SOME APPLICATIONS OF STOCHASTIC DIFFERENTIAL EQUATIONS TO OPTIMAL NONLINEAR FILTERING*

W. M. WONHAM†

1. Introduction. A current problem in control theory is that of estimating the dynamical state of a physical system, on the basis of data perturbed by noise. Solution of the estimation problem is usually immediate if one knows the probability distribution of the system state at each instant of time, conditional on the data available up to that instant. It is therefore of interest to ask how this *posterior* probability distribution evolves with time, and if possible to specify the dynamical structure of a filter (i.e., analog device) which generates the posterior distribution when its input is the time function actually observed.

In the present paper, filters of this type are defined by means of stochastic differential equations[1] for the posterior distribution in which the observed time function appears as a forcing term. Differential equations for this purpose were introduced in 1960 by Stratonovič [1], who also indicated their application to stochastic control problems [2]. When the dynamical system under observation is linear and the noise is white Gaussian it has been shown [3] that Stratonovič's equation can be solved formally to yield the stochastic differential equation of the optimal (linear) filter. When the function to be estimated is a Markov step process and the noise is white Gaussian the optimal (nonlinear) filter equations were stated in [4]. The latter equations are discussed in more detail in §3, below; they differ from those of Stratonovič in a sense to be noted in the sequel. For one example, discussed in §3, performance of the optimal nonlinear filter is evaluated numerically and is found to be somewhat better than that of the simpler Wiener filter, particularly when the noise intensity is low.

In §4, the equations of §3 are generalized heuristically to the case where the state space of the step process is continuous, and in §5 some tentative remarks are made on the form of the solutions.

Some parallel work on noisy observation of a diffusion process has been reported by Kushner in a recent paper [5].

[1] A brief review of stochastic differential equations is given in Appendix 1.

2. Noisy measurement of an unknown constant. The basic idea of a "functional" filter is illustrated by the following simple estimation problem. Let x be a discrete, real-valued random variable with range of values a_1, $\cdots$, a_K and a priori probability distribution $\{p_j(0), j = 1, \cdots, K\}$ at $t = 0$. Suppose that one observes the function

$$(1) \qquad y(t) = xt + \int_0^t \beta(s) \, dw(s), \qquad\qquad t \geqq 0,$$

where the function β is known to the observer,[2] and $\{w(t), t \geq 0\}$ is a Wiener process which is independent of x, with[3] $P\{w(0) = 0\} = 1$. The process $y(t)$ defined by (1) can also be written as the solution of the stochastic differential equation

$$(2) \qquad dy(t) = x \, dt + \beta(t) \, dw(t), \quad y(0) = 0.$$

Dividing formally by dt one obtains the possibly more familiar version

$$(3) \qquad \dot{y}(t) = x + \beta(t)\dot{w}(t), \quad y(0) = 0,$$

where $\dot{w}$ represents Gaussian white noise. Since w is not differentiable in the ordinary sense we shall use instead the differential notation of (2) and interpret (2) to mean the integral representation (1) (see Appendix 1).

Let us now introduce the posterior distribution

$$(4) \qquad p_j(t) = P\{x = a_j | \, y(s), 0 \leqq s \leqq t\}, \qquad j = 1, \cdots, K.$$

Evaluation of $p_j(t)$ is straightforward (see Appendix 2); the result is[4]

$$(5) \qquad p_j(t) = \frac{p_j(0) \exp\left[a_j \int_0^t \beta(s)^{-2} \, dy(s) - \frac{1}{2} a_j^2 \int_0^t \beta(s)^{-2} \, ds \right]}{\sum_{k=1}^{K} p_k(0) \exp\left[a_k \int_0^t \beta(s)^{-2} \, dy(s) - \frac{1}{2} a_k^2 \int_0^t \beta(s)^{-2} \, ds \right]}.$$

The stochastic integral in (5) is well-defined [6, Chap. IX, §2]. Now write $p(t) = [p_1(t), \cdots, p_K(t)]$ and consider the joint process $\{x, p(t), t \geqq 0\}$, where we regard x as a fixed random variable with distribution $\{p_j(0), j = 1, \cdots, K\}$. Since almost every $w(t)$ sample function is continuous, the same is true of the $p(t)$ sample functions (by [6, Chap. IX, Theorem

---

[2] We shall assume that β is continuously differentiable, and bounded away from 0 for $t \geqq 0$.

[3] It is assumed that an underlying probability space with elementary events ω is given. Probability measure on this space is denoted by P. To simplify notation, ω-dependence of random variables will not be indicated. In the Appendices "Borel field" means "Borel field of ω-sets".

[4] The qualification "with probability 1" on equalities between conditional probabilities is to be understood.

5.2]). Moreover, it is easily seen (Appendix 3) that the $\{x, p(t)\}$ process is Markov.

Our aim is to describe the evolution in time of the p_j's by means of a system of stochastic differential equations. Having verified the existence of the limits (7) and (8) written below, one can apply to the Markov process $\{x, p(t)\}$ a representation theorem of Doob [6, Chap. VI, Theorem 3.3]. Alternatively, since the $p_j(t)$ are known explicitly, it is more direct to apply a result of Dynkin (Appendix 1 or [7, Theorem 7.2]), and this gives

$$(6) \qquad dp_j(t) = m_j[t, x, p(t)] \, dt + \sigma_j[t, x, p(t)] \, dw(t), \qquad j = 1, \cdots, K.$$

The functions m_j and σ_j in (6) have the probabilistic meaning

$$(7) \qquad m_j(t, \xi, \pi) = \lim_{h \to 0} E\left\{ \frac{p_j(t+h) - p_j(t)}{h} \,\middle|\, x = \xi, p(t) = \pi \right\},$$

and

$$\sigma_i(t, \xi, \pi)\sigma_j(t, \xi, \pi)$$
$$(8) \quad = \lim_{h \to 0} E\left\{ \frac{[p_i(t+h) - p_i(t)][p_j(t+h) - p_j(t)]}{h} \,\middle|\, x = \xi, p(t) = \pi \right\};$$
$$i, j = 1, \cdots, K.$$

In Appendix 3 the limits (7) and (8) are computed from (5), using Dynkin's formulas; the results are

$$(9) \qquad\qquad m_j(t, x, p) = \beta(t)^{-2}(x - \bar{x})(a_j - \bar{x})p_j,$$

$$(10) \qquad\qquad \sigma_j(t, x, p) = \beta(t)^{-1}(a_j - \bar{x})p_j,$$

$j = 1, \cdots, K$, where

$$(11) \qquad\qquad \bar{x} = \sum_{k=1}^{K} a_k p_k.$$

Writing out (6) in full and noting (2), we obtain finally

$$
(12) \quad
\begin{aligned}
dp_j(t) = {}& -\beta(t)^{-2}\bar{x}(t)[a_j - \bar{x}(t)]p_j(t) \, dt \\
& + \beta(t)^{-2}[a_j - \bar{x}(t)]p_j(t) \, dy(t), \quad j = 1, \cdots, K.
\end{aligned}
$$

The system of stochastic differential equations (12) is the desired result. It can be interpreted as specifying the dynamical structure of a filter (or analog device) which continuously generates the posterior distribution $p(t)$ when the input is the observed function $y(t)$. From a practical viewpoint this interpretation might be useful when the function actually observed is not y, but rather some approximation to the "function" $\dot{y}$ indicated in (3). In that case formal division by dt in (12) yields a system of

nonlinear differential equations for the p_j's, in which $\dot{y}$ appears as a forcing term. This system of equations can be simulated by an analog device with input $\dot{y}$ and output p, and thus represents the filter.

The Ito equation (12) is solved (Appendix 1) by replacing (12) by an integral equation and constructing the solution q by successive approximations. It is plausible that the structure of the analog feedback device should be chosen to model that of the successive approximation scheme. However, it is not yet known precisely in what sense a physical process must approximate a $\dot{y}$ "process" in order that the filter output be a useful approximation to the desired posterior distribution. Experimental work along these lines would be of considerable interest.

