)\leq 2K\{(T+1)(1+\eta^2)||\Gamma||+\int_0^t E|L(s, \pi_s x)|^2 ds\} <+\infty$$

making use of (5.4). Here $||\Gamma||$ denotes the total mass of the measure Γ. Therefore, y_t^n converges to a continuous process y_t and it satisfies (5.5) by a standard argument. It is clear that y_t is $\sigma\{x_s, w_s; s\leq t\}$-measurable. The uniqueness can be proved similarly.

As before we shall denote the σ-fields $\sigma\{y_s;\ 0\leq s\leq t\}$ as $\mathcal{F}_t$ and $\mathcal{G}_t=\sigma\{x_s, w_s; 0\leq s\leq t\}$. Let us further introduce the following conditions.

(5. 6) $b(t, g, f)$ does not depend on $g\in D$.

(5. 7) The determinant of $N\times N$-matrix $c=bb^t$ is not zero for all t and f, where b^t is the transpose of b.

(5. 8) $\int_0^T E|h'(t, \pi_t x, \pi_t y)|^2\, dt<\infty$, where $h'(t, g, f)=c^{-1/2}(t, f)\, a(t, g, f)$.

Lemma 5.2. *Assume* (5.2)~(5.8). *Set*

$$w'_t = \int_0^t c^{-1/2}(s, \pi_s y)b(t, \pi_s y)dw_s\,, \tag{5.9}$$

$$z'_t = \int_0^t h'(s, \pi_s x, \pi_s y)ds + w'_t\,. \tag{5.10}$$

Then w'_t is an N-vector standard Wiener process adapted to $(\mathcal{G}_t)$ such that $\sigma\{w'_v - w'_u;\ s \le u \le v < T\}$ is independent of $\mathcal{G}_s$. Furthermore three processes (w'_t, h'_t, z'_t) satisfies conditions (2.1)~(2.4).

Proof. Since $w_t - w_s$ is independent of $\mathcal{G}_s$ by the assumption, w_t is a $\mathcal{G}_t$-martingale. Since $c_s^{-1/2}b_s$ is $\mathcal{G}_s$-measurable by Lemma 5.1, the process w'_t is also a $\mathcal{G}_t$-martingale. Now by the definition of the stochastic integral, we have

$$\begin{aligned} E[(w'^i_t - w'^i_s)(w'^j_t - w'^j_s) \mid \mathcal{G}_s] &= E\left[\int_s^t (c_u^{-1/2} b_u b_u^t c_u^{-1/2})_{ij} du \mid \mathcal{G}_s\right] \\ &= (t-s)\delta_{ij}\,. \end{aligned}$$

Hence we have obtained the first assertion. Since $\{h'_u: u \le s\}$ is $\mathcal{G}_s$-measurable, condition (2.4) is satisfied for (h'_t, w'_t). Condition (2.3) for the process h'_t is obvious from the assumption (5.8), thus proving the latter assertion of the lemma.

Set

$$\mathcal{F}'_t = \sigma\{z'_s;\ 0 \le s \le t\} \tag{5.11}$$

and define

$$\nu'_t = z'_t - \int_0^t \hat{h}'_s ds \qquad \text{where} \quad \hat{h}'_s = E[h'_s \mid \mathcal{F}'_s]\,. \tag{5.12}$$

Then Lemma 2.1 shows that ν'_t is a standard Wiener process adapted to $(\mathcal{F}'_t)$. We prove

Lemma 5.3. $\mathcal{F}_t = \mathcal{F}'_t$ *for all* $0 \le t \le T$.

