JOURNAL OF DIFFERENTIAL EQUATIONS 3, 179-190 (1967)

Dynamical Equations for Optimal Nonlinear Filtering

H. J. KUSHNER*

Center Dynamical Systems, Division of Applied Mathematics, Brown University, Providence, Rhode Island

1. INTRODUCTION

In this paper, we prove a result in optimal nonlinear filtering (and representation of a conditional expectation as a solution to a stochastic differential equation) which we derived formally in [*1*] and [*2*]. Some possible computational methods are briefly discussed.

Write the vector stochastic differential (Itô) equations

$$dx = f(x, t)\, dt + V^{1/2}(x, t)\, dz, \tag{1}$$

$$y_t = \int_0^t g(x_s, s) ds + \int_0^t \Sigma_s^{1/2}\, d\hat{w}_s , \tag{2}$$

where z_s and $\hat{w}_s$ are independent vector Wiener processes. The matrices $V^{1/2}$ and $\Sigma^{1/2}$ are square roots of the nonnegative-definite V and positive-definite Σ, respectively. Let $dw = \Sigma^{1/2} d\hat{w}$, and suppose that z_s is independent[1] of w_s. $E^{\mathcal{A}}$ is the expectation conditioned on a σ-field $\mathcal{A}$, and $\mathcal{B}(\cdot)$ is the completion of the minimal σ-field over which the random variables the parenthesis are measurable.

The function y_t represents observations on the process x_s. (From one "engineering" point of view, the observation is $\dot{y} = g(x, t) + \xi$, where ξ is "white Gaussian" noise.) Write $\mathcal{F}_t = \mathcal{B}(y_s, s \leqslant t)$, and suppose that there is a "conditional probability *density*" $P^{\mathcal{F}_t}(x, t)$ of x_t conditioned on $\mathcal{F}_t$. Then, the *formal* results in [*2*] are that the conditional "density" and expectation have the representations

$$\begin{aligned} dP^{\mathcal{F}_t}(x, t) = P^{\mathcal{F}_t}(x, t)(dy_t - E^{\mathcal{F}_t} g(x_t, t)dt)' \Sigma_t^{-1}(g(x, t) \\ - E^{\mathcal{F}_t} g(x_t, t)) + L^* P^{\mathcal{F}_t}(x, t) dt \end{aligned} \tag{3}$$

* Part of this research was done while the author was a consultant to the RAND Corp., Santa Monica, California, and part was supported by the United States Air Force through the Air Force Office of Scientific Research under Grant No. AF-AFOSR-693-64.

[1] A derivation may also be carried out without this assumption.

$$d(E^{\mathscr{F}_t}h(x_t)) = (dy_t - E^{\mathscr{F}_t}g(x_t, t)dt)'\Sigma_t^{-1}(E^{\mathscr{F}_t}h(x_t)g(x_t, t) - E^{\mathscr{F}_t}g(x_t, t)E^{\mathscr{F}_t}h(x_t)) + E^{\mathscr{F}_t}Lh(x_t)dt \tag{4}$$

L^* is the formal adjoint of the differential generator

$$L = \sum_i f_i(x, t) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j} V_{ij}(x, t) \frac{\partial^2}{\partial x_i \partial x_j}$$

If $\Sigma_s^{-1} = 0$, then the observations are valueless, and (3) reduces to the Fokker–Planck equation. The right sides of (3) and (4) are linear in the (incremental) observation dy_t.

In this paper, we prove (4) under explicit conditions [(A1)–(A11) below] on the x_t process. Note also that the proof is valid if we suppose that the f and V of (1) are general nonanticipative functions, and (A1)–(A11) holds, and (A1) is uniform in ω. (3) has not yet been proved without the implicit assumption that $P^{\mathscr{F}_t}(x_t, t)$ is sufficiently differentiable and has suitable properties for large $\| x \|$, which we have not been able to very from conditions on the x_t and y_t processes. If $f(x, t) = 0$ and $V(x, t) = 0$, then $L^* = 0$ and (3) may be proved.

From a Baysian point of view, (4) is a complete description of the optimum filter. It gives, *in principle*, a recursive method for computing the conditional moments of x_t, given the observations y_s, $s \leqslant t$; i.e., a particular sample path of a version of $E^{\mathscr{F}_t}h(x_t)$ may be obtained as a function of time, as the observations become available. (The estimate is the output of a dynamical system whose input is the "observation.") As such, it would be expected to be significant from the point of view of the practical problems of filtering. Recursive methods for computing $E^{\mathscr{F}_t}x_t$ (a linear differential equation whose forcing term is linear in the observation) for the case of linear f, g, and V independent of x are available [9] and widely used. The practical usefulness of the result depends on the constructability of physical apparatus which provide useful approximations to the system described by (4). The numerical work will be reported on in detail later. Some of the ideas which we feel to be novel and worthwhile are discussed here. For some of the nonlinear systems [(1) and (2)] studied, our methods yield consistently better results than currently existing methods based on "linearization," and the use of methods for the linear problem.