Example. It is worth emphasizing that the ordinary rules of integration cannot be applied to the stochastic equation (12) in an attempt to regain the explicit solution (5). To illustrate this fact let x have possible values $a_1 = +1$, $a_2 = -1$ and let $p_1(0) = p_2(0) = \frac{1}{2}$, $\beta \equiv 1$. Since $p_1(t) + p_2(t) \equiv 1$, it is sufficient to consider

$$(13) \qquad\qquad q(t) = p_1(t) - p_2(t).$$

From (11), $\bar{x}(t) = q(t)$; and from (12) the filter is defined by

$$(14) \qquad\qquad dq = -q(1 - q^2)\, dt + (1 - q^2)\, dy.$$

On the other hand (5) yields the evaluation

$$(15) \qquad\qquad q(t) = \tanh\,[y(t)],$$

and formal differentiation of (15) gives

$$(16) \qquad\qquad dq = (1 - q^2)\, dy.$$

We observe that the first term on the right side of (14) is absent from (16). The point is simply that stochastic differential equations of Ito type cannot be manipulated by the usual formal rules (cf. [6, Chap. IX, §5]); also, it is plausible that an analog device for generating q should be set up according to the Ito equation (14), and not according to the "formal" equation (16).

It must be noted finally that Stratonovič's procedure [1] applied to this example leads to (16) and not to (14) (cf. [2, (9)]). In general it appears that the coefficient of dt in Stratonovič's equations (when these are written in differential notation) cannot be given the interpretation indicated in (7).

3. Noisy observations of a Markov step process. Let $\{x(t),\ t \geqq 0\}$ be a stationary Markov step process[5] with a finite number of states (distinct

---

[5] For a discussion of the process see [6, Chap. VI, §1], where our ν_i, ν_{ij} are denoted by q_i, q_{ij}.

step levels) $a_1, \cdots, a_K$. Denote the transition probabilities by

$$p_{ij}(h) = P\{x(t + h) = a_j | x(t) = a_i\}.$$

We assume that

$$(17) \qquad p_{ij}(h) = \begin{cases} 1 - \nu_i h + o(h), j = i & (h \to 0) \\ \nu_{ij} h + o(h), j \neq i & (h \to 0), \end{cases}$$

where the $\nu_{ij} \geqq 0$ are constants and

$$(18) \qquad \nu_i = \sum_{\substack{j=1 \\ j \neq i}}^{K} \nu_{ij}, i = 1, \cdots, K.$$

Let the distribution of $x(0)$ be $\{p_j(0), j = 1, \cdots, K\}$. To avoid triviality we assume that $\{p_j(0)\}$ is not concentrated on any absorbing state among the a_j's.

As in §2, suppose that the process observed is $\{y(t), t \geqq 0\}$ defined by

$$(19) \qquad dy(t) = x(t)\, dt + \beta(t)\, dw(t), \qquad\qquad t \geqq 0,$$

where $P\{y(0) = 0\} = 1$ and the Wiener process $\{w(t), t \geqq 0\}$ is independent of the $x(t)$ process. Introduce the posterior probabilities

$$(20) \qquad p_j(t) = P\{x(t) = a_j | y(s), 0 \leqq s \leqq t\}, \quad j = 1, \cdots, K.$$

We now seek a system of stochastic differential equations for the $p_j(t)$. In contrast to the situation in §2, a representation of the $p(t)$ process as an explicit functional of the $y(t)$ process (cf. (5)) is apparently no longer available. However, the appropriate generalization of (12) can be obtained by an application of Doob's theorem [6, Chap. VI, Theorem 3.3]. The equations, derived in Appendix 4, are

$$(21) \qquad \begin{aligned} dp_j(t) = {}& [-\nu_j p_j(t) + \sum_{\substack{i=1 \\ i \neq j}}^{K} \nu_{ij} p_i(t)]\, dt - \beta(t)^{-2} \bar{x}(t)[a_j - \bar{x}(t)]p_j(t)\, dt \\ & + \beta(t)^{-2}[a_j - \bar{x}(t)]p_j(t)\, dy(t), \quad j = 1, \cdots, K. \end{aligned}$$

Equation (21) is the main result of this paper. The previous remarks on the interpretation of (12) apply also to (21): the equation defines the structure of an ideal analog device for generating the p_j's from the data, y. Comparison of (21) with (12) shows that the only new term in (21) is the first term on the right side. This term is of the form $\mathcal{L}^+[p]\, dt$ where $\mathcal{L}^+$ is the forward operator of the $x(t)$ process, and thus represents a change in p due to the observer's a priori knowledge of how the $x(t)$ process evolves.

We shall now discuss a simple special case of (21) in detail.

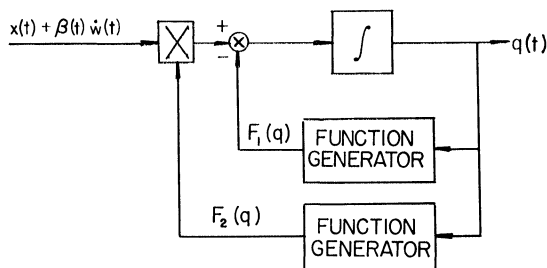*Example.* Let $\beta \equiv$ const. and suppose that x is a "random telegraph sig-

FIG. 1. *Optimal nonlinear filter: Example, §3.*

$$F_1(q) = 2\nu q + \beta^{-2}q(1 - q^2)$$

$$F_2(q) = \beta^{-2}(1 - q^2)$$

nal" [11]; that is,

(22)
$$a_1 = +1, \quad a_2 = -1,$$
$$\nu_i = \nu_{ij} = \nu, \quad i, j = 1, 2.$$

The parameter ν is the expected number of jumps of $x(\cdot)$ in unit time. Let

(23)
$$q(t) = p_1(t) - p_2(t).$$

From (21) and (22),

(24)
$$dq = -2\nu q \, dt - \beta^{-2}q(1 - q^2) \, dt + \beta^{-2}(1 - q^2) \, dy,$$

or equivalently

(25)
$$dq = [-2\nu q - \beta^{-2}(1 - q^2)(q - x)] \, dt + \beta^{-1}(1 - q^2) \, dw.$$

A block diagram of the optimal filter is shown in Fig. 1.

We shall evaluate the filter performance in terms of mean square estimation error. Thus the optimal estimate of $x(t)$ is

(26)
$$\hat{x}(t) = E\{x(t) \mid y(s), 0 \leqq s \leqq t\}$$
$$= a_1 p_1(t) + a_2 p_2(t)$$
$$= q(t).$$

In Appendix 4 it is shown that the joint process $\{x(t), q(t), t \geqq 0\}$ is Markov. It will be assumed that this process has stationary densities $\pi^{\pm}(q)$, $-1 \leqq q \leqq 1$, defined by

(27)
$$\pi^{\pm}(q) \, dq = P\{x(t) = \pm1, q(t) \in (q, q + dq)\}.$$

Then the mean square estimation error is

$$(28) \qquad \sigma^2 = \int_{-1}^{1} (1 - q)^2 \pi^+(q) \, dq + \int_{-1}^{1} (1 + q)^2 \pi^-(q) \, dq.$$

It remains to calculate the densities $\pi^\pm$. From (25), the stationary Kolmogorov forward equation [6, Chap. VI] of the $\{x, q\}$ process is

$$(29) \qquad \tfrac{1}{2}\beta^{-2}[(1 - q^2)^2 \pi^\pm(q)]'' - [\beta^{-2}(\pm 1 - q)(1 - q^2)\pi^\pm(q)$$
$$- 2\nu q \pi^\pm(q)]' \pm \nu[\pi^-(q) - \pi^+(q)] = 0,$$

where $'$ denotes d/dq. With the symmetry condition $\pi^-(q) = \pi^+(-q)$, (29) has the unique solution

$$(30) \qquad \pi^\pm(q) = c(1 \pm q)(1 - q^2)^{-2} \exp [-2\mu(1 - q^2)^{-1}],$$

where

$$(31) \qquad c = \left[2 \int_1^\infty z^{1/2}(z - 1)^{-1/2} e^{-2\mu z} \, dz \right]^{-1}$$

and

$$(32) \qquad \mu = \beta^2 \nu.$$

From (28), (30) the stationary error variance is then

$$(33) \qquad \sigma^2 = \frac{\displaystyle\int_0^\infty z^{-1/2}(z + 1)^{-1/2} e^{-2\mu z} \, dz}{\displaystyle\int_0^\infty z^{-1/2}(z + 1)^{1/2} e^{-2\mu z} \, dz}$$
$$= \begin{cases} -2\mu \log 2\mu + o(\mu \log \mu) & \text{as } \mu \to 0, \\ 1 - (4\mu)^{-1} + O(\mu^{-2}) & \text{as } \mu \to \infty. \end{cases}$$