Proof. Since y_t and z'_t are related by

$$y_t = \eta + \int_0^t c_s^{1/2} dz'_s \quad \text{or} \quad z'_t = \int_0^t c_s^{-1/2} dy_s \tag{5.13}$$

and $c_s^{-1/2}$ is $\mathcal{F}_s$-measurable, it is clear that $\mathcal{F}'_t \subset \mathcal{F}_t$. For the proof of the converse relation, we shall apply Lemma 3.1. Set $\beta_t = \nu'_t$ and $\varphi_t = -\hat{h}'_t$. The lemma states that $z'^n_t = \nu'_t + \int_0^{t \wedge T_n} \hat{h}'_s ds$ is an $(\mathcal{F}'_t, \tilde{P}_n)$-standard Wiener process. Then the solution y^n_t of the stochastic differential equation

$$y_t^n = \eta + \int_0^t c^{1/2}(s, \pi_s y^n) dz_s'^n$$

considered in $(\mathcal{F}_t', \tilde{P}_n)$ exists uniquely and $\sigma\{y_s^n; s \leq t\} \subset \sigma\{z_s'^n; s \leq t\}$ by Lemma 5.1, since $c^{1/2}(t, f)$ is Lipschitz continuous in the sense of (5.3) and (5.4) at least locally. On the other hand, since $z_t'^n = z_t'$ holds for $t < T_n$, $y_t = y_t^n$ holds for $t < T_n$ by the uniqueness of the stochastic equation. This implies $\mathcal{F}'_{t \wedge T_n} \supset \mathcal{F}_{t \wedge T_n}$. Since $T_n \uparrow T$ as $n \to \infty$, we get $\mathcal{F}_t' \supset \mathcal{F}_t$. The proof is complete.

We have thus reduced the filtering problem for the observation process y_t to that for the new observation process z_t'. Theorem 4.1 is then modified as follows.

Theorem 5.1. *Assume* (4.1) *and* (5.2)~(5.8). *If f belongs to $\tilde{\boldsymbol{D}}(\hat{A})$ and satisfies $\int_0^T E|f(x_t)h_t'|^2 dt < \infty$, then $E^t[f(x_t)] = E[f(x_t)|\mathcal{F}_t]$ satisfies the following stochastic differential equation*

$$E^t[f(x_t)] = E[f(x_0)] + \int_0^t E^s[\hat{A}_s f] ds + \int_0^t (c_s^{-1/2}[E^s(f(x_s)a_s) - E^s(f(x_s))E^s(a_s) + b_s E^s(\tilde{D}_s f)], dv_s') \tag{5.14}$$

where v_t' is the standard Wiener process determined by (5.12).

Proof. The only difference between (4.12) and (5.14) is in the term corresponding to $\tilde{D}_s f$. In (5.14), the N-vector $\tilde{D}_s f = (\tilde{D}_s^1 f, \cdots, \tilde{D}_s^N f)$ is defined as the Radon-Nikodym derivative of $\langle M(f), w \rangle_t$ with respect to t, which is related to that of $\langle M(f), w' \rangle_t$ in the following form (See [17, p 458] or [15, p 79])

$$\langle M(f), w \rangle_t = \int_0^t c_s^{-1/2} b_s d\langle M(f), w' \rangle_s .$$

Therefore $\tilde{D}_s f$ in (4.12) corresponds to $c_s^{-1/2} b_s \tilde{D}_s f$ in (5.14).

EXAMPLE. Let us consider the case when the system and observation processes are solutions of a stochastic differential equations of the type considered by Fleming and Nisio [3].

$$dx_t = A(t, \pi_t x, \pi_t y) dt + B(t, \pi_t x, \pi_t y) dw_t , \tag{5.15}$$

$$dy_t = a(t, \pi_t x, \pi_t y) dt + b(t, \pi_t y) dw_t . \tag{5.16}$$

Here w_t is an M-vector standard Wiener process, a and A are N and $(M\text{-}N)$-vector functionals $(M > N)$ respectively, and b and B are $N \times M$ and $(M\text{-}N) \times M$-matrix functionals respectively. We assume similar Lipschitz conditions as (5.2) and (5.3) for both of (a, b) and (A, B). For the initial random variables, we shall assume that $y_0 = 0$ and that

(5.17) x_0 is independent of $\sigma\{w_s, 0<s<T\}$.

