The Annals of Probability
1985, Vol. 13, No. 4, 1033–1107

# SPECIAL INVITED PAPER

## WHITE NOISE CALCULUS AND NONLINEAR FILTERING THEORY[1]

BY G. KALLIANPUR AND R. L. KARANDIKAR

*University of North Carolina, Chapel Hill*

A self-contained outline of recent work on the white noise approach to nonlinear filtering is given together with the necessary background of white noise calculus.

**1. Introduction.** An informal description of the filtering problem is the following: Let the unobserved signal process $X = (X_u)$ be a stochastic (usually assumed to be Markov) process taking values in $R^d$. Information concerning $X$ is provided by a process $y = (y_t)$ which is continuously observed over time

$$y_t = h_t(X_t) + n_t, \qquad 0 \le t \le T, \tag{1.1}$$

where $h_t$ is a known function and $n = (n_t)$ is a noise process. The problem is to obtain the "best" estimate of $X_t$ given by the set of observations $\{y_s, 0 \le s \le t\}$ or, equivalently, to find the conditional distribution of $X_t$ given $\{y_s, 0 \le s \le t\}$.

A rigorous formulation of (1.1) requires the choice of a mathematical model for the noise process. The latter is an extremely important question when continuous time processes are involved. The usual procedure is to regard the noise $n_t$, heuristically, as the "derivative" $\dot{W}_t$ of the Wiener process and to replace (1.1) by its "integrated" version

$$Y_t = \int_0^t h_s(X_s)\,ds + W_t, \qquad 0 \le t \le T. \tag{1.2}$$

In the martingale approach to filtering theory (which we will also refer to as the stochastic calculus or conventional theory) (1.2) is taken to be the canonical observation model and it is assumed that $X = (X_t)$, $W = (W_t)$ are defined on a probability space $(\Omega, \mathscr{A}, \Pi)$, $h$: $[0,T] \times \mathbf{R}^d \to \mathbf{R}^m$ is a measurable function satisfying

$$\int_0^T |h_u(X_u)|^2\,du < \infty, \qquad \Pi\text{-a.s.}, \tag{1.3}$$

and $W$ is a standard $m$-dimensional Wiener process.

Martingale calculus and the theory of Itô stochastic differential equations (SDEs) have been applied with spectacular success in solving the problem of nonlinear filtering based on (1.2). The first phase of this development culminating

Received July 1984; revised March 1985.

[1] Research supported by AFOSR Contract No. F49620 82 C0009.

*AMS* 1980 *subject classifications.* Primary 60A05, 60G35; secondary 60H15.

*Key words and phrases.* Finitely additive white noise, optimal filter, lifting, quasicylindrical mapping, conditional expectation, Bayes formula, robustness.

in the derivation of the most general form of the SDE for the optimal filter was essentially given in the papers of Mortenson (1966), Kushner (1967), Kallianpur and Striebel (1969), Zakai (1969) and Fujisaki *et al.* [FKK] (1972).

In the important case when the signal process is an $\mathbb{R}^d$-valued diffusion most recent work has led to the study of the existence of unique solutions of stochastic partial differential equations (SPDEs) of Itô type satisfied by the unnormalized conditional density of the filtering problem [Rozovskii (1975), Krylov–Rozovskii (1978, 1981) and Pardoux (1979, 1982)].

Despite its elegance, the power of its mathematical techniques, and the enormous stimulus it has provided to the theory of SDEs, the practical validity of the stochastic calculus theory of nonlinear filtering is open to serious criticism. In a series of papers in which he was the first to advocate a finitely additive (f.a.) white noise approach, Balakrishnan has expressed the view (shared by some engineers) that the model (1.2) is not suitable for applications because the results obtained cannot be instrumented [Balakrishnan (1977, 1980)]. While this objection to the role of the Wiener process in physical problems may not be new, this criticism has stimulated a search for a rigorous theoretical framework for nonlinear filtering that is faithful to the observed phenomena.

The purpose of this article is to present a connected and self-contained account of the work we have done over the last two years in constructing a rigorous white noise calculus and applying it to develop nonlinear filtering theory.