The result (33) will be compared with the error variance of a Wiener filter which is optimal for the same input. For the Wiener filter a standard computation yields

$$(35) \qquad \begin{aligned} \sigma_w{}^2 &= 2\mu[(1 + \mu^{-1})^{1/2} - 1] \\ &= \begin{cases} 2\mu^{1/2} + o(\mu^{1/2}) \text{ as } \mu \to 0, \\ 1 - (4\mu)^{-1} + O(\mu^{-2}) \text{ as } \mu \to \infty. \end{cases} \end{aligned}$$

Numerical results are given in Fig. 2. Since the nonlinear filter generates the estimate $\hat{x}(t)$ defined by (26) it is necessarily optimal (with respect to error variance) in the class of all filters which operate on the present and past of the data y. Thus $\sigma^2 \leq \sigma_w{}^2$; in fact, $\sigma^2/\sigma_w{}^2$ is substantially less than unity when the noise intensity is low.
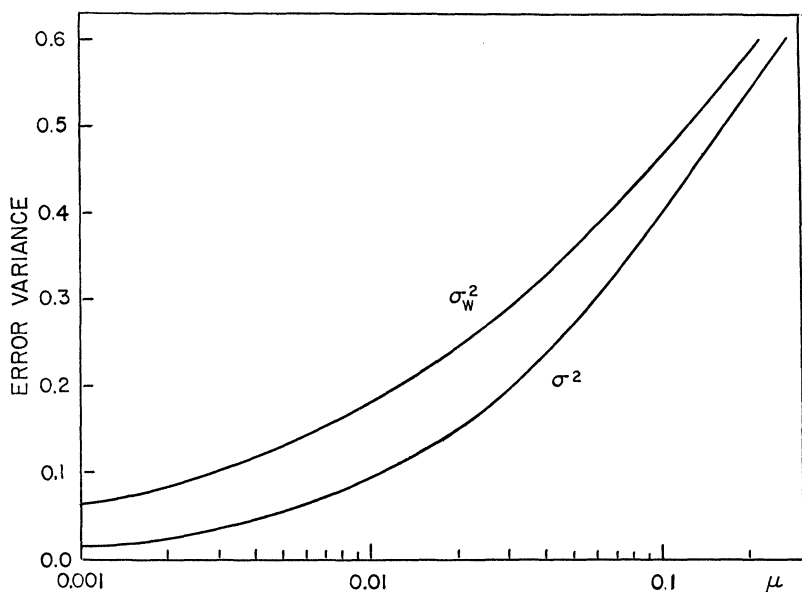
FIG. 2. *Numerical results for example*, §3.

4. Heuristic generalization of (21) to a continuous state space. The differential equations (12), (21) were derived on the assumption that the state space of the $x(t)$ process is a finite set. The following is a heuristic generalization to a continuous state space. Let $\{x(t), t \geqq 0\}$ be a real-valued Markov step process with state space X, where X is a closed finite interval. Let $\nu(\xi)$, $\nu(\xi, A)$ be defined for ξ in X and A a Borel subset of X; the functions $\nu(\cdot)$, $\nu(\cdot, \cdot)$ are assumed to be a "standard pair" in the sense of Doob [6, Chap. VI, §2][6]. In addition, $\nu(\cdot)$ is assumed to be bounded on X; the $x(t)$ sample functions are then almost all step functions [6]. In analogy to (17) one has

$$
\begin{aligned}
p(h, \xi, A) &= P\{x(t+h) \in A \mid x(t) = \xi\} \\
(36) \qquad &= \begin{cases} \nu(\xi, A)h + o(h), \ \xi \notin A, \\ 1 - \nu(\xi)h + o(h), \ A = \{\xi\}. \end{cases}
\end{aligned}
$$

As in §3, suppose next that

$$
(37) \qquad dy(t) = x(t) \, dt + \beta(t) \, dw(t), \qquad\qquad t \geqq 0,
$$

and introduce the posterior probability

$$
(38) \qquad p(t, A) = P\{x(t) \in A \mid y(s), 0 \leqq s \leqq t\}.
$$

[6] The functions ν are denoted in [6] by q.

Then inspection of (21) suggests the generalization

$$dp(t, A) = \left[-\int_A \nu(\xi, X - A)p(t, d\xi) + \int_{X-A} \nu(\xi, A)p(t, d\xi) \right] dt$$

$$(39) \qquad\qquad - \left[\beta(t)^{-2} \bar{x}(t) \int_A [\xi - \bar{x}(t)]p(t, d\xi) \right] dt$$

$$+ \left[\beta(t)^{-2} \int_A [\xi - \bar{x}(t)]p(t, d\xi) \right] dy(t).$$

In (39),

$$(40) \qquad\qquad \bar{x}(t) = \int_X \xi p(t, d\xi),$$

and A is an arbitrary Borel subset of X.

Just as in the case of a (finite-dimensional) Ito equation, (39) might plausibly be interpreted by starting from the corresponding integral equation, obtained by integrating both sides of (39) with respect to t, and seeking the solution as the limit of successive approximations. Unlike (12) and (21), however, the general equation (39) cannot readily be interpreted as specifying the dynamics of a practically realizable filter for generating p from the data y.

5. Sufficient statistics. We have seen in §4 that the stochastic differential equation for the posterior distribution of $x(t)$ cannot be used directly, in general, to design an optimal filter. In practice, construction of the posterior distribution must be reduced to the evaluation of a "small" number of functionals (sufficient statistics) on the observed function y. For example, inspection of (5) shows that the K-dimensional stochastic system (12) can be replaced by the 1-dimensional system

$$(41) \qquad\qquad dz(t) = \beta(t)^{-2} dy(t), \quad z(0) = 0,$$

which generates the sufficient statistic

$$(42) \qquad\qquad z(t) = \int_0^t \beta(s)^{-2} dy(s).$$

On the other hand the writer knows of no similar reduction of the system (21) or of (39).

Even if a nontrivial sufficient statistic $z(t)$ for the determination of p exists, it may be impossible to write z as an explicit functional of y. It would be enough to know, however, that z satisfies a stochastic differential equation

$$(43) \qquad\qquad dz = \zeta_1(t, z)dt + \zeta_2(t, z) dy,$$

where the functions ζ_1, ζ_2 are known; then in principle z could be obtained as the output of an analog device set up according to (43). Thus the "solution" of an equation of type (21), (39) might take the form of a (known) function of a statistic z which satisfies a (known) equation of type (43). The investigation of solutions of this type (if they exist) would be of considerable interest.

Acknowledgment. The author is indebted to H. J. Kushner (RIAS) and to J. Ternan (Cambridge University) for helpful discussions.

Appendix 1. Stochastic differential equations.

1. Since stochastic differential equations are not yet widely used in engineering applications we mention here some definitions and known results. For a detailed account the reader is referred to Doob [6, Chap. VI, §3] and Dynkin [7, Chaps. 7, 11].

Let $\{z(t),\ t \geqq 0\}$ be a stochastic process in K-dimensional Euclidean space R^K; we write $z = (z_1, \cdots, z_K)$, where the z_i are real-valued, and put $\|z\| = (\sum_1^K z_i^2)^{1/2}$. Let $\{w(t),\ t \geqq 0\}$ be a Wiener process (Brownian motion process) in R^J; that is, $w(t) = [w_1(t), \cdots, w_J(t)]$ where the $w_i(t)$ are independent Wiener processes in R^1 and, for $t \geqq s \geqq 0$,

$$(44) \qquad E\{[w_i(t) - w_i(s)][w_j(t) - w_j(s)]\} = \begin{cases} t - s, j = i, \\ 0, j \neq i. \end{cases}$$

(See [6, Chap. II, §9] for the definition of the Wiener process in R^1.)