Then the equation (5.15) and (5.16) has an unique solution (x_t, y_t) which is measurable with respect to $\sigma\{x_0, w_s; s<t\}$. If additional conditions (5.7) and (5.8) are imposed on a and b, Theorem 5.1 can be applied to this case. We shall obtain explicit representations of $\hat{A}_t$ and $\tilde{D}$. The results are that if f is a C^2-class function on R^{M-N}, it belongs to $\tilde{\boldsymbol{D}}(\hat{A})$ and

$$\hat{A}_t f(\omega) = \sum_{i=1}^{M-N} A_i(t, \pi_t x, \pi_t y) f'_{x_i}(x_t) + \frac{1}{2}\sum_{i,j=1}^{M-N} (B^t B)_{ij}(t, \pi_t x, \pi_t y) f''_{x_i x_j}(x_t) \tag{5.18}$$

and

$$\tilde{D}^t_t f(\omega) = \sum_{j=1}^{M-N} B_{ij}(t, \pi_t x, \pi_t y) f'_{x_j}(x_t) . \tag{5.19}$$

To prove this, apply Ito's formula to $f(x_t)$, we get

$$f(x_t) - f(x_0) = \sum_{i,j} \int B_{ij}(s, \pi_s x, \pi_s y) f'_{x_i}(x_s) dw^j_s + \int_0^t \bar{A}_s f(\omega) ds ,$$

where the integrand in the last term is the right hand side of (5.18). Since the first term on the right hand side is a $\mathcal{G}_t$-martingale, we see that

$$M_t(f) \equiv f(x_t) - f(x_0) - \int_0^t \bar{A}_s f ds$$

is a $\mathcal{G}_t$-martingale. This proves (5.18). The proof of (5.19) is immediate from

$$\begin{aligned}\langle M(f), w^i \rangle_t &= \sum_{k,j} \int B_{kj}(s, \pi_s x, \pi_s y) f'_{x_k}(x_s) d\langle w^j, w^i \rangle_s \\ &= \sum_k \int B_{ki}(s, \pi_s x, \pi_s y) f'_{x_k}(x_s) ds .\end{aligned}$$

Using the vector notation, $c_s^{-1/2} b_s \tilde{D}_s f = c_s^{-1/2} b B^t f'$, where B^t is the transpose of B and $f' = (f'_{x_1}, \cdots, f'_{x_N})$. Therefore, the term involving $\tilde{D}_s f$ disappears if and only if $bB^t \equiv 0$.

REMARK 6.1. There are many other ways of choosing $N \times N$-matrices $c^{1/2}$ and $c^{-1/2}$ in discussions of this section. In fact, in case where the dimmension M is equal to N, it is more natural to replace $c^{1/2}$ and $c^{-1/2}$ by b and b^{-1} respectively. More generally, if we choose an $N \times N$-matrix d with the Lipschitz condition (5.3) such that $c = dd^t$, then all discussions are valid replacing $c^{1/2}$ and $c^{-1/2}$ by d and d^{-1} respectively. It should be noted that the condition (5.8) does not depend on the choice of such d. Although the the innovation process (5.12) is changed by a such replacement, the the expression (5.14) does not depend on the replacement. In fact, the last member of (5.14) is equal to

$$\int_0^t (E^s[f(x_s)a_s]-E^s[f(x_s)]E^s[a_s]+b_sE^s[\tilde{D}_sf],\ c_s^{-1}dy_s-c_s^{-1}E^s[a_s]ds)\,. \tag{5.20}$$

6. Case of Markov Process

We now specialize the discussion of the previous section to the case where the pair (x_t, y_t) of the signal process x_t and the observation process y_t is Markov with repsect to $(\mathcal{G}_t, P)$. In this context the stochastic differential equation has a more definite meaning. We shall show in this section that $\hat{A}_t f(\omega)$ is replaced by $A_t f(x_t, y_t)$ where A_t is the generator of the process (x_t, y_t) and that with additional conditions $\tilde{D}^i f(t, \omega)$ is replaced by $D^i f(t, x_t, y_t)$ where $D^i f(t, x, y)$ is a measurable function in $[0, T]\times S\times R^N$. D^i may be regarded as a first order linear differential operator if $\eta_t=(x_t, y_t)$ is a diffusion Markov process.