Two things should be pointed out in connection with our theory:

(i) The natural space of observations and of the noise is a Hilbert space (or a suitable subspace of it) of Wiener measure zero. An insistence on countably additive probabilistic techniques would require an enlargement of this space either to a Wiener space of continuous functions $C([0,T];\mathbb{R}^m)$ or to an even larger space such as $\mathscr{S}'(\mathbb{R}^d)$, the Schwarz space of tempered distributions. (The latter is an interesting possibility yet to be fully explored.) The alternative is to resort to a finitely additive white noise measure on Hilbert space.

(ii) The role of the Wiener process as a model for noise is customarily that given in (1.2). However, Wiener himself, in a conversation with one of the authors several years ago at the Indian Statistical Institute, stated that $n(f)$ given by the stochastic integral $\int_0^1 f(t)\,dW_t$ can be used to define Gaussian white noise over the Hilbert space $L^2[0,1]$. The latter is indeed an example of a (finitely additive) cylinder measure or weak distribution used in later work by I. E. Segal (1956) and by L. Gross (1960, 1962). It is the model adopted by us in the definition of the abstract statistical model to be given in Section 4.

The plan of the paper is as follows: We begin by giving, in Section 2, a concise description of the main results of nonlinear filtering theory based on stochastic calculus. For more details see Part I of our survey paper [Kallianpur and Karandikar (1983b)]. Other and more comprehensive accounts of this theory are available in books. (See the references in Section 2.)

Section 3 and part of Section 4 contain preparatory material on the theory of f.a. quasicylindrical probability measures (QCPs), for which we cannot find any reference in the literature. The notion of a lifting with respect to a QCP is, of course, based on ideas of Segal and Gross but it has been found necessary to develop these further for our purpose. The definition of absolute continuity for finitely additive measures and the version of the Radon–Nikodym theorem given in the book of Dunford and Schwartz (1958) are not suitable from our point of view. The same is true of the definitions given in Gross's paper [Gross (1962)]. We have introduced new—and for us, the right—definitions of all these concepts. The development presented is entirely self-contained and culminates in the abstract nonlinear filtering model and the finitely additive version of the Kallianpur–Striebel Bayes formula.

In Section 5, the white noise versions of the Zakai and FKK equations are obtained. We then specialize to the case when the signal is an $\mathbb{R}^d$-valued diffusion process. When the unnormalized conditional density exists it is shown to satisfy a partial differential equation (PDE) which we call the Zakai equation for the unnormalized density. The other main results of this section (Theorems 5.5 and 5.6) establish the existence of a unique solution of the Zakai PDE under conditions which do not assume the boundedness of $h$. As mentioned earlier, the Zakai and FKK equations as well as the PDE are *not* stochastic differential equations but "ordinary" differential (or partial differential) equations in which the observed path occurs as a parameter in the coefficients.

In Section 6 we study the f.a. white noise theory in a more general framework that includes applications to signal and observation processes taking values in *infinite-dimensional* separable Hilbert spaces. In the model (1.2) the state space of $X$ is assumed to be a complete separable, metric (Polish) space $S$, $e$ is $\mathscr{K}$-valued white noise, where $\mathscr{K}$ is possibly infinite-dimensional Hilbert space, and $h$: $[0,T]\times S\to\mathscr{K}$ is a measurable function such that $E\int_0^T\|h_s\|_{\mathscr{K}}^2\,ds<\infty$. The chief difficulty here is that we have no conditional density since there is no Lebesgue measure (or any natural measure) in Hilbert space. Instead of the PDEs of the preceding section we obtain f.a. analogues of measure-valued equations of FKK, Zakai, and Kunita types. The equations (of which the first two are differential equations) are derived, and the equivalence and the uniqueness of solutions of these equations are established in Theorems 6.1–6.4. An approximation procedure is also set forth involving convergence in variation norm.

It seems worth remarking that the white noise calculus approach is seen to its best advantage in its treatment of problems where infinite-dimensional signal processes are encountered. This is an important area of application to filtering and prediction of random fields (though a great deal still remains to be done). At present, the stochastic calculus treatment has not gone beyond deriving the SDEs for the optimal filter [see, e.g., Korezlioglu and Martias (1984) and the references therein]. The difficulties in the way of proving existence of unique solutions seem to be formidable.