The stochastic differential equation of interest here is written

$$(45) \qquad dz(t) = m[t, z(t)]\, dt + \sigma[t, z(t)]\, dw(t), \qquad\qquad t \geqq 0,$$

where m is a K-vector and σ is a $K \times J$ matrix. Loosely interpreted, (45) states that in a small time interval $(t,\ t + dt)$ the vector $z(t)$ suffers a "dynamical" displacement $m[t,\ z(t)]\, dt$ plus a random displacement $\sigma[t, z(t)]\, dw(t)$, where the latter is a Gaussian random vector with mean 0 and covariance matrix $\sigma[t, z(t)]\sigma'[t, z(t)]\, dt$ ($'$ denotes transpose).

Dividing both sides of (45) formally by dt one obtains

$$(46) \qquad\qquad \dot{z} = m(t, z) + \sigma(t, z)\dot{w} \qquad\qquad (\cdot = d/dt),$$

where $\dot{w}$ is a J-vector whose components are independent "Gaussian white noise processes". The notation of (46) has been more common in the engineering literature than that of (45) but it is objectionable for two reasons: (a) almost all $w(t)$ sample functions are nondifferentiable almost everywhere, a fact which is intuitively plausible if we note that $E\|dw(t)\|$ is proportional to $(dt)^{1/2}$;

(b) when applied to (46) the usual formal rules of integration lead in general to results which are definitely incorrect. This statement will be illustrated later.

2. The stochastic differential equation (45) does not specify the value of a derivative. According to Gihman [9], such an equation can be defined by giving an explicit procedure for constructing the solution, but then has meaning only insofar as this construction can be carried out. An alternative interpretation of (45) due to Ito [8] is the following. Replace (45) by the stochastic integral equation

$$(47) \qquad z(t) = z(0) + \int_0^t m[s, z(s)] \, ds + \int_0^t \sigma[s, z(s)] \, dw(s).$$

The stochastic integrals on the right side of (47) are defined [6, Chap. IX, §2, §5] under suitable restrictions on the (random) functions $m[\cdot, z(\cdot)]$ and $\sigma[\cdot, z(\cdot)]$. Ito's construction of a $z(t)$ process which satisfies (47) is carried out by successive approximation; one sets $z^{(0)}(t) \equiv 0$ and

$$(48) \qquad z^{(n+1)}(t) = z(0) + \int_0^t m[s, z^{(n)}(s)] \, ds + \int_0^t \sigma[s, z^{(n)}(s)] \, dw(s),$$

$$n = 0, 1, 2, \cdots.$$

Conditions which guarantee that the sequence $\{z^{(n)}\}$ converges for t in a finite interval $[0, T]$ are given by Doob [6, IX, §3][7] and are, mainly, that the functions $m(t, z)$, $\sigma(t, z)$ satisfy a uniform Lipschitz condition in z, and are bounded in norm by $C(1 + \| z \|^2)^{1/2}$ where C is some constant. Then there exists a process $\{z(t), 0 \leqq t \leqq T\}$ with the following properties. (i) $\lim_{n \to \infty} z^{(n)}(t) = z(t)$ uniformly in t with probability 1; and the $z(t)$ process is unique, in the sense that the difference of two solutions is zero, with probability 1, for each t.
(ii) The $z(t)$ sample functions are, with probability 1, continuous in $[0, T]$.
(iii) For each $t \in [0, T]$, (47) is true with probability 1.
(iv) If the initial value $z(0)$ is a random variable which is independent of the increments $\{w(t_2) - w(t_1), t_1, t_2 \in [0, T]\}$, then $\{z(t), 0 \leqq t \leqq T\}$ is a Markov process.

In addition, the $z(t)$ process has the following local properties, which make precise the interpretation of (45) given earlier:

$$(49) \quad \text{(v)} \quad \lim_{h \to 0} E \left\{ \frac{z(t + h) - z(t)}{h} \, \middle| \, z(t) = \zeta \right\} = m(t, \zeta),$$

$$\text{(vi)} \quad \lim_{h \to 0} E \left\{ \frac{[z_i(t + h) - z_i(t)][z_j(t + h) - z_j(t)]}{h} \, \middle| \, z(t) = \zeta \right\}$$

$$(50) \qquad = \sum_{r=1}^{J} \sigma_{ir}(t, \zeta) \sigma_{jr}(t, \zeta)$$

$$= b_{ij}(t, \zeta), \quad \text{say}; \qquad i, j = 1, \cdots, K.$$

---

[7] Doob's treatment for $K = J = 1$ can be generalized after replacing $| m |$ by $\| m \|$ and $| \sigma |$ by $\| \sigma \| = (\sum_i \sum_j \sigma_{ij}^2)^{\frac{1}{2}}$ (cf. [7, Chap. 7]).

3. Assume for the moment that the Markov process $\{z(t), 0 \leqq t \leqq T\}$ constructed above has a transition probability density $p = p(s, z; t, \zeta)$, defined for $0 \leqq s < t \leqq T$ and $z, \zeta \in R^K$. It would be convenient if the hypotheses made on m and σ (strengthened to include the differentiability needed below) guaranteed that the density p exists and satisfies the Kolmogorov equations

$$(51) \qquad -\frac{\partial p}{\partial s} = \frac{1}{2} \sum_{i=1}^{K} \sum_{j=1}^{K} b_{ij}(s, z) \frac{\partial^2 p}{\partial z_i\, \partial z_j} + \sum_{i=1}^{K} m_i(s, z) \frac{\partial p}{\partial z_i}$$

$$(52) \qquad \frac{\partial p}{\partial t} = \frac{1}{2} \sum_{i=1}^{K} \sum_{j=1}^{K} \frac{\partial^2}{\partial \zeta_i\, \partial \zeta_j} [b_{ij}(t, \zeta)p] - \sum_{i=1}^{K} \frac{\partial}{\partial \zeta_i} [m_i(t, \zeta)p].$$

Unfortunately it does not seem possible to establish (51) and (52) without making a priori assumptions on the smoothness of p (see the discussion in [6, Chap. VI, §3]). Nevertheless, if the $z(t)$ process obtained by solving the Ito equation (45) has a transition density which satisfies the Kolmogorov equations, then the coefficients which appear in the latter are related to the functions m and σ of the Ito equation according to (49)–(52) above. Used heuristically, this correspondence between (45) and (51), (52) is convenient in engineering applications (see, e.g., [10]).

4. The following example[8] shows that Ito equations cannot be manipulated by the ordinary rules of integration (cf. also [6, p. 443]). Let $\{w(t), t \geqq 0\}$ be a Wiener process in R^1 with $P\{w(0) = 0\} = 1$. Let

$$(53) \qquad\qquad\qquad z(t) = e^{w(t)}, \qquad\qquad\qquad t \geqq 0.$$

It can be shown that the $z(t)$ process can be represented as the solution of an Ito equation (45). From (49), (50) we find

$$(54) \qquad\qquad\qquad m(\zeta) = \tfrac{1}{2}\zeta, \quad \sigma(\zeta) = \zeta,$$

so that

$$(55) \qquad\qquad dz(t) = \tfrac{1}{2}z(t)\, dt + z(t)\, dw(t), \qquad\qquad t \geqq 0,$$

with initial condition $P\{z(0) = 1\} = 1$. On the other hand if (55) is replaced by

$$(56) \qquad\qquad\qquad \dot{z} = \tfrac{1}{2}z + z\dot{w}, \qquad z(0) = 1,$$

and the last equation integrated formally, the result is

$$(57) \qquad\qquad\qquad z(t) = e^{\frac{1}{2}t + w(t)}.$$

5. It is useful to know under what conditions a given process $\{\psi(t), t \geqq 0\}$ can be represented as the solution of an Ito equation (45) or equivalently

---

[8] Suggested to the writer by H. J. Kushner.

of the integral equation (47). One such representation theorem is given by Doob [6, Chap. VI, Theorem 3.3]. We will state here a slightly specialized version of a theorem of Dynkin [7, Theorem 7.2].