Let us first investigate conditions for $\eta_t=(x_t, y_t)$ to be Markov.

Lemma 6.1. *Suppose that (x_t) is a Markov process which is completely independent of (w_t). Assume further that coefficients $a(t, g, f)$ and $b(t, g, f)$ of (5.2) depend only on the values $f(0)$ and $g(0)$ and that the measure Γ is concentrated at the point $\{0\}$. Then $((x_t, y_t), \mathcal{G}_t, P)$ is a Markov process.*

Since the proof is carried out by a standard argument of Markov processes, we shall state only the outline. Let us first notice that the latter condition of the lemma states that $a(t, \pi_t x, \pi_t y)=a(t, x_t, y_t)$ and $b(t, \pi_t y)=b(t, y_t)$. Then the equation (5.5) is written as

$$y_t = y_s+\int_s^t a(u, x_u, y_u)dy+\int_s^t b(u, y_u)dw_u\,. \tag{6. 1}$$

Now let $\{Q_\omega\}$ be the regular conditional distribution relative to $(\mathcal{G}_s, P)$, i.e., $Q_\omega(A)$, $A\in\mathcal{A}$ is $\mathcal{G}_s$-measurable for each A, a probability measure for each ω and that $Q_\omega(A)=P(A|\mathcal{G}_s)$ a.e. (See [20]). Since $\sigma\{x_s;\, s\leq T\}$ and $\sigma\{w_v-w_u;\, s\leq u<v\leq T\}$ are independent relative to Q_ω for a.s.ω, (y_t) $t>s$ may be considered as an observation process related to the signal process (x_t, Q_ω), $t>s$, by the formula (6.1). Then the uniqueness of the above stochastic differential equation and the Markov property of x_t proves that the joint distribution of (x_t, y_t), $t>s$ relative to the measure Q_ω depends only on the initial value (x_s, y_s) (together with the transition probability function of x_t and coefficients $a(s, x, y)$ and $b(s, y)$). This shows the Markov property of $((x_t, y_t), \mathcal{G}_t, P)$.

REMARK 6.1. The conditions of Lemma 6.1 are not always necessary for the Markov property of (x_t, y_t). For example, in the case of Fleming-Nisio, if we assume that all coefficients a, b, A, B depend only on t, the values $f(0)$ and $g(0)$, then the process (x_t, y_t) is Markov as is well known. However the process x_t is obviously not Markov.

Let $P(s, \eta;\, t, B)$ $(\eta=(x, y))$ be the transition probability function of (η_t)

assumed jointly measurable in (s, η, t) and let

$$P_s^t f(\eta) = \int P(s, \eta; t, d\eta') f(\eta') .$$

A family of linear operators A_t, $t \in [0, T]$ defined in the space of real valued measurable functions on $S \times R^N$ is called an extended generator if

$$P_s^t f(\eta) - f(\eta) = \int_s^t P_s^u A_u f(\eta) du \tag{6. 2}$$

is satisfied for all $0 < s < t < T$. We denote by $\boldsymbol{D}(A)$ the set of all f depending only on the first variable x and satisfying (6.1) together with

$$E|f(\eta_t)|^2 < +\infty \qquad \text{for each } t \text{ in } [0, T] , \tag{6. 3}$$

$$\int_0^T E|A_t f(\eta_t)|^2 dt < \infty . \tag{6. 4}$$

Set

$$M_t(f)(\omega) \equiv f(x_t) - f(x_0) - \int_0^t A_s f(\eta_s) ds . \tag{6. 5}$$

Lemma 6.2. *$M_t(f)$ defined by (6.5) is a square integrable $(\mathcal{G}_t, P)$-martingale.*

Proof. Since $f \in \boldsymbol{D}(A)$ we write $f(x_t)$ for $f(\eta_t)$. Note, however, that $A_u f(\eta)$ need not involve x. From this observation, the fact that $(\eta_t, \mathcal{G}_t, P)$ is Markov, and (6.2)~(6.4) we have the following chain of relations. For $0 < s < t$,

$$E[M_t(f) | \mathcal{G}_s] = M_s(f) + E\Big[f(x_t) - f(x_s) - \int_s^t A_u f(\eta_u) du \,|\, \mathcal{G}_s \Big] . \tag{6. 6}$$