In a recent note, Bucy [3] put the problem in [2] in a form in which a formal application of Itô's Lemma yields (3). This work, still formal, is less intuitive, but more satisfying mathematically than our approach in [2]. Some relevant results for Poisson processes are given by Wonham [4]. Although the work in [1] and [2] was independent, Stratonovich [5] had, from a formal point of view, considered the same problem earlier. These results are not consistent with the Itô interpretation of the stochastic integral. However, Stratonovich has recently described a stochastic calculus [8] (somewhat different from Itô's), with respect to which his earlier results must be interpreted. Then the formal continuous time results in [2] and [5] appear to be equivalent, at least in the scalar case with independent z_s and w_s processes.

2. Assumptions

Functions of t only are written with the argument as a subscript. Otherwise we use whatever form appears most convenient.

(A1) The components of $f(\cdot, \cdot)$ and $V^{1/2}(\cdot, \cdot)$ are Baire functions and satisfy a uniform Lipschitz condition in the variable x, and are bounded, in absolute value, by $K(1 + x'x)^{1/2}$ for some real positive number K. $E\| x_0 \|^2 < \infty$, and x_0 is independent of z_s, $0 \leqslant s \leqslant T$.

(A2) The components of $g(\cdot, \cdot)$ and the scalar valued $h(\cdot)$ are Baire functions of all their arguments for $0 \leqslant t \leqslant T$, $\| x \| < \infty$. $h(\cdot)$ has continuous second partial derivatives at each finite x.

(A3) Σ_t is positive-definite and continuous at each t in the finite interval $[0, T]$. Σ_s does not depend on x (see remark at the end of the proof).

(A4) $$Eq(x_s) \exp[(1 + b) \int_0^T g'(x_s, s)\Sigma_s^{-1} g(x_s, s) ds] < \infty, \; Eg'(x_t, t)g(x_t, t) < \infty$$

in $[0, T]$, where $q(x_s)$ is either 1 or $| h(x_s)|^{1+b}$, for some $b > 0$.

(A5) $E| h(x_t)| < \infty$, $t \leqslant T$.

(A6) $E| Lh(x_t)| < \infty$, $E\| g(x_t, t)\, h(x_t)\| < \infty$, $t \leqslant T$.

(A7) The z_t process is independent of the w_t process.[2]

(A8) $$\int_0^T Eq_t^2 g'(x_t, t)\Sigma_t^{-1} g(x_t, t) \exp\left[3 \int_0^t g'(x_s, s)\Sigma_s^{-1} g(x_s, s) ds\right] dt < \infty,$$

where $q_t = h(x_t)$ or 1.

(A9) $$\int_0^T E\,| Lh(x_t) |\, \exp\left[\int_0^t g'(x_s, s)\Sigma_s^{-1} g(x_s, s) ds\right] dt < \infty.$$

[2] Introduced only to allow a simpler proof.

(A10) $$\int_0^T E \exp\left[3 \int_0^t g'(x_s, s)\Sigma_s^{-1} g(x_s, s)ds\right]$$
$$\times (\mathbf{grad}\, h(x_t))' V(x_t, t)(\mathbf{grad}\, h(x_t))dt < \infty.$$

(A11) $$\int_0^T Eq_t \exp\left[\int_0^t g'(x_s, s)\Sigma_s^{-1} g(x_s, s)ds\right] g'(x_t, t)\Sigma_s^{-1} g(x_t, t)dt < \infty,$$

where q_t is either $| h(x_t) |$ or 1.

3. Proof of Eq. (4)

THEOREM. *Assume* (A1) *to* (A11). *Then, a version of* $E^{\mathscr{F}_t} h(x_t)$ *satisfies the stochastic differential equation* (4).

Proof. (1) Under (A1), the process x_t, $t \leqslant T$, is defined and continuous with probability one (w.p.l). Fix $t \leqslant T$ until mentioned otherwise. For each positive integer k, define the partition of $[0, t]$:

$$0 = t_{k0} \leqslant t_{k1} \leqslant \cdots \leqslant t_{k(n_k+1)} = t;$$

$$I_{ki} = \{t: t_{k,i+1} > t \geqslant t_{ki}\};$$

$$\delta y_{ki} = \int_{I_{ki}} dy_s = y_{t_{k,i+1}} - y_{t_{ki}};$$

$$G_{ki} = \int_{I_{ki}} g(x_s, s)ds, \ \delta w_{ki} = \int_{I_{ki}} dw_s = \int_{I_{ki}} \Sigma_s^{1/2} dw_s.$$

Then $\delta y_{ki} = G_{ki} + \delta w_{ki}$. Write $\mathscr{F}_k = \mathscr{B}(\delta y_{k0}, ..., \delta y_{kn_k})$, $\mathscr{F}_t = \mathscr{B}(y_s, s \leqslant t)$, and $\mathscr{G}_k = \mathscr{B}(G_{k0}, ..., G_{kn_k})$.