THEOREM. *Let the process* $\{z(t), t \geqq 0\}$ *satisfy an integral equation of form* (47) *(in particular we can have* $z(t) \equiv w(t)$*) and let* $\phi = \phi(t, \zeta)$ *be a numerical function, twice continuously differentiable in* (t, ζ) *for* $t \geqq 0$ *and* ζ *in* R^K. *Put* $\psi(t) \equiv \phi[t, z(t)]$. *Then the process* $\{\psi(t), t \geqq 0\}$ *satisfies the integral equation*

$$(58) \qquad \psi(t) = \psi(0) + \int_0^t \tilde{m}[s, z(s)] \, ds + \int_0^t \sum_{r=1}^J \tilde{\sigma}_r[s, z(s)] \, dw_r(s).$$

The functions $\tilde{m}$ *and* $\tilde{\sigma}_r$ *are given by*

$$(59) \qquad \tilde{m}(t, \zeta) = \frac{\partial \phi(t, \zeta)}{\partial t} + \sum_{i=1}^K \frac{\partial \phi(t, \zeta)}{\partial \zeta_i} m_i(t, \zeta) \\ + \frac{1}{2} \sum_{i=1}^K \sum_{j=1}^K \frac{\partial^2 \phi(t, \zeta)}{\partial \zeta_i \partial \zeta_j} b_{ij}(t, \zeta),$$

where

$$(60) \qquad b_{ij}(t, \zeta) = \sum_{r=1}^J \sigma_{ir}(t, \zeta) \sigma_{jr}(t, \zeta),$$

and

$$(61) \qquad \tilde{\sigma}_r(t, \zeta) = \sum_{i=1}^K \frac{\partial \phi(t, \zeta)}{\partial \zeta_i} \sigma_{ir}(t, \zeta), \qquad r = 1, \cdots, J.$$

Equation (58) is (by definition) equivalent to the stochastic differential equation

$$(62) \qquad d\psi(t) = \tilde{m}[t, z(t)] \, dt + \sum_{i=1}^J \tilde{\sigma}_r[t, z(t)] \, dw_r(t).$$

Equation (62) is more general than (45) in the sense that $\tilde{m}$ and the $\tilde{\sigma}_r$ may not be expressible as functions of (t, ψ); however, by adjoining (62) to (45) we obtain a new system of the same type as before.

Finally, it is seen that (62) can also be written

$$(62') \qquad d\psi(t) = \left\{ \frac{\partial \phi[t, z(t)]}{\partial t} + \frac{1}{2} \sum_{i=1}^K \sum_{j=1}^K \frac{\partial^2 \phi[t, z(t)]}{\partial \zeta_i \partial \zeta_j} b_{ij}[t, z(t)] \right\} dt \\ + \sum_{i=1}^K \frac{\partial \phi[t, z(t)]}{\partial \zeta_i} dz_i(t).$$

Equation (62') exhibits the "chain rule" explicitly.

Appendix 2. Derivation of (5). Let $t > 0$ be fixed and put $s_{rn} = rt/n$, $r = 0, 1, \cdots, n$. It will be verified that $\hat{p}_j(t)$, defined by

$$(63) \qquad \hat{p}_j(t) = \underset{n \to \infty}{\text{l.i.m.}} \, P\{x = a_j \mid y(s_{rn}), \, r = 0, 1, \cdots, n\},$$

is given by the expression on the right side of (5); and then that $\hat{p}_j(t)$ is actually the conditional probability $p_j(t)$ defined by (4). Put $\eta_{rn} = y(s_{rn}) - y(s_{r-1,n})$, $(r = 1, \cdots, n; n = 1, 2, \cdots)$. Then from (1),

$$(64) \qquad \eta_{rn} = \frac{xt}{n} + \int_{(r-1)t/n}^{rt/n} \beta(s) \, dw(s).$$

Thus for each n the random variables $\eta_{rn} - xt/n$, $r = 1, \cdots, n$, are independent and Gaussian with mean 0 and variance

$$(65) \qquad v_{rn} = \int_{(r-1)t/n}^{rt/n} \beta(s)^2 \, ds.$$

Defining

$$(66) \qquad \hat{p}_j^{(n)}(t) = P\{x = a_j \mid y(s_{rn}), \, r = 0, 1, \cdots, n\},$$

we have

$$
\begin{aligned}
(67) \qquad \hat{p}_j^{(n)}(t) &= P\{x = a_j \mid \eta_{rn}, \, r = 1, \cdots, n\} \\
&= \frac{p_j(0) \exp\left[-\sum_{r=1}^{n} \left(\eta_{rn} - a_j \frac{t}{n} \right)^2 (2v_{rn})^{-1} \right]}{\sum_{k=1}^{K} p_k(0) \exp\left[-\sum_{r=1}^{n} \left(\eta_{rn} - a_k \frac{t}{n} \right)^2 (2v_{rn})^{-1} \right]} \\
&= \frac{p_j(0) \exp\left[a_j \left(\frac{t}{n} \right) \sum_{r=1}^{n} \eta_{rn} v_{rn}^{-1} - a_j^2 \sum_{r=1}^{n} \left(\frac{t}{n} \right)^2 (2v_{rn})^{-1} \right]}{\sum_{k=1}^{K} p_k(0) \exp\left[a_k \left(\frac{t}{n} \right) \sum_{r=1}^{n} \eta_{rn} v_{rn}^{-1} - a_k^2 \sum_{r=1}^{n} \left(\frac{t}{n} \right)^2 (2v_{rn})^{-1} \right]}
\end{aligned}
$$

By definition of the stochastic integral ([6, Chap. IX, §2]) and the continuity of $\beta(s)^{-2}$, $0 \leqq s \leqq t$, there follows

$$(68) \qquad \underset{n \to \infty}{\text{l.i.m.}} \left(\frac{t}{n} \right) \sum_{r=1}^{n} \eta_{rn} v_{rn}^{-1} = \int_0^t \beta(s)^{-2} \, dy(s)$$

and

$$(69) \qquad \lim_{n \to \infty} \sum_{r=1}^{n} \left(\frac{t}{n} \right)^2 v_{rn}^{-1} = \int_0^t \beta(s)^{-2} \, ds.$$

From (67)–(69) it follows that $\hat{p}_j(t)$ coincides with the expression given by (5). It is clear that the same result is obtained with any sequence

$\{s_{rn}\}$ such that $0 = s_{0n} < s_{1n} < \cdots < s_{nn} = t \; (n = 1, 2, \cdots)$, and

$$\max_{1 \leqq r \leqq n} (s_{rn} - s_{r-1,n}) \to 0 \quad \text{as} \quad n \to \infty.$$

Let $\mathfrak{F}_t$ be the smallest Borel field[9] with respect to which the random variables $y(s)$, $0 \leqq s \leqq t$, are measurable. To see that $\hat{p}_j(t) = p_j(t)$, note first that $\hat{p}_j(t)$ is certainly measurable relative to $\mathfrak{F}_t$. Now suppose that $0 \leqq s_\nu \leqq t$ and that A_ν is a Borel subset of R^1, $\nu = 1, \cdots, N$. If Λ is the event $[y(s_\nu) \in A_\nu, \; \nu = 1, \cdots, N]$ then by adjoining the s_ν's to each of the sets $(s_{1n}, \cdots, s_{nn})$, $n = 1, 2, \cdots$, and including the $y(s_\nu)$ as conditioning variables in (66), we can write

$$(70) \qquad \int_\Lambda \hat{p}_j^{(n)}(t) \, dP = P\{\Lambda[x = a_j]\}, \qquad n = 1, 2, \cdots.$$

Since $\hat{p}_j^{(n)} \to \hat{p}_j(t)$ in the mean we obtain, on letting $n \to \infty$,

$$(71) \qquad \int_\Lambda \hat{p}_j(t) \, dP = P\{\Lambda[x = a_j]\}.$$

Since (71) holds for every Λ of the form described, it holds for every Λ in the Borel field $\mathfrak{F}_t$ (cf. [6, Chap. I, §7]). That is, we have verified that $\hat{p}_j(t)$ has the defining properties of the conditional probability $p_j(t)$.

Appendix 3. Properties of the $p(t)$ process (5). A simple computation from (5) shows that, for $0 < \tau < t$,

$$(72) \quad p_j(t) = \frac{p_j(\tau) \exp\left[a_j \int_\tau^t \beta(s)^{-2} \, dy(s) - \frac{1}{2} a_j^2 \int_\tau^t \beta(s)^{-2} \, ds \right]}{\sum_{k=1}^K p_k(\tau) \exp\left[a_k \int_\tau^t \beta(s)^{-2} \, dy(s) - \frac{1}{2} a_k^2 \int_\tau^t \beta(s)^{-2} \, ds \right]}.$$

Consider the joint process $\{x, p(t), t \geqq 0\}$ where x is regarded as a fixed random variable with distribution $\{p_j(0)\}$. From (72), the vector $p(t)$ depends only on x, $p(\tau)$, and the $w(s)$ increments for $\tau < s < t$. The latter increments are independent of $p(\tau)$, and of x and the $w(s)$ increments for $0 < s < \tau$, on which $p(\tau)$ depends. It follows that the conditional distribution of x, $p(t)$ given x, $p(s)$, $0 \leqq s \leqq \tau$, is a function of x, $p(\tau)$ alone; that is, the process $\{x, p(t), t \geqq 0\}$ is Markov.