Section 7 is a brief digression devoted to likelihood ratios in the f.a. framework, a topic whose applications to statistical problems will be considered elsewhere. Formulas for likelihood ratios for random fields are also given, since in principle, the white noise theory poses no special difficulties for multiparameter processes.

There have been attempts in recent years to modify the stochastic calculus filtering theory so as to bring it closer to applications. The work on pathwise or "robust" solutions to the SDEs of the filtering problems [Clark (1978), Davis (1979, 1980), Pardoux (1979), (1982)] as well as the investigations of the approximations to Itô SDEs and their solutions [discussed in Ikeda and Watanabe (1981)] may be viewed in this light. The white noise approach is entirely different in spirit in that it avoids the complexities inherent in the stochastic calculus treatment of the subject. Our lack of familiarity with f.a. measure theory necessitates a careful definition of concepts well known in countably additive probability theory: conditional expectations, absolute continuity, Radon–Nikodym derivatives, change of variables formula, etc. Once this is done, however, the simplicity and the advantages of the new approach are apparent.

The question that naturally arises, however, is whether the white noise calculus gives the same results as the martingale-theoretic approach to the subject. This important point is discussed in some detail in Section 8 where we show that our theory is consistent with the conventional theory. In fact, the robust filter of the latter can be recovered by using the theorems of Section 8. In Section 9 we list some of the open problems on which we and some of our colleagues are working.

**2. A brief survey of nonlinear filtering theory based on stochastic calculus.** The stochastic calculus approach to nonlinear filtering theory takes the model (1.2) as its starting point. Since detailed expositions of the theory are now available in books and survey articles [Elliott (1982), Kallianpur (1980), Liptser and Shiryaev (1977), Kunita (1983)], we give here only a highly condensed account of the principal results concerning the derivation of the stochastic differential equation (SDE) of the optimal filter and on the existence and uniqueness of the solution. Our object is to give the reader a flavour of the martingale approach and to enable him to contrast it with the theory described in the later sections.

In the model (1.2)

$$Y_t = \int_0^t h_u\,du + W_t, \qquad 0 \le t \le T,$$

$h_t$ is measurable wrt $\mathscr{F}_t^X$ where $\mathscr{F}_t^X$ is the $\sigma$ field generated by the family $\{X_u, 0 \le u \le t\}$ augmented by the inclusion of zero probability sets. Here $X = (X_t)$, called the signal process, is usually taken to be a Markov process with state space $\mathbf{R}^d$, though more general infinite dimensional spaces can be considered (see Sec. 6). The process $h = (h_t)$ is taken to be $R^m$-valued and contains information about $X = (X_t)$. Furthermore, it is assumed that

$$E\int |h_t|^2\,dt < \infty.$$

The noise in (1.2) is given by $W = (W_t)$, an $\mathbf{R}^m$-valued standard Wiener process. The most general assumptions on the noise and signal process are that for each $t$, the $\sigma$ fields $\sigma\{X_u, W_u, 0 \le u \le t\}$ and $\sigma\{W_v - W_u, t \le u < v \le T\}$ are independent. An important special case to be treated in greater detail later is the independence of signal and noise, i.e., of $\sigma\{X_t, 0 \le t \le T\}$ and $\sigma\{W_t, 0 \le t \le T\}$. [The processes $(X_t)$, $(h_t)$, and $(W_t)$ are defined on a countably additive probability space $(\Omega, \mathscr{A}, \Pi)$.]

Given the observations $\{Y_s, 0 \le s \le t\}$, the basic nonlinear filtering problem is to find the conditional distribution $\Pi\{X_t \in (\cdot)|\mathscr{F}_t^Y\}$ of the "present" state of the system. A crucial requirement is that the solution be a recursive estimate since in most applications the data are observed continuously in time so that it is important to be able to compute the conditional distribution at time $t + \delta t$ ($\delta t > 0$) using its value at time $t$. In other words, one seeks a SDE for the conditional distribution or, equivalently, for the conditional expectation $\Pi_t(f) = E[f(X_t)|\mathscr{F}_t^Y]$ for a wide enough class of functions $f$: $\mathbf{R}^d \to \mathbf{R}$. In most cases of interest, e.g., if $X$ is a diffusion process it is enough to consider $f \in C_b^2$.