The δy_{ki}, $i \leqslant n_k$, are conditionally independent (with respect to $\mathscr{G}^k$) normally distributed random variables with mean G_{ki} and finite variance

$$S_{ki} = \int_{I_{ki}} \Sigma_s \, ds,$$

$$P^{\mathscr{G}_k}(\delta y^k \in A) = C \int_A N(G^k, S^k, a)\, da,$$

where C is a normalizing constant and $a = (a_1, ..., a_{n_k})$, $G^k = \{G_{ki}, i \leqslant n_k\}$, $\delta y^k = \{\delta y_i, i \leqslant n_k\}$, and

$$N(G^k, S^k, a) = \exp -\frac{1}{2} \sum_{i=1}^{n_k} (a_i - G_{ki})' S_{ki}^{-1} (a_i - G_{ki}).$$

Let the Wiener processes $\tilde{z}_s$ and z_s be independent. Let $\tilde{x}_s$, $s \leqslant T$, correspond to $\tilde{z}_s$ via (1). Then the processes $\tilde{x}_s$ and x_s are independent, but have the same distribution. Define $\tilde{G}_{ki} = \int_{I_{ki}} g(\tilde{x}_s, s)\, ds$ and $\tilde{G}^k = \{\tilde{G}_{ki}, i \leqslant n_k\}$.

Let $P(dG^k)$ and $P(dG^k \times dx_t)$ be the measures on the Euclidean range spaces of G^k and the pair (G^k, x_t), respectively. Considered as an ω function (since δy^k is an ω function), H_t^k is obviously a version of $E^{\mathscr{F}_k} h(x_t)$.

$$H_t^k = \frac{\iint h(x_t) N(G^k, S^k, \delta y^k) P(dG^k \times dx_t)}{\int N(G^k, S^k, \delta y^k) P(dG^k)}. \tag{6}$$

Since δy^k is held fixed in the integrations in (6), we may change notation to a more convenient form by substituting $\tilde{x}_t$ and $\tilde{G}^k$ for x_t and G^k, respectively. Then, w.p.l. (recall that $\tilde{x}_t$ has the same law as x_t, but is independent of x_s, $s \leqslant T$).

$$H_t^k = \frac{E^{\mathscr{F}_k} h(\tilde{x}_t) N(\tilde{G}^k, S^k, \delta y^k)}{E^{\mathscr{F}_k} N(\tilde{G}^k, S^k, \delta y^k)} = E^{\mathscr{F}_k} h(x_t). \tag{7}$$

We now multiply both terms of (7) by the $\mathscr{F}_k$-measurable function $\exp \frac{1}{2} \sum_0^{n_k} \delta y_{ki}' S_{ki}^{-1} \delta y_{ki} \geqslant 1$ [which is finite w.p.l.)] yielding

$$\begin{aligned} H_t^k &= E^{\mathscr{F}_k} h(x_t) = E^{\mathscr{F}_k} h(\tilde{x}_t) \exp R_k / E^{\mathscr{F}_k} \exp R_k, \\ R_k &= \sum_0^{n_k} (\delta y_{ki}' S_{ki}^{-1} \tilde{G}_{ki} - \tfrac{1}{2} \tilde{G}_{ki}' S_{ki}^{-1} \tilde{G}_{ki}). \end{aligned} \tag{8}$$

Define

$$\begin{aligned} R_t &= \int_0^t dy_s' \Sigma_s^{-1} g(\tilde{x}_s, s) - \frac{1}{2} \int_0^t g'(\tilde{x}_s, s) \Sigma_s^{-1} g(\tilde{x}_s, s)\, ds \\ &= \int_0^t dw_s' \Sigma_s^{-1/2} g(\tilde{x}_s, s) + \int_0^t g'(x_s, s) \Sigma_s^{-1} g(\tilde{x}_s, s)\, ds \\ &\quad - \frac{1}{2} \int_0^t g'(\tilde{x}_s, s) \Sigma_s^{-1} g(\tilde{x}_s, s)\, ds, \\ H_t &= E^{\mathscr{F}_t} h(\tilde{x}_t) \exp R_t / E^{\mathscr{F}_t} \exp R_t. \end{aligned} \tag{9}$$

(2) As k increases, let $\mathscr{F}_k \subset \mathscr{F}_{k+1}$ and $\max_i$ (length of I_{ki}) $\to 0$. Owing to the w.p.l. continuity of y_s on $[0, T]$, $\mathscr{F}_k \uparrow \mathscr{F}_t = \cup \mathscr{F}_k$.[3] Next we prove that $H_t^k \to H_t$ (w.p.l.), and H_t is a version of $E^{\mathscr{F}_t} h(x_t)$; t is still fixed.

The sequence of conditional expectations H_t^k is a martingale, and $E|H_t^k| = E|h(x_t)| < \infty$. By the martingale convergence theorem there is

[3] $\cup \mathscr{F}_k$ is the completion of the minimal σ-field containing $\mathscr{F}_1, \ldots$.

an $\mathscr{F}_t$-measurable random variable η, with $E|\eta| = E|h(x_t)|$, such that $H_t^k \to \eta$ w.p.l. as $k \to \infty$ and $H_t^1, H_t^2, \ldots, \eta$ is a martingale. In fact, we also have $E^{\mathscr{F}_t} h(x_t) = \eta$.