The stochastic differential equation (7) for the $p(t)$ process can be established by applying either a representation theorem of Doob [6, Chap. VI, Theorem 3.3] or a related theorem of Dynkin [7, Theorem 7.2]. In either case the theorem mentioned must be extended slightly to take account of the fact that only the $p(t)$-component of the joint $\{x, p(t)\}$

[9] See footnote 3 in §2.

process is of diffusion type (alternatively the constant component x can be regarded as a trivial diffusion process). We shall apply Dynkin's theorem, extended to the present case. From (5) we see that $p_j(t)$ is of the form

$$(73) \qquad p_j(t) = \phi_j[t, z(t)],$$

where

$$(74) \qquad \begin{aligned} z(t) &= \int_0^t \beta(s)^{-2} \, dy(s) \\ &= \int_0^t x\beta(s)^{-2} \, ds + \int_0^t \beta(s)^{-1} \, dw(s) \end{aligned}$$

and

$$(75) \qquad \phi_j(t, z) = \frac{p_j(0) \exp\left[a_j z - \frac{1}{2} a_j^{\,2} \int_0^t \beta(s)^{-2} \, ds \right]}{\sum_{k=1}^K p_k(0) \exp\left[a_k z - \frac{1}{2} a_k^{\,2} \int_0^t \beta(s)^{-2} \, ds \right]}.$$

Since $\phi_j(t, z)$ is twice continuously differentiable in (t, z), there follows (by Appendix 1 or [7, Theorem 7.2])

$$(76) \quad p_j(t) - p_j(0) = \int_0^t m_j[s, x, p(s)] \, ds + \int_0^t \sigma_j[s, x, p(s)] \, dw(s).$$

The functions m_j, σ_j are given by

$$(77) \quad m_j(t, x, p) = \frac{\partial}{\partial t} \phi_j(t, z) + x\beta(t)^{-2} \frac{\partial}{\partial z} \phi_j(t, z) + \frac{1}{2} \beta(t)^{-2} \frac{\partial^2}{\partial z^2} \phi_j(t, z)$$

and

$$(78) \qquad \sigma_j(t, x, p) = \beta(t)^{-1} \frac{\partial}{\partial z} \phi_j(t, z).$$

The probabilistic meaning of m_j, σ_j is expressed by (7) and (8). The expressions (9) and (10) are computed directly from (75), (77) and (78). Finally, the stochastic differential equation (12) is equivalent (by definition) to the integral equation (76).

Appendix 4. Derivation of (21).
1. We first evaluate $p_j(t)$. To simplify the writing of certain conditional expectations it is convenient to adjoin to the probability space of the $\{x(t), w(t)\}$ process, a "dummy" step process $\{\tilde{x}(t), t \geqq 0\}$, defined to have the same range, initial distribution and transition probabilities as the $x(t)$ process, but independent of $\{x(t)\}$ and $\{w(t)\}$.

Now let $s_{rn} = rt/n$ $(r = 0, 1, \cdots, n; n = 1, 2, \cdots)$ and put

$$\eta_{rn} = y(s_{rn}) - y(s_{r-1,n}),$$

(79)
$$\xi_{rn} = \int_{(r-1)t/n}^{rt/n} x(s)\, ds,$$

$$\tilde{\xi}_{rn} = \int_{(r-1)t/n}^{rt/n} \tilde{x}(s)\, ds.$$

By (19),

$$\eta_{rn} = \xi_{rn} + \int_{(r-1)t/n}^{rt/n} \beta(s)\, dw(s);$$

and for each fixed n the random variables $\eta_{rn} - \xi_{rn}$ $(r = 1, \cdots, n)$ are independent and Gaussian with mean 0 and variance

(80)
$$v_{rn} = \int_{(r-1)t/n}^{rt/n} \beta(s)^2\, ds.$$

Using this fact we can write

$$p_j^{(n)}(t) = P\{x(t) = a_j \mid y(s_{rn}), r = 0, 1, \cdots, n\}$$

$$= P\{x(t) = a_j \mid \eta_{rn}, r = 1, \cdots, n\}$$

(81)
$$= \frac{\displaystyle\sum_{i=1}^{K} p_i(0)p_{ij}(t) \cdot E\left\{\exp\left[-\sum_{r=1}^{n} (c_{rn} - \xi_{rn})^2 (2v_{rn})^{-1}\right] \middle| x_0 = a_i, x_t = a_j\right\}_{c_{rn}=\eta_{rn}}}{\displaystyle\sum_{k=1}^{K}\sum_{i=1}^{K} p_i(0)p_{ik}(t) \cdot E\left\{\exp\left[-\sum_{r=1}^{n} (c_{rn} - \xi_{rn})^2 (2v_{rn})^{-1}\right] \middle| x_0 = a_i, x_t = a_k\right\}_{c_{rn}=\eta_{rn}}}.$$

In (81) the c_{rn} are arbitrary real numbers and the conditional expectation is regarded as a function of the c_{rn} evaluated at the (random) argument $c_{rn} = \eta_{rn}$ $(r = 1, \cdots, n)$. The last line of (81) follows by Bayes' rule and the fact that the conditional expectations thus evaluated are simply conditional densities of the η_{rn}. From now on a normalizing factor will be denoted by the generic symbol N. Then

(82)
$$p_j^{(n)}(t) = N \sum_{i=1}^{K} p_i(0)p_{ij}(t) \cdot E\left\{\exp\left[\sum_{r=1}^{n} c_{rn}\xi_{rn}v_{rn}^{-1} - \frac{1}{2}\sum_{r=1}^{n} \xi_{rn}^2 v_{rn}^{-1}\right] \middle| x_0 = a_i, x_t = a_j\right\}_{c_{rn}=\eta_{rn}}.$$

By our assumptions on the $\tilde{x}(t)$ process, (82) can also be written

$$
(83) \quad
\begin{aligned}
p_j^{(n)}(t) &= N \sum_{i=1}^{K} p_i(0) p_{ij}(t) \\
&\cdot E\left\{ \exp\left[\sum_{r=1}^{n} \eta_{rn} \xi_{rn} v_{rn}^{-1} - \frac{1}{2} \sum_{r=1}^{n} \xi_{rn}^2 v_{rn}^{-1} \right] \Big| \tilde{x}_0 = a_i, \tilde{x}_t = a_j, \eta_{1n}, \cdots, \eta_{nn} \right\}.
\end{aligned}
$$

Since almost every $\tilde{x}(t)$ sample function is a step function, the limit

$$
(84) \quad \lim_{n \to \infty} \sum_{r=1}^{n} \xi_{rn}^2 v_{rn}^{-1} = \int_0^t \beta(s)^{-2} \tilde{x}(s)^2 \, ds
$$

exists with probability 1 and hence, by bounded convergence, in mean square. Further

$$
(85) \quad
\begin{aligned}
\sum_{r=1}^{n} \eta_{rn} \xi_{rn} v_{rn}^{-1} &= \sum_{r=1}^{n} \left[\xi_{rn} \tilde{\xi}_{rn} + \xi_{rn} \int_{(r-1)t/n}^{rt/n} \beta(s) \, dw(s) \right] v_{rn}^{-1} \\
&\to \int_0^t \beta(s)^{-2} x(s) \tilde{x}(s) \, ds \\
&\qquad\qquad + \int_0^t \beta(s)^{-1} \tilde{x}(s) \, dw(s) \qquad (n \to \infty) \\
&= \int_0^t \beta(s)^{-2} \tilde{x}(s) \, dy(s),
\end{aligned}
$$

where the integral is again a limit in mean square ([6, Chap. IX, §2]). Let $\theta_n(t)$ be the random variable in the brackets in (83) and put

$$
(86) \quad \theta(t) = \int_0^t \beta(s)^{-2} \tilde{x}(s) \, dy(s) - \frac{1}{2} \int_0^t \beta(s)^{-2} \tilde{x}(s)^2 \, ds.
$$