The following simple result is the natural starting point of the martingale theoretic approach to the subject.

PROPOSITION 2.1. *Let $\xi_t = \xi_0 + A_t + M_t$, $t \in [0, T]$ be a d-dimensional semimartingale relative to $(\mathscr{G}_t)$. Assume the following conditions hold:*

(i) *$(A_t)$ is a right-continuous $(\mathscr{G}_t)$-adapted process of bounded variation in $[0, T]$ with $A_0 = 0$.*

(ii) $E|\text{Var}_T A|^2 < \infty$.

(iii) *$(M_t, \mathscr{G}_t)$ is a right-continuous $L^2$ martingale with $M_0 = 0$.*

(iv) $E|\xi_0|^2 < \infty$.

*Then $\Pi_t(\xi) = E(\xi_t|\mathscr{F}_t)$ is an $L^2$-semimartingale relative to $\mathscr{F}_t$, where $(\mathscr{F}_t)$ is a right-continuous family of $\sigma$ fields such that $(\mathscr{F}_t) \subseteq (\mathscr{G}_t)$ for all $t$. Further,*

$$\Pi_t(\xi) = \Pi_0(\xi_0) + \bar{A}_t + \bar{M}_t; \tag{2.1}$$

(2.2) *$\bar{A}_t$ is the dual predictable projection of $A_t$ relative to $(\mathscr{F}_t)$ with $\bar{A}_0 = 0$;*

$$E|\text{Var}_T \bar{A}|^2 < \infty; \tag{2.3}$$

(2.4) *$(\bar{M}_t, \mathscr{F}_t)$ is a right-continuous martingale with $\bar{M}_0 = 0$.*

Now suppose that the signal process is an $\mathbf{R}^d$-valued diffusion process with generator $\mathscr{L}$, given by

$$\mathscr{L}f = \frac{\partial f}{\partial t} + \mathscr{L}_t f$$

for $f \in C_0^{1,2}([0, T] \times R^d)$ and for $g \in C_0^2(R^d)$.

$$(\mathscr{L}_t g)(x) = \tfrac{1}{2} \sum_{i,j=1}^{d} a_{ij}(t, x) \frac{\partial^2 g}{\partial x^i \partial x^j}(x) + \sum_{i=1}^{d} b_i(t, x) \frac{\partial g}{\partial x^i}(x), \tag{2.5}$$

for suitable measurable functions $a_{ij}$, $b_i$ such that the matrix $((a_{ij}(t, x)))$ is nonnegative definite.

The process $(X_t)$ can be realized as a (weak) solution to the SDE

$$dX_t = b(t, X_t)\,dt + \sigma(t, X_t)\,dB_t, \tag{2.6}$$

where $B_t$ is an $\mathbb{R}^d$-valued Brownian motion and $\sigma$ is such that $a(t,x) = \sigma(t,x)\sigma^*(t,x)$.

For an integrable function $f$ on $(\Omega, \mathscr{A}, \Pi)$, introduce the notation

$$\hat{\Pi}_t(f) = E_{\Pi}(f \mid \mathscr{F}_t^Y)$$

and for a function $g\colon \mathbb{R}^d \to \mathbb{R}$ such that $E_{\Pi}|g(X_t)| < \infty$,

$$\Pi_t(g) = \hat{\Pi}_t(g(X_t)) = E_{\Pi}(g(X_t) \mid \mathscr{F}_t^Y).$$

By Itô's formula (or Lemma A.1), we have that for $g \in C_0^2(\mathbb{R}^d)$,

$$M_t = g(X_t) - \int_0^t (\mathscr{L}_s g)(X_s)\,ds \tag{2.7}$$

is an $\mathscr{F}_t^X$ martingale. In view of the assumption on the independence of $\sigma(X_u, W_u, u \le t)$ and $\sigma(W_u - W_t, t \le u \le T)$, this implies that $M_t$ is an $\mathscr{F}_t^{XY}$ martingale.