(3) By (A3) and (A4), and for small $\delta = t_{ki+1} - t_{ki}$, we write

$$\left(\int_{I_{ki}} \Sigma_s ds\right)^{-1} = \frac{1}{\delta} (\Sigma_s^{-1/2} + \epsilon_1(\delta, s))'(\Sigma_s^{-1/2} + \epsilon_1(\delta, s)) \quad \text{where} \quad \epsilon_1(\delta, s)$$

is uniformly small in s and in i. Then

$$\int_{I_{ki}} g'(\bar{x}_s, s)\, ds \left(\int_{I_{ki}} \Sigma_s\, ds\right)^{-1} \int_{I_{ki}} g(\bar{x}_s, s)\, ds$$

$$= \frac{1}{\delta} \int_{I_{ki}} g'(\bar{x}_s, s)(\Sigma_s^{-1/2} + \epsilon_1(\delta, s))'\, ds \int_{I_{ki}} (\Sigma_s^{-1/2} + \epsilon_1(\delta, s)) g(\bar{x}_s, s)\, ds \equiv M_{ki}.$$

(A4) and Fubini's theorem imply that, as a function of $s, g'(\bar{x}_s, s)\, \Sigma_s^{-1} g(\bar{x}_s, s)$ is integrable (w.p.l.) on $[0, T]$. The Schwarz inequality yields

$$\sum_i M_{ki} \leqslant \int_0^t g'(\bar{x}_s, s)\Sigma_s^{-1} g(\bar{x}_s, s)\, ds(1 + \epsilon(\delta)).$$

where $\epsilon(\delta) \to 0$ as $\delta \to 0$. The sequence of functions M_k with values M_{ki}/δ in I_{ki} tends to $g'(\bar{x}_s, s)\, \Sigma_s^{-1} g(\bar{x}_s, s)$ almost everywhere on $[0, T]$ (w.p.l.). Then, an application of Fatou's lemma yields.

$$\liminf \sum_i M_{ki} = \liminf \int_0^t M_k\, ds \geqslant \int_0^t g'(\bar{x}_s, s)\Sigma_s^{-1} g(\bar{x}_s, s)\, ds,$$

which implies equality in the limit of (10) (w.p.l.). Similarly it may be shown that the other sums in R_k converge w.p.l. to the corresponding integrals in R_t.

For small b, by definition of R_k,

$$E\,|\,h(\bar{x}_t) \exp R_k\,|^{1+b} = E\Big\{|\,h(x_t)\,|^{1+b} \exp\,(1 + b) \sum_i \delta w_{ki}' S_{ki}^{-1} \bar{G}_{ki}$$

$$\cdot \exp\,(1 + b)\Big(\sum_i G_{ki}' S_{ki}^{-1} \bar{G}_{ki} - \frac{1}{2} \sum_i \bar{G}_{ki}' S_{ki}^{-1} \bar{G}_{ki}\Big)\Big\}$$

$$\leqslant E\,\Big\{|\,h(\bar{x}_t)\,|^{1+b} \exp\left(\frac{(1 + b)^2}{2} - \frac{(1 + b)}{2} + \frac{(1 + b)}{2}\right) \sum_i \bar{G}_{ki}' S_{ki}^{-1} \bar{G}_{ki}$$

$$\cdot \exp \frac{(1 + b)}{2} \sum_i G_{ki}' S_{ki}^{-1} G_{ki}\Big\}\,. \tag{12}$$

The last step in (12) makes use of the facts that the expectation of $\exp(1 + b) \sum_i \delta w_{ki} S_{ki}^{-1} \bar{G}_{ki}$, given the $\bar{G}_{ki}$, is $\exp \frac{1}{2}(1 + b)^2 \sum_i \bar{G}_{ki}' S_{ki}^{-1} \bar{G}_{ki}$, and also of the inequality $2\bar{G}'_{ki}S_{ki}^{-1}\bar{G}_{ki} \leqslant G'_{ki}S_{ki}^{-1}G_{ki} + \bar{G}'_{ki}S_{ki}^{-1}\bar{G}_{ki}$. By (A4) and (10), the integrand on the right side of (12) is uniformly bounded by an integrable function, and we may conclude that $h(\bar{x}_t) \exp R_k$ are in L_r for some $r > 1$, and are uniformly integrable. Since, in addition, $\exp R_k \to \exp R_t$ w.p.l., $h(\bar{x}_t) \exp R_k \to h(\bar{x}_t) \exp R_t$ in L_r, $r > 1$. Thus, since $\mathscr{F}_k \uparrow \mathscr{F}_t$, $E^{\mathscr{F}_k} h(\bar{x}_t) \exp R_k \to E^{\mathscr{F}_t} h(\bar{x}_t) \exp R_t$ in probability (Loeve [12], p. 409, para. 10a). Similarly $E^{\mathscr{F}_k} \exp R_k \to E^{\mathscr{F}_t} \exp R_t$ in probability Thus $H_t^k \to H_t$ in probability. Since limits in probability and w.p.l. limits are the same (w.p.l.), $H_t^k \to H_t$ w.p.l.