The conditional expectation on the right equals

$$E[f(x_t) | \mathcal{G}_s] - f(x_s) - E\Big[\int_s^t A_u f(\eta_u) du \,|\, \mathcal{G}_s \Big] . \tag{6. 7}$$

But this is zero since

$$\begin{aligned} E[f(x_t) | \mathcal{G}_s] - f(x_s) &= P_s^t f(\eta_s) - f(\eta_s) \\ &= \int_s^t P_s^u A_u f(\eta_s) du \\ &= \int_s^t E[A_u f(\eta_u) | \mathcal{G}_s] du . \end{aligned} \tag{6. 8}$$

Hence the right hand side of (6.6) equals $M_s(f)$. The square integrability of $M_t(f)$ is obvious from (6.3) and (6.4). The lemma is proved.

The above lemma shows that $\boldsymbol{D}(A) \subset \boldsymbol{D}(\hat{A})$ and that $A_t f(\eta_t) = \hat{A}_t f(\omega)$, where $\widehat{A}_t$ is the operator defined is Section 4. In order to investigate the property of the operators D^i, it is necessary to quote rather delicate but deep results concerning additive functionals in the theory of Markov processes. The following terminology and results are from Motoo-Watanabe [17] and Meyer [15]. It is well known that the space time process (t, x_t, y_t) is stationary Markov. We assume

(6. 9) (t, x_t, y_t) is a Hunt process with Meyer's Hypothesis (L),

(6.10) $M_t(f)$ is $\sigma\{x_s, w_s;\ s\leq t\}$-measurable.

By the above two conditions, each i-th component w_t^i of w_t together with $M_t(f)$ is an additive functional of the process η_t. Then the process $\langle M(f), w^i\rangle_t$ introduced in Lemma 4.2 is again an additive functional that is absolutely continuous with respect to t for a.s.ω. Then there exists a jointly measurable function $D^i f(t, x, y)$ such that

$$\langle M(f), w^i\rangle_t = \int_0^t D^i f(s, x_s, y_s)ds \qquad \text{a.s.} \tag{6.11}$$

(Such $D^i f(s, x, y)$ is determined uniquely a.s. relative to a suitable measure called "canonical". See [17], Theorem 7.2.)

Theorem 5.1 yields the following result for the Markov process case.

Theorem 6.1. *Let* $\eta_t=(x_t, y_t)$ *where* y_t *is given by* (6.1) *be a Markov process. Assume condition* (4.1) *and let the coefficients a and b satisfy* (5.2), (5.3), (5.4), (5.7) *and* (8.8). *If f belongs to* $\boldsymbol{D}(A)$ *and* $\int_0^T E|f(x_t)h_t'|^2dt<\infty$, *then* $E^t f(x_t)$ *satisfies the stochastic differential equation* (5.14) *where* $\widehat{A}_t$ *and* $\widetilde{D}_t$ *are replaced by* A_t *the generator of* (x_t, y_t) *and by the operator* D_t *whose components* D_t^i *are defined by* (6.11).

REMARK 6.2. If, as in Lemma 6.1, (x_t) and (w_t) are completely independent it follows from (4.10) of Lemma 4.2 that the term involving D_t in the stochastic differential equation of Theorem 6.1 disappears leading us to the case treated in [10].

EXAMPLE (c.f. Liptzer-Shiryaev [13]). In case that coefficients a, b, A and B depend on $(t, f(0), g(0))$ in Fleming-Nisio's case, the operators A_t and D are given by

$$A_t f(x, y) = \sum_{i=1}^{M-N} A_i(t, x, y)f'_{x_i}(x)+\frac{1}{2}\sum_{i,j=1}^{M-N}(B^tB)_{ij}\,(t, x, y)f''_{x_ix_j}(x)$$

$$D^i f(t, x, y) = \sum_{j=1}^{M-N} B_{ji}(t, x, y)f'_{x_j}(x)\,,$$

where f is a C^2-class function in R^{M-N}.