Applying Proposition 2.1 with $\xi_t = g(X_t)$, $A_t = \int_0^t (\mathscr{L}_s g)(X_s)\,ds$, $\mathscr{G}_t = \mathscr{F}_t^{XY}$, and $\mathscr{F}_t = \mathscr{F}_t^Y$, we get

$$\Pi_t(g) = \Pi_0(g) + \bar{A}_t + \bar{M}_t, \tag{2.8}$$

where $\bar{M}_t$ is an $\mathscr{F}_t^Y$ martingale. It can be shown that $\bar{A}_t = \int_0^t \Pi_s(\mathscr{L}_s g)\,ds$ so that

$$\Pi_t(g) = \Pi_0(g) + \int_0^t \Pi_s(\mathscr{L}_s g)\,ds + \bar{M}_t. \tag{2.9}$$

It will be seen from (2.9) that the problem of obtaining an Itô-type SDE for $\Pi_t(f)$ reduces to the problem of expressing $\bar{M}_t$ as a stochastic integral with respect to an $(\mathscr{F}_t^Y)$-adapted Wiener process. That such a Wiener process exists is shown by the following result.

**Proposition 2.2.**

(i) *The process $\nu_t$ defined by*

$$\nu_t = Y_t - \int_0^t \hat{\Pi}_s(h_s)\,ds \tag{2.10}$$

*is a Wiener martingale with respect to $(\mathscr{F}_t^Y)$;*

(ii) *For any $t$, $\sigma[\nu_v - \nu_t, t \le v \le T] \perp \mathscr{F}_t^Y$.*

See FKK (1972) for the proof. The Wiener martingale $(\nu_t, \mathscr{F}_t^Y)$ is called an *innovation process*.

Clearly from the definition $\mathscr{F}_t^\nu \subseteq \mathscr{F}_t^Y$. The equality of the $\sigma$ fields (for each $t$) holds in the case of linear (Kalman-Bucy) filtering [Kailath (1968)]. For nonlinear filtering, Allinger and Mitter (1981) extending an earlier result due to Clark (1969), have shown that equality holds if the processes $(h_t)$ and $W_t$ are independent and $E\int_0^T |h_s|^2\,ds < \infty$.

The significance of this result from the point of view of nonlinear filtering is obvious for the representation of the martingale $\bar{M}_t$ of (1.9) as a stochastic integral, with respect to $\nu_t$, is then an immediate consequence of the celebrated results of Cameron and Martin and of Itô [see Kallianpur (1980)]. However, a counterexample due to Tsirel'son shows that equality cannot hold in general. Although the problem is of independent interest, Fujisaki, Kallianpur, and Kunita have shown that the desired stochastic integral representation is available without having to prove first that $\mathscr{F}_t^{\nu} = \mathscr{F}_t^Y$.

PROPOSITION 2.3 [Fujisaki *et al.* (1972)]. *Let $(\bar{M}_t, \mathscr{F}_t^Y)$ be a right-continuous (or separable) $L^2$-martingale with $\bar{M}_0 = 0$. Then there is a process $\phi_t$ which is $\mathscr{F}_t^Y$-predictable such that*

$$\bar{M}_t = \int_0^t (\phi_s, d\nu_s) = \sum_{i=1}^m \int_0^t \phi_s^i \, d\nu_s^i, \tag{2.11}$$

*Furthermore,* $\int_0^T E|\phi_s|^2 \, ds < \infty$.

The process $\phi_s$ appearing in (2.11) is related to the processes $(X_t), (Y_t)$ as follows [see Kallianpur (1980)]. Let $\tilde{D}_s^{k,i}$ be the $\mathscr{F}_t^{X,Y}$-predictable process given by

$$\langle B^k, W^i \rangle_t = \int_0^t \tilde{D}_s^{k,i} \, ds \tag{2.12}$$

and let

$$(\tilde{D}_s^i g)(x) = \sum_{k=1}^d \sum_{j=1}^d \tilde{D}_s^{k,i} \sigma_{k,j}(s, x) \frac{\partial}{\partial x^j} g(x). \tag{2.13}$$

Then

$$\phi_s^i = \tilde{\Pi}_s(g(X_s) h_s^i) - \Pi_s(g) \tilde{\Pi}_s(h_s^i) + \tilde{\Pi}_s(\tilde{D}_s g(X_s)). \tag{2.14}$$

The basic SDE given in the next result is a consequence of the above arguments and Proposition 2.3.