(4) Now, we show that (9) satisfies (4) for each t (w.p.l.) We use the martingale definition of the stochastic integral.

The maximum values (in $[0, T]$) of the ordinary integrals in R_t are finite w.p.l. Since $\int_0^T Eg'(\bar{x}_s, s)\, \Sigma_s^{-1} g(\bar{x}_s, s)\, ds < \infty$, the stochastic integral $\int_0^t g(\bar{x}_s, s)\, \Sigma_s^{-1/2}\, d\hat{w}_s$ is continuous in $[0, T]$ w.p.l., and, hence, is bounded there w.p.l. Thus $\infty > R_t > -\infty$ and $\infty > \exp R_t > 0$ for all t in $[0, T]$ w.p.l.

Now, since the function $\exp(u)$ is twice continuously differentiable at each $u \in (-\infty, \infty)$, and $-\infty < R_t < \infty$ w.p.l., Itô's lemma (see [7], Theorem 7.2) implies that $\exp R_t$ is a stochastic integral with

$$\begin{aligned}
\exp R_t &= \exp R_0 + \int_0^t \frac{\partial(\exp R_s)}{\partial R_s}\, dR_s \\
&\quad + \frac{1}{2}\int_0^t \frac{\partial^2(\exp R_s)}{\partial R_s^2}\, g'(\bar{x}_s, s)\Sigma_s^{-1} g(\bar{x}_s, s)\, ds, \\
&= \exp R_0 + \int_0^t (\exp R_s) g'(\bar{x}_s, s)\Sigma_s^{-1} dy_s, \\
& \qquad\qquad dy_s = \Sigma_s^{1/2} d\hat{w}_s + g(x_s, s)\, ds .
\end{aligned} \tag{13}$$

Since $h(x)$ has continuous second derivatives and, by (A1),

$$\max_{t \leqslant T} \| \bar{x}_t \|^2 < \infty$$

w.p.l., $h(\bar{x}_t)$ is also a stochastic integral,

$$h(\bar{x}_t) - h(\bar{x}_0) = \int_0^t Lh(\bar{x}_s)\, ds + \int_0^t (\operatorname{grad} h(\bar{x}_s))' V_s^{1/2} d\bar{z}_s .$$

$(\exp R_t)\, h(\bar{x}_t)$ is also a stochastic integral. Using the independence of w_t and $\bar{z}_t$,

$$\begin{aligned}
h(\bar{x}_t) \exp R_t - h(\bar{x}_0) \exp R_0 &= \int_0^t (\exp R_s) h(\bar{x}_s) g'(\bar{x}_s, s)[\Sigma_s^{-1}\, dy_s] \\
&+ \int_0^t (Lh(\bar{x}_s)) \exp R_s\, ds + \int_0^t (\exp R_s)(\mathbf{grad}\, h(\bar{x}_s))'\, V_s^{1/2}\, d\bar{z}_s .
\end{aligned} \tag{14}$$

By (A10) and the independence of $\tilde{z}_s$ and y_s, $s \leqslant T$, the expectation with respect to $\mathscr{F}_t$, of the last term in (14) is zero w.p.l. The term will be omitted hence forth.

It will be proved (5) that for all integrals in (13) and (14) we have

$$E^{\mathscr{F}t} \int_0^t k_s \, dy_s = \int_0^t [E^{\mathscr{F}s} k_s] \, dy_s \quad \text{and} \quad E^{\mathscr{F}t} \int_0^t k_s \, ds = \int_0^t [E^{\mathscr{F}s} k_s] \, ds \text{ w.p.l.}$$

and that the maximum values (in $[0, T]$)of the conditional expectation of each term in (14) with respect to $\mathscr{F}_t$ is finite w.p.l. We also have $E^{\mathscr{F}t} \exp R_t > 0$ w.p.l. With this interchange of the order of the ordinary and stochastic integration with the conditional expectation, H_t is the ratio of stochastic integrals; $H_t = A_t/B_t = E^{\mathscr{F}t} h(\tilde{x}_t) \exp R_t / E^{\mathscr{F}t} \exp R_t$, where

$$\begin{aligned} dA_s &= [E^{\mathscr{F}s}(\exp R_s) Lh(\tilde{x}_s)] \, ds + [E^{\mathscr{F}s}(\exp R_s) h(\tilde{x}_s) g'(x_s, s)] \Sigma_s^{-1} \, dy_s \\ dB_s &= [E^{\mathscr{F}s}(\exp R_s) g'(\tilde{x}_s, s)] \Sigma_s^{-1} \, dy_s . \end{aligned} \tag{15}$$

Applying Itô's Lemma to the ratio $A_t/B_t = H_t$ yields

$$\begin{aligned} H_t = H_0 + \int_0^t \Big(\frac{\partial H_s}{\partial A_s} \, dA_s &+ \frac{\partial H_s}{\partial B_s} \, dB_s + \frac{\partial^2 H_s}{\partial A_s \partial B_s} (dA_s)'(dB_s) \\ &+ \frac{1}{2} \frac{\partial^2 H_s}{\partial B_s^2} (dB_s)'(dB_s) \Big] . \end{aligned}$$