REMARK 6.3. It is possible and sometimes convenient to regard the pair of processes (x_t) and (w_t) as a Markov process under more general settings by enlarging the state space. Let $x^t=(x_s, s\leq t)$ and $y^t=(y_s, s\leq t)$. Then we can prove that $((x^t, y^t), \mathcal{G}_t, P)$ is a Markov process by a similar argument as in Lemma 6.1, making use of the uniqueness of the solution (5.5). Also, the case that $((x_t, y^t), \mathcal{G}_t, P)$ is Markov is discussed in the problems of stochastic control based on a partially observable process. Such a case occurs if the coefficients a, b, A and B of Fleming-Nisio depend on t, $f(0)$ and g. It will be obvious that the discussions of this section can be applied to these cases.

KOBE UNIVERSITY OF COMMERCE
UNIVERSITY OF MINNESOTA
NAGOYA UNIVERSITY

References

[1] J.M.C. Clark: *The representation of functionals of Brownian motion by the stochastic integrals*, Ann. Math. Statist. **41** (1970), 1282–1295.
[2] J.L. Doob: Stochastic Processes, Wiley, New York, 1952.
[3] W.H. Fleming and M. Nisio: *On the existence of optimal stochastic controls*, J. Math. Mech. **15** (1966), 777–794.
[4] I.I. Gikhman and A.V. Skorokhod: Introduction to the Theory of Random Processes, W.B. Saunders Co., Philadelphia, 1969.
[5] I.V. Girsanov: *On transforming a certain class of stochastic processes by absolutely continuous substitution of measures*, Theor. Probability Appl. **5** (1960), 285–301.
[6] K. Ito and M. Nisio: *On stationary solution of a stochastic differential equation*, J. Math. Kyoto Univ. **4** (1964), 1–75.
[7] T. Kailath: *An innovations approach to least squares estimation, Part I: Linear filtering with additive white noise*, IEEE Trans. Automatic Control, AC–**13**, Number 6 (1968), 646–655.
[8] T. Kailath: *A further note on a general likelihood formula for random signals in Gaussian noise*, IEEE Trans. Information Theory, IT–**16**, (1970).
[9] G. Kallianpur and C. Striebel: *Estimation of stochastic processes: Arbitrary system process with additive white noise observation errors*, Ann. Math. Statist. **39** (1968), 785–801.
[10] G. Kallianpur and C. Striebel: *Stochastic differential equations occurring in the estimation of continuous parameter stochastic processes*, Theor. Probability Appl. **4** (1969), 597–622.
[11] G. Kallianpur and C. Striebel: *Stochastic differential equations in stochastic estimation problems*, Multivariate Analysis II, Academic Press, 1969.
[12] H. Kunita and S. Watanabe: *On square integrable martingales*, Nagoya Math. J. **30** (1967), 209–245.

[13] R.S. Liptzer and A.N. Shiryaev: *Non linear filtering of Markov diffusion processes,* Trudy Mat. Inst. Steklov **104** (1968), 135–180.

[14] H.P. Mckean. Jr.: Stochastic Integrals, Academic Press, 1969.

[15] P.A. Meyer: Probability and Potentials, Blaisdell Pub. Co., 1966.

[16] P.A. Meyer: *Integrales stochastiques,* Sémianire de Probabilités I, Lecture Notes in Math. 39, Springer, 1967.

[17] M. Motoo and S. Watanabe: *On a class of additive functionals of Markov processes,* J. Math. Kyoto Univ. **4** (1965), 429–469.

[18] A.N. Shiryaev: *Stochastic equations of non linear filtering of Markovian jump processes,* Problemy Peredaci Informacii **3** (1966), 3–22.

[19] A.V. Skorokhod: *Limit theorems for stochastic processes,* Theor. Probability Appl. **1** (1956), 261–290.

[20] D. Stroock and R. Varadhan: *Diffusion processes with continuous coefficients,* I. II, Comm. Pure. Appl. Math. **22** (1969), 345–400, 479–530.