THEOREM 2.4 [FKK (1972)]. *Under the conditions stated earlier, for all* $f \in C_b^2(R^d)$,

$$\begin{aligned} \Pi_t(f) = \Pi_0(f) &+ \int_0^t \Pi_s(h_s f) \, ds \\ &+ \sum_{i=1}^m \int_0^t [\tilde{\Pi}_s(f(X_s) h_s^i) - \Pi_s(f) \tilde{\Pi}_s(h_s^i) + \tilde{\Pi}_s(\tilde{D}_s^i f(X_s))] \, d\nu_s^i. \end{aligned} \tag{2.15}$$

*Stochastic equations for the conditional distribution* $\Pi_t(X_t \in \cdot | \mathscr{F}_t^Y)$. Suppose that the functions $h$ and $\tilde{D}$ [appearing in (2.12)] are given by

$$h_s^i(\omega) = h_s^i(X_s(\omega)) \quad \text{and} \quad \tilde{D}_s^{k,i}(\omega) = D_s^{k,i}(X_s(\omega)) \tag{2.16}$$

for suitable measurable functions $h_s^i$, $D_s^{k,i}$ from $\mathbf{R}^+ \times \mathbf{R}^d$ to $\mathbf{R}^1$. Let

$$D_s^i g(x) = \sum_{k=1}^{d} \sum_{j=1}^{d} D_s^{k,i}(x)\sigma_{k,j}(s,x)\frac{\partial}{\partial x^i}g(x). \tag{2.17}$$

In view of (2.16), the equation (2.15) reduces to the following equation, which is a SDE for the conditional distribution $\Pi_t(X_t \in \cdot|\mathscr{F}_t^Y)$ [recall: $\Pi_t(f) = \int f(x)\Pi_t(X_t \in dx|\mathscr{F}_t^Y)$].

$$\begin{aligned}\Pi_t(f) &= \Pi_0(f) + \int_0^t \Pi_s(\mathscr{L}_s f)\,ds \\ &\quad + \sum_{i=1}^{m} \int_0^t \left[\Pi_s(fh_s^i) - \Pi_s(f)\Pi_s(h_s^i) + \Pi_s(D_s^i f)\right] d\nu_s^i.\end{aligned} \tag{2.18}$$

This equation was first derived by Kushner (1967).

Zakai (1969) obtained a SDE equivalent to (2.18) which is easier to handle. It may be derived from (2.18) as follows: Let

$$\alpha_t = \exp\left(\int_0^t \sum_{i=1}^{m} \Pi_s(h_s^i)\,dY_s^i - \tfrac{1}{2}\int_0^t \sum_{i=1}^{m}\left[\Pi_s(h_s^i)\right]^2 ds\right). \tag{2.19}$$

Then,

$$\alpha_t = \exp\left(\int_0^t \sum_{i=1}^{m} \Pi_s(h_s^i)\,d\nu_s^i + \tfrac{1}{2}\int_0^t \sum_{i=1}^{m}\left[\Pi_s(h_s^i)\right]^2 ds\right).$$

Itô's formula gives

$$d\alpha_t = \alpha_t \sum_{i=1}^{m} \left(\Pi_t(h_t^i)\right)^2 + \alpha_t \sum_{i=1}^{m} \Pi_t(h_t^i)\,d\nu_t^i. \tag{2.20}$$

Using (2.18) and applying Itô's formula to $\sigma_t(f) = \Pi_t(f)\cdot\alpha_t$, we get

$$d\sigma_t(f) = \sigma_t(\mathscr{L}_t f) + \sum_{j=1}^{m} \sigma_t\left(D_t^j f + h_t^j f\right) dY_t^j. \tag{2.21}$$

The equation (2.21) is called the Zakai equation and is a measure-valued SDE. From the relation $\sigma_t(f) = \Pi_t(f)\cdot\alpha_t$, we have $\sigma_t(1) = \Pi_t(1)\cdot\alpha_t$ and since $\alpha_t > 0$ a.s., we get

$$\Pi_t(f) = \frac{\sigma_t(f)}{\sigma_t(1)}. \tag{2.22}$$

Thus, $\sigma_t(f)$ is called the unnormalized conditional expectation of $f$.