Thus,

$$\begin{aligned} dH_t = {} & \{[E^{\mathscr{F}t}(\exp R_t) h(\tilde{x}_t) g'(\tilde{x}_t, t)] \Sigma_t^{-1} dy_t + E^{\mathscr{F}t}[(\exp R_t) Lh(\tilde{x}_t)] dt\} / B_t \\ & - [E^{\mathscr{F}t} h(\tilde{x}_t) \exp R_t][E^{\mathscr{F}t}(\exp R_t) g'(\tilde{x}_t, t) \Sigma_t^{-1} dy_t] / B_t^2 \\ & - [E^{\mathscr{F}t}(\exp R_t) h(\tilde{x}_t) g'(\tilde{x}_t, t) \Sigma_t^{-1}] \cdot [E^{\mathscr{F}t} g(\tilde{x}_t, t) \exp R_t] \, dt / B_t^2 \\ & + [E^{\mathscr{F}t} h(\tilde{x}_t) \exp R_t][E^{\mathscr{F}t}(\exp R_t) g'(\tilde{x}_t, t)] \, \Sigma_t^{-1} [E^{\mathscr{F}t} g(\tilde{x}_t, t) \exp R_t] dt / B_t^3 . \end{aligned}$$

Since $E^{\mathscr{F}t} k(\tilde{x}_t) \exp R_t / B_t = E^{\mathscr{F}t} k(x_t)$ whenever $E |k(x_t)| < \infty$, we obtain finally

$$\begin{aligned} dE^{\mathscr{F}t} h(x_t) = {} & (dy_t - E^{\mathscr{F}t} g(x_t, t) dt)' \Sigma_t^{-1} (E^{\mathscr{F}t} h(x_t) g(x_t, t) - E^{\mathscr{F}t} h(x_t) E^{\mathscr{F}t} g(x_t, t)) \\ & + E^{\mathscr{F}t} Lh(x_t) \, dt, \end{aligned}$$

which is (4).

It only remains to prove the statement in the second paragraph below (14).

(5) Let D_s be a vector-valued measurable (s, ω) function which is Lebesgue-measurable for almost all fixed ω. Let D_s be independent of $\hat{w}_t - \hat{w}_s$, all $t > s$. Let $\hat{w}_s$, $s \leqslant t$ be measurable over the σ-field $\mathscr{A}_t$, and let

$$E^{\mathscr{A}t} D_s = E^{\mathscr{A}s} D_s \tag{16}$$

w.p.l., $t > s$. First we show that if

$$\int_0^T ED_s'D_s\,ds < \infty, \tag{17}$$

then, w.p.l., for each t,

$$E^{\mathcal{O}_t}\int_0^t D_s'\,d\hat{w}_s - \int_0^t (E^{\mathcal{O}_s}D_s')\,d\hat{w}_s = 0. \tag{18}$$

Under (16), (18) is obviously true if D_s is a step function with fixed points of discontinuity $s = t_1, \ldots$. By (17) we may approximate D_s by a sequence of nonanticipative right continuous step functions D_s^n satisfying (see Doob [6], IX, pp. 440–441)

$$\int_0^T E(D_s - D_s^n)'(D_s - D_s^n)\,ds < 2^{-n}. \tag{19}$$

Finally, it is straightforward to prove that, w.p.l.

$$E^{\mathcal{O}_t}\int_0^t D_s'd\hat{w}_s = \lim_n E^{\mathcal{O}_t}\int_0^t (D_s^n)'d\hat{w}_s = \lim_n \int_0^t [E^{\mathcal{O}_s}D_s^n]'d\hat{w}_s$$
$$= \int_0^t [E^{\mathcal{O}_s}D_s']d\hat{w}_s\,.$$

By a similar argument, if

$$\int_0^T E\,\|\,D_s\,\|\,ds < \infty, \tag{20}$$

then, w.p.l.

$$\int_0^t E^{\mathcal{O}_s}D_s ds = E^{\mathcal{O}_t}\int_0^t D_s ds\,. \tag{21}$$

Let $\mathcal{O}_t = \mathscr{B}(y_s, \hat{w}_s, s \leqslant t)$. Note that $\bar{x}_s, s \leqslant T$, is independent of all random variables which are measurable over $\mathcal{O}_t$. The integrands of all the integrals of (22) satisfy (16).

$$\begin{aligned}
&\text{(a)} && \int_0^t (\exp R_s)g'(\bar{x}_s, s)\,\Sigma_s^{-1/2}d\hat{w}_s\,,\\
&\text{(b)} && \int_0^t (\exp R_s)h(\bar{x}_s)g'(\bar{x}_s, s)\Sigma_s^{-1/2}\,d\hat{w}_s\,,\\
&\text{(c)} && \int_0^t (\exp R_s)g'(\bar{x}_s, s)\Sigma_s^{-1}g(x_s, s)\,ds,\\
&\text{(d)} && \int_0^t (\exp R_s)h(\bar{x}_s)g'(x_s, s)\Sigma_s^{-1}g(x_s, s)\,ds,\\
&\text{(e)} && \int_0^t (\exp R_s)Lh(\bar{x}_s)\,ds.
\end{aligned} \tag{22}$$

Under (A8) the integrands in (22a) and (22b) satisfy (17). Under (A9) and (A11) the integrands (22c), (22d), and (22e) satisfy (20). Hence, for these integrands the appropriate result, either (18) or (21), is true. The fact that (22a) and (22b) are martingales together with (A8)–(A11) imply that the maximum, over $t \leqslant T$, of the expectations of all terms in (22), conditional on $\mathcal{O}\!\ell_t$, are finite w.p.1.