From (84) and (85), l.i.m. $\theta_n(t) = \theta(t)$. Furthermore

$$
\begin{aligned}
| e^{\theta_n(t)} - e^{\theta(t)} | &\leq \tfrac{1}{2} | \theta_n(t) - \theta(t) | [e^{\theta_n(t)} + e^{\theta(t)}] \\
&\leq | \theta_n(t) - \theta(t) | e^{\lambda |w(t)| + \mu t}
\end{aligned}
$$

for some constants λ, $\mu > 0$; and since $E\{[e^{\lambda |w(t)|}]^2\} < \infty$,

$$
(87) \quad \underset{n \to \infty}{\text{l.i.m.}} \; e^{\theta_n(t)} = e^{\theta(t)}.
$$

Denote by $\mathfrak{F}_t^n$ (resp. $\mathfrak{F}_t$) the smallest Borel field[10] relative to which the random variables $\tilde{x}(0)$, $\tilde{x}(t)$, η_{1n}, $\cdots$, η_{nn}, (resp. $\tilde{x}(0)$, $\tilde{x}(t)$, $y(s)$, $0 \leq s \leq t$) are measurable. Now

$$
(88) \quad
\begin{aligned}
&E\{e^{\theta_n(t)} \mid \mathfrak{F}_t^n\} - E\{e^{\theta(t)} \mid \mathfrak{F}_t\} \\
&\quad = E\{e^{\theta_n(t)} \mid \mathfrak{F}_t^n\} - E\{e^{\theta_n(t)} \mid \mathfrak{F}_t\} + E\{e^{\theta_n(t)} \mid \mathfrak{F}_t\} - E\{e^{\theta(t)} \mid \mathfrak{F}_t\}.
\end{aligned}
$$

---

[10] See footnote 3 in §2.

Since $\mathfrak{F}_t{}^n \subset \mathfrak{F}_t$ and since, by inspection of $\theta_n(t)$, the random variable $E\{e^{\theta_n(t)} \mid \mathfrak{F}_t\}$ is measurable relative to $\mathfrak{F}_t{}^n$, there follows ([6, Chap. I, Theorem (8.1)])

$$(89) \qquad E\{e^{\theta_n(t)} \mid \mathfrak{F}_t{}^n\} - E\{e^{\theta_n(t)} \mid \mathfrak{F}_t\} = 0$$

with probability 1. Finally, (87) implies

$$(90) \qquad \underset{n \to \infty}{\mathrm{l.i.m.}}\ E\{e^{\theta_n(t)} \mid \mathfrak{F}_t\} - E\{e^{\theta(t)} \mid \mathfrak{F}_t\} = 0.$$

Thus we have shown that

$$(91) \qquad \underset{n \to \infty}{\mathrm{l.i.m.}}\ p_j{}^{(n)}(t) = N \sum_{i=1}^{K} p_i(0) p_{ij}(t)$$
$$\cdot E\{e^{\theta(t)} \mid \tilde{x}(0) = a_i, \tilde{x}(t) = a_j; y(s), 0 \leqq s \leqq t\}.$$

By exactly the same argument as in Appendix 2 the left side of (91) can be identified with $p_j(t)$.

2. Next we show that the process $\{x(t), p(t), t \geqq 0\}$ is Markov. For $0 \leqq \tau \leqq t$ write

$$(92) \quad \phi(\tau, t) = \exp\left[\int_\tau^t \beta(s)^{-2}\tilde{x}(s)\, dy(s) - \frac{1}{2}\int_\tau^t \beta(s)^{-2}\tilde{x}(s)^2\, ds\right],$$

and let $\mathfrak{K}_\tau{}^t$ be the smallest Borel field relative to which the random variables $y(s)$, $\tau \leqq s \leqq t$, are measurable. Thus $\phi(\tau, t)$ is measurable relative to the Borel field generated by $\mathfrak{K}_\tau{}^t$ and the $\tilde{x}(s)$'s for $\tau \leqq s \leqq t$. With this notation,

$$(93) \quad p_j(t) = N \sum_{i=1}^{K} p_i(0) p_{ij}(t) E\{\phi(0, t) \mid \tilde{x}_0 = a_i, \tilde{x}_t = a_j, \mathfrak{K}_0{}^t\}.$$

Now

$$p_j(t + h)$$
$$= N \sum_{k=1}^{K} p_k(0) p_{kj}(t + h) E\{\phi(0, t + h) \mid \tilde{x}_0 = a_k, \tilde{x}_{t+h} = a_j, \mathfrak{K}_0{}^{t+h}\}$$
$$(94) \qquad = N \sum_{k=1}^{K} p_k(0) p_{kj}(t + h) \sum_{i=1}^{K} P\{\tilde{x}_t = a_i \mid \tilde{x}_0 = a_k, \tilde{x}_{t+h} = a_j\}$$
$$\cdot E\{\phi(0, t)\phi(t, t + h) \mid \tilde{x}_0 = a_k, \tilde{x}_t = a_i, \tilde{x}_{t+h} = a_j, \mathfrak{K}_0{}^{t+h}\}$$
$$= N \sum_{i=1}^{K} \sum_{k=1}^{K} p_k(0) p_{ki}(t) p_{ij}(h) E\{\phi(0, t) \mid \tilde{x}_0 = a_k, \tilde{x}_t = a_i, \mathfrak{K}_0{}^t\}$$
$$\cdot E\{\phi(t, t + h) \mid \tilde{x}_t = a_i, \tilde{x}_{t+h} = a_j, \mathfrak{K}_t{}^{t+h}\},$$

where we have used the fact that the $\tilde{x}(t)$ process is Markov and is inde-

pendent of the $y(t)$ process. Comparing (93) and (94) we obtain

$$
(95) \quad p_j(t + h) = N \sum_{i=1}^{K} p_i(t) p_{ij}(h)
$$
$$
\cdot E\{\phi(t, t + h) \mid \tilde{x}_t = a_i, \tilde{x}_{t+h} = a_j, \mathfrak{IC}_t^{\,t+h}\}.
$$

Write $p(t) = [p_1(t), \cdots, p_K(t)]$ and consider the joint process $\{x(t), p(t), t \geqq 0\}$. Equations (92) and (95) show that $x(t + h), p(t + h)$ are fully determined by $x(t)$, $p(t)$ and the $w(s)$ increments for $t \leqq s \leqq t + h$. Reasoning as in Appendix 2 we conclude that the joint process $\{x(t), p(t), t \geqq 0\}$ is Markov.

3. We now evaluate functions m_j and b_{ij} defined by

$$
(96) \quad m_j(t, x, p) = \lim_{h \to 0} E\left\{\frac{p_j(t + h) - p_j(t)}{h} \,\middle|\, x(t) = x, p(t) = p\right\}
$$

and

$$
(97) \quad b_{ij}(t, x, p) = \lim_{h \to 0} E
$$
$$
\cdot \left\{\frac{[p_i(t + h) - p_i(t)][p_j(t + h) - p_j(t)]}{h} \,\middle|\, x(t) = x, p(t) = p\right\}.
$$

To simplify computation note that the conditional expectation in (95) is readily evaluated, if it is modified by including the extra condition that no jump of $\tilde{x}(\cdot)$ occurs in the interval $(t, t + h)$; and the conditional probability that no jump occurs, given $\tilde{x}_t = \tilde{x}_{t+h} = a_i$, is

$$
(98) \quad \frac{e^{-\nu_i h}}{p_{ii}(h)} = 1 + O(h^2) \qquad\qquad (h \to 0).
$$

Also, since $\beta(s)^{-1}$ is assumed bounded, we have

$$
(99) \quad \phi(t, t + h) \leqq f(h) e^{\lambda|\Delta y|},
$$

where $\Delta y = y(t + h) - y(t)$, $\lambda > 0$ is constant, and $f(h)$ is bounded as $h \to 0$. Put

$$
(100) \quad \theta_j = a_j \int_t^{t+h} \beta(s)^{-2}\, dy(s) - \frac{1}{2} a_j^2 \int_t^{t+h} \beta(s)^{-2}\, ds.
$$