A derivation of the Zakai equation that is closer in spirit to the white noise approach of this paper, is based on what is known as the reference probability method. It is the approach followed by Pardoux (1979) and several other writers.

The following representation for $\Pi_t(f)$ can be obtained under the assumption that $h$ is bounded (or a suitable integrability condition). Let

$$N_t = \exp\left\{\int_0^t \sum_{i=1}^{m} h_s^i(X_s)\,dY_s^i - \tfrac{1}{2}\int_0^t |h_s(X_s)|^2\,ds\right\},$$

and let $\Pi'$ be a measure on $(\Omega, \mathscr{A})$ given by

$$d\Pi' = N_T^{-1}\, d\Pi.$$

By Girsanov's theorem, $\Pi'$ is a probability measure and $\{Y_t\}$ is a Wiener process under $\Pi'$. It can be shown that

$$\Pi_t(f) = \frac{E_{\Pi'}(f(X_t)N_t|\mathscr{F}_t^Y)}{E'_\Pi(N_t|\mathscr{F}_t^Y)} \tag{2.23}$$

and that the numerator in (2.23) is $\sigma_t(f)$. $\Pi'$ is called the reference probability.

The Zakai equation (2.21) can now be obtained directly from (2.23) which may be regarded as an "extended" Bayes formula. In Theorem 2.7 below, we employ this method to give an explicit derivation of (2.21) for the case of interest to us, viz., when the signal and noise are independent. Formula (2.23) then takes a simpler form equivalent to the Bayes formula obtained in Theorem 2.6.

*Stochastic PDE for the unnormalized conditional density.* When the (unconditional) distribution of $X_t$ admits a density wrt the Lebesgue measure on $\mathbf{R}^d$, then (under certain conditions) the functional $\sigma_t(f)$ also admits a density, i.e.,

$$\sigma_t(f) = \int_{R^d} f(x)p_t(x)\,dx \tag{2.24}$$

for a suitable $\mathscr{F}_t^Y$-adapted process $p_t(x)$. In view of (2.22), we have

$$E_\Pi(f(X_t)|\mathscr{F}_t^Y) = \frac{\int f(x)p_t(x)\,dx}{\int p_t(x)\,dx} \tag{2.25}$$

and hence $p_t(x)$ (if it exists) is called the unnormalized conditional density of $X_t$ given $\mathscr{F}_t^Y$. It follows from (2.21) that $p_t(x)$ satisfies the following stochastic PDE (SPDE)

$$dp_t(x) = \mathscr{L}_t^* p_t(x)\,dt + \sum_{j=1}^m (D_t^{j*} + h_t^j)p_t(x)\,dY_t^j, \tag{2.26}$$

where $\mathscr{L}_t^*$ and $D_t^{j*}$ are formal adjoints of $\mathscr{L}_t$ and $D_t^j$.

The problem of identifying the unnormalized conditional density as the unique solution to a SPDE has been studied by Krylov-Rozovskii (1981) and Pardoux (1979). We state Pardoux's result using his notation.