By adding the results for (22a) and (22c) and for (22b) and (22d),

$$\text{(a)}\quad E^{\mathcal{O}\!\ell_t}\int_0^t (\exp R_s) g'(\bar{x}_s, s)\Sigma_s^{-1}\, dy_s = \int_0^t [E^{\mathcal{O}\!\ell_t}(\exp R_s) g'(\bar{x}_s, s)\Sigma_s^{-1}]\, dy_s,$$

$$\text{(b)}\quad E^{\mathcal{O}\!\ell_t}\int_0^t (\exp R_s) h(\bar{x}_s) g'(x_s, s)\Sigma_s^{-1}\, dy_s = \int_0^t [E^{\mathcal{O}\!\ell_t}(\exp R_s) h(\bar{x}_s) g'(\bar{x}_s, s)\Sigma_s^{-1}]\, dy_s, \tag{23}$$

$$\text{(c)}\quad E^{\mathcal{O}\!\ell_t}\int_0^t (\exp R_s) Lh(\bar{x}_s)\, ds = \int_0^t E^{\mathcal{O}\!\ell_t}[(\exp R_s) Lh(\bar{x}_s)]\, ds.$$

Now, note that, on the right sides of (23), the expectation $E^{\mathcal{O}\!\ell_s}$ is equivalent to the expectation $E^{\mathscr{F}_s}$, and the Theorem is proved.

Remark. The case where Σ_s is a function of x is degenerate. The value $\Sigma(s, x_s)$ at time 0 may be determined by observing y_s, $0 \leqslant s \leqslant \tau$, where τ is arbitrarily small. Divide $[0, \tau]$ into N_k units of length Δ_k, $N_k\Delta_k = \tau_k$. Form

$$Q_k = \frac{1}{N_k}\sum_1^N \frac{\delta y_{ki}^2}{\Delta_k} = \frac{1}{N_k}\sum_1^N (G_{ki} + \delta w_{ki})^2/\Delta_k .$$

Under mild additional hypothesis, it can be shown, via the strong law of large numbers, that, as $\tau_k \to 0$, $\Delta_k \to 0$, $N_k \to \infty$, $Q_k \to \Sigma(0, x_0)$ w.p.1. The problem is essentially one of computing the variance of a normally distributed variate when infinitely many independent observations are available. The degeneracy arises owing to the fact that the observation noise is "white."

5. Remarks on Construction of Physical Systems Corresponding to Eq. (4)

Let $h(x_t) = x_{it}$, the ith component of the vector x_t. Then $Lh(x) = f_i(x, t)$ and the equation for the conditional mean $E^{\mathscr{F}_t} x_{it} = m_{it}$ is easily obtained from (4).

$$dm_{it} = (dy_t - E^{\mathscr{F}_t} g(x_t, t)dt'\Sigma_t^{-1}(E^{\mathscr{F}_t}(x_{it} g(x_t, t)) - m_{it}E^{\mathscr{F}_t} g(x_t, t)) + E^{\mathscr{F}_t} f_i(x_t, t)dt.$$

Similarly, the equations for $c_{ijt} \equiv E^{\mathscr{F}t}x_{it}x_{jt}$ can be obtained. Then, the equation for the covariances $m_{ijt} \equiv E^{\mathscr{F}t}(x_{it} - m_{it})(x_{jt} - m_{jt})$ are obtained from $d[c_{ijt} - m_{it}m_{jt}]$ and Itô's Lemma (to obtain the differential of the product of stochastic integrals $m_{it}m_{jt}$). This procedure is valid and can be carried further (higher moments obtained) provided that (A1)–(A11) hold for the necessary $h(x)$ functions.

Let (A1)–(A11), corresponding to functions $h(x) = \{h_i(x), i = 1,..., M\}$, hold. First assume: (B1); that the right side of the equations (4) for $d(E^{\mathscr{F}t}h_i(x_t))$, $i = 1,..., M$, involve only functions of $E^{\mathscr{F}t}h(x_t)$; i.e., $E^{\mathscr{F}t}g(x_t, t) = F(E^{\mathscr{F}t}h(x_t))$ for some function $F(\cdot)$, etc. Then, the system (4) of equations for the $d(E^{\mathscr{F}t}h_i(x_t))$ has the usual form of the vector Itô equation. If the uniform Lipschitz and growth conditions are satisfied, the system (4) has a unique solution which is a version of the conditional expectation of the vector $h(x_t)$.