From (92) and (98)–(100),

$$
(101) \quad E\{\varphi(t, t + h) \mid \tilde{x}_t = a_i, \tilde{x}_{t+h} = a_i, \mathfrak{IC}_t^{\,t+h}\} = e^{\theta_i} + h^2 \omega_i(h),
$$

and if $j \neq i$,

$$
(102) \quad E\{\phi(t, t + h) \mid \tilde{x}_t = a_i, \tilde{x}_{t+h} = a_j, \mathfrak{IC}_t^{\,t+h}\} = 1 + h\omega_{ij}(h),
$$

where (for a suitable $f(\cdot)$)

$$(103) \qquad 0 \leq \omega_i(h), \quad \omega_{ij}(h) \leq f(h)e^{\lambda|\Delta y|}.$$

Thus

$$p_j(t + h) = N[p_j(t)p_{jj}(h)\{e^{\theta_j} + h^2\omega_j(h)\}$$

$$(104) \qquad\qquad + \sum_{\substack{i=1 \\ i \neq j}}^{K} p_i(t)p_{ij}(h)\{1 + h\omega_{ij}(h)\}].$$

A simple calculation from (100) yields

$$(105) \qquad \lim_{h \to 0} h^{-1}E\{\theta_j \mid x(t) = x\} = a_j x\beta(t)^{-2} - \tfrac{1}{2}a_j^2\beta(t)^{-2},$$

$$(106) \qquad \lim_{h \to 0} h^{-1}E\{\theta_i\theta_j \mid x(t) = x\} = a_i a_j\beta(t)^{-2}.$$

We note that in (104), N^{-1} is the sum over j of the terms in brackets. Writing out (104) explicitly, subtracting $p_j(t)$ and then using (105), (106), we can compute the limits (96) and (97). The results are

$$(107) \quad m_j(t, x, p) = -\nu_j p_j + \sum_{\substack{i=1 \\ i \neq j}}^{K} \nu_{ij} p_i + \beta(t)^{-2}(x - \bar{x})(a_j - \bar{x})p_j$$

and

$$(108) \quad b_{ij}(t, x, p) = [\beta(t)^{-1}(a_i - \bar{x})p_i][\beta(t)^{-1}(a_j - \bar{x})p_j],$$

where

$$(109) \qquad\qquad \bar{x} = \sum_{i=1}^{K} a_i p_i.$$

4. Define

$$(110) \qquad \sigma_j(t, x, p) = \beta(t)^{-1}(a_j - \bar{x})p_j, \qquad j = 1, \cdots, K.$$

It will be shown finally that the $p(t)$ process can be represented as the solution of the stochastic differential system

$$dp_j(t) = m_j[t, x(t), p(t)]\, dt$$

$$(111) \qquad\qquad + \sigma_j[t, x(t), p(t)]\, dw(t), \quad t \geq 0, j = 1, \cdots, K,$$

where $\{w(t),\ t \geq 0\}$ is the Wiener process introduced in (19). We shall apply a representation theorem of Doob [6, Chap. VI, Theorem 3.3], generalized to allow for the fact that only the $p(t)$ component of the $\{x(t), p(t)\}$ process is continuous (that almost every $p(t)$ sample function is continuous follows from (104)).

The $p(t)$ process is obviously bounded; hence by (107), (110) the conditions usually imposed on m and σ [6, Chap. VI, §3] are satisfied. Now let $\mathcal{G}_t$ be the smallest Borel field with respect to which the random variables $x(s)$, $p(s)$, $0 \leqq s \leqq t$ are measurable. Then, since the $\{x, p\}$ process is Markov, the evaluations (107), (108) are unchanged if the conditional expectations in (96), (97) are defined relative to $\mathcal{G}_t$. Reasoning as in the proof[11] of [6, Chap. VI, Theorem 3.3], we conclude that each process

$$
(112) \quad \left\{ p_j(t) - p_j(0) - \int_0^t m_j[s, x(s), p(s)] \, ds, \mathcal{G}_t \, ; \, t \geqq 0 \right\},
$$

$$
j = 1, \cdots, K,
$$

is a martingale which satisfies the conditions of [6, Chap. IX, Theorem 5.3]. That is, if $\{\tilde{p}_j(t), \mathcal{G}_t \, ; \, t \geqq 0\}$ is the martingale defined by (112), if $\{\sigma_j\}^{-1} = \sigma_j^{-1}$ where $\sigma_j \neq 0$, and if $\{\sigma_j\}^{-1} = 0$ where[12] $\sigma_j = 0$, then the equation

$$
(113) \quad \hat{w}_j(t) = \int_0^t \{\sigma_j[s, x(s), p(s)]\}^{-1} \, d\tilde{p}(s)
$$

defines a Wiener process $\hat{w}_j(t)$, $t \geqq 0$.

It remains to identify each $\hat{w}_j(t)$ process with the $w(t)$ process of (19). Let

$$
(114) \quad \begin{aligned} \tilde{y}(t) &= y(t) - \int_0^t x(s) \, ds \\ &= \int_0^t \beta(s) \, dw(s). \end{aligned}
$$

Using (104)–(106) and (110) we find that

$$
(115) \quad \begin{aligned} \lim_{h \to 0} h^{-1} E\{ [\tilde{y}(t+h) - \tilde{y}(t)][p_j(t+h) - p_j(t)] \mid \mathcal{G}_t \} \\ = \beta(t)\sigma_j[t, x(t), p(t)]. \end{aligned}
$$

Reasoning as before we conclude that the process $\{\tilde{y}(t), \mathcal{G}_t \, ; \, t \geqq 0\}$ is a martingale. Since

$$
w(t) = \int_0^t \beta(s)^{-1} \, d\tilde{y}(s),
$$

it follows by (113) and (115) that

$$
(116) \quad E\{ [w(t) - w(s)][\hat{w}_j(t) - \hat{w}_j(s)] \mid \mathcal{G}_s \} = t - s, \qquad 0 < s < t.
$$

[11] The Borel field $\mathcal{G}_t$ plays the role of the $\mathcal{F}_t$ of Doob's theorem.

[12] σ_j cannot vanish on a t-interval. Otherwise $\bar{x}(t)$ is constant on this interval, which can be shown to imply that the posterior variance of $x(t)$ vanishes, hence that all but one $p_k(t)$ vanishes, in contradiction to (86) and (91).

Hence for each $t > 0$, $\hat{w}_j(t) = w(t)$ with probability 1; by continuity this implies that the processes $\hat{w}_j(t)$, $w(t)$ are essentially identical. The integrated form of (21) now follows from (113) by inversion.

REFERENCES

[1] R. L. STRATONOVIČ, *Conditional Markov processes*, Theor. Probability Appl., 5 (1960), pp. 156–178.

[2] G. E. KOLOSOV AND R. L. STRATONOVIČ, *A problem of synthesis of an optimal regulator by methods of dynamic programming*, Avtomat. i Telemeh., 24 (1963), pp. 1165–1173.

[3] W. M. WONHAM, *Stochastic problems in optimal control*, RIAS TR 63-14, 1963.

[4] R. E. KALMAN AND W. M. WONHAM, *Investigation of filter functions*, Final TR, Contract No. DA-36-034-ORD-3708Z, U.S. Army Ordnance Missile Command, Redstone Arsenal, Alabama, 1964.

[5] H. J. KUSHNER, *On the differential equations satisfied by conditional probability densities of Markov processes, with applications*, this Journal, 2 (1964), pp. 106–119.

[6] J. L. DOOB, *Stochastic Processes*, John Wiley, New York, 1953.

[7] E. B. DYNKIN, *Markovskie Processy*, Fizmatgiz, Moscow, 1963.

[8] K. ITO, *On Stochastic Differential Equations*, Mem. Amer. Math. Soc., 4, 1951.

[9]. I. I. GIHMAN, *On the theory of differential equations of stochastic processes, I.* Ukrain. Mat. Z., 2 (1950), pp. 37–63.

[10] J. F. BARRETT, *Application of Kolmogorov's equations to randomly disturbed automatic control systems*, Proc. IFAC, Automatic and Remote Control, vol. 2, Butterworth, London, 1961, pp. 724–733.

[11] S. O. RICE, *Mathematical analysis of random noise*, Bell System Tech. J., 23, 24 (1944); reprinted in Noise and Stochastic Processes, N. Wax, ed., Dover, New York, 1954.