Let the system process $(X_t)$ and observation $(Y_t)$ be given by

$$dX_t = b(t, X_t)\,dt + \sigma(t, X_t)\,dW_t,$$
$$dY_t = h(t, X_t)\,dt + g(t)\,dW_t + g'(t)\,dW_t',$$

where $W_t$ and $W_t'$ are independent standard Wiener processes with values in $\mathbf{R}^d$ and $\mathbf{R}^m$, respectively, $\sigma_{ij}(t,x)$ are continuous, bounded on $[0,T]\times\mathbf{R}^d$, $i, j = 1,\dots,d$; $b_i(t,x), h_k(t,x)$ are Borel-measurable, bounded on $[0,T]\times\mathbf{R}^d$, $i = 1,\dots,d$; $k = 1,\dots,m$; $g_{k,i}(t), g'_{k,j}(t)$ are continuous on $\mathbf{R}^+$, $k, j = 1,\dots,m$; $i = 1,\dots,d$; and $g(t)g^*(t) + g'(t)g'^*(t) \equiv I$. Further, assume that $\sigma\sigma^*$ and

$g'g'^*$ are uniformly positive definite and for $i, j = 1, \dots, d$:

$$\frac{\partial}{\partial x^i}\sigma_{ij} \quad \text{exists and is bounded.}$$

Let $a = \sigma\sigma^*$, $c = g^*$, and $\bar{h}_t^k = h_t^k - \sum_{i=1}^d \{[\partial c_{k,i}(t)]/\partial x^i\}$. With these notations, the Zakai equation (2.26) reduces to

$$dp_t = \mathscr{L}_t^* p_t\, dt + [\bar{h}_t p_t - c_t \cdot \nabla p_t, dY_t], \qquad p_0 = \phi, \tag{2.27}$$

where $\phi$ is the density of $X_0$ and $\mathscr{L}_t$ is given by (2.5). Let

$$W^{1,2} = \left\{ u \in L^2(\mathbf{R}^d) \colon \frac{\partial u}{\partial x^i} \in L^2(\mathbf{R}^d) \right\}$$

and let $\mathscr{L}_t, \mathscr{L}_t^*$ be defined by

$$\langle \mathscr{L}_t u, v\rangle = \langle L_t^* v, u\rangle = -\tfrac{1}{2} \sum_{i,j=1}^d \int a_{ij}(t,x) \frac{\partial u}{\partial x^i}\frac{\partial v}{\partial x^j}\, dx + \int \sum_{i=1}^d a_i(t,x) \frac{\partial u}{\partial x^i}\, dx,$$

where $u, v \in W^{1,2}$ and

$$a_i = b_i - \frac{1}{2}\sum_{i=1}^d \frac{\partial a_{ij}}{\partial x^j}.$$

The equation (2.27) is to be taken as an equation for a $W^{1,2}$-valued process. Pardoux (1979) has proved the following result.

THEOREM 2.5. *The SPDE* (2.27) *has a unique solution $p_t(x)$ in the class*

$$L^2(\Omega \times (0,T), W^{1,2}) \cap L^2(\Omega, C[0,T], L^2(\mathbf{R}^d)).$$

*Further, $p_t(x)$ is the unnormalized conditional density of $X_t$ given $\mathscr{F}_t^Y$ [i.e., $p_t(x)$ satisfies* (2.25)].

*Bayes formula and Zakai equation in the signal-noise independent case.* We now specialize to the important case of nonlinear filtering when the signal process and noise are independent. Formula (2.23) reduces to a simpler form and can be looked upon as a Bayes formula. This gives the conditional expectations in terms of function space integrals. The SDEs for the optimal filter can be derived directly from this formula without using the sophisticated martingale-theoretic arguments. [See Kallianpur–Striebel (1969).]

In the model (1.2), let $(X_t)$ and $(W_t)$ be independent and let $h_t = h_t(X_t(\omega))$, where $h_t(x)$ satisfies

$$\int_0^T |h_t(X_t(\omega))|^2\, dt < \infty, \qquad \Pi\text{-a.s.} \tag{2.28}$$

Let $Z = (Z_t)$ be the canonical coordinate process on $\Omega_0 = C([0,T], \mathbf{R}^m)$, $\Pi_0$, the standard Wiener measure, and $\mathscr{A}_0$ be the Borel $\sigma$ field on $\Omega_0$, augmented with $\Pi_0$ null sets.

THEOREM 2.6 [Kallianpur–Striebel (1968)]. *Let $g$ be a $\mathscr{F}_T^X$-measurable integrable random variable. Then we have*

$$E_\Pi(g|\mathscr{F