Let samples $\delta y_i = y_{t_{i+1}} - y_{ti}$ be available for small $\Delta_i = t_{i+1} \equiv t_i$. Then, by writting all differentials in the Itô equation as finite differences, the Itô equation transforms into a difference equation. If a continuous parameter process is obtained (from the solution of the difference equation) by a suitable interpolation, then, as the difference interval goes to zero, the result of the interpolation converges to the solution of the Itô equation w.p.l., for each t [*11*]. This suggests that, for "sufficiently small" Δ_i, the difference scheme, applied to (4), would yield a useful approximation to the conditional expectation.

Dynamical systems for constructing the sample solutions of the Itô equation do not appear to be available. Although the introduction of the Wiener process w_t seems necessary for careful theoretical work, the true physical observation may be of the form $g(x_t, t) + \psi_t$, where ψ_t is a well-defined process—unlike dw_t/dt. If $\int_0^t \psi_s \, ds$ has a distribution close to that of w_t, but ψ_t is still a well-defined process, then using (4), valid or not, we may divide (4) by $dt(dy/dt = g + \psi)$ and obtain a differential equation. Since the right side of (4) contains, by (B1), only functions of '$E^{\mathscr{F}t}h(x_t)$', a dynamical system corresponding to the resulting equation may now be built. The observation dy/dt occurs as a "driving term." One would like to assert that the solution process approximates the process $E^{\mathscr{F}t}h(x_t)$. The validity of such an assertion is closely connected to the relation between the solution of Itô's equation using Itô's constructive method, and the "solution" when the rules of ordinary integration are used. We mention only that a very similar question, on the relation between solutions to equations interpreted in the ordinary and in the Itô sense, has been treated in [*10*]. The general conclusion is that, if $\int^t \psi_s \, ds$ is close to w_t in distribution, then to each equation interpreted in the Itô sense, there is a second equation, possibly containing extra terms, to which the application of the ordinary calculus yields a solution with a distribution "close" to that of the solution to the Itô equation. This question is also related to the difference in results between [2] and [5].

Now, drop Assumption (B1). Then the right side of (4) contains terms $E^{\mathscr{F}_t}Q$ which are not functions of $E^{\mathscr{F}_t}h(x_t)$. A number of interesting possibilities for approximation of the $E^{\mathscr{F}_t}Q$ arise. Some of these will hopefully be discussed elsewhere, in connection with results of some current numerical and experimental studies. There are the obvious approaches of either neglecting such terms or using an approximation (e.g., by a truncated Taylor series) of $E^{\mathscr{F}_t}Q$ in terms of $E^{\mathscr{F}_t}h(x_t)$. Both involve serious pitfalls where "nonlinearities" in g, f and V are "large."

A seemingly promising discrete-parameter type of approximation, which is currently under study follows; namely, compute the equations for dm_{it}, dm_{ijt}, and, perhaps, dm_{ijkt}. Convert the set to finite-difference form. Arbitrarily assume a multivariate distribution D_n, e.g., normal, uniform, etc. Compute the parameters of the distribution from m_{it_n}, m_{ijt_n}, and perhaps m_{ijkt_n}. Then compute the necessary expectations of all terms Q with respect to $D_n(E_{D_n}Q)$ and let $E_{D_n}Q$ replace $E\mathscr{F}_{tn}Q$. Then compute the "conditional moments" at t_{n+1}, etc. Distributions D_n, suitable for the problem, must, of course, be found.

References

1. Kushner, H. J., On the dynamical equations of conditional probability density functions, with applications to optimal stochastic control theory. *J. Math. Anal. Appl.* **8** (1964), 332-344.
2. Kushner, H. J., On the differential equation satisfied by conditional probability densities of Markov processes. *J. SIAM Control,* Ser. A. **2** (1964), 106-119.
3. Bucy, R. C., Nonlinear filtering theory. *IEEE Trans. Automatic Control* **10** (1965), 198-199.
4. Wonham, W. M., Some applications of stochastic differential equations to optimal nonlinear filtering. *J. SIAM Control, Ser.* A. **2** (1964), 347-369.
5. Stratonovich. R. I., Conditional Markov processes. *Theory Prob. Appl.* **5** (1960), 156-178.
6. Doob, J. L., "Stochastic Processes." Wiley, New York, 1953.
7. Dynkin, E. B., "Markov Processes." Springer-Verlag, Berlin, 1965.
8. Stratonovich, R. I., A new representation for stochastic integrals and equations. *Vestn. Moskov. Univ., Ser. Mat. Mekhan.* **1** (1964), 3-12.
9. Kalman, R. E. and Bucy, R. C., New results in linear filtering and prediction theory. *Trans. ASME, J. Basic Engr.* **83**D (1961).
10. Wong, E. and Zakai, M., On the convergence of ordinary integrals to stochastic integrals. *Ann. Math. Stat.* **36** (1965), 1560-1564.
11. Maruyama, G., Continuous Markov processes and stochastic equations. *Rend. Circ. Mat. Palermo* **4** (1955), 48-49.
12. Loeve, M., "Probability Theory," 3rd ed. Van Nostrand, Princeton, New Jersey, 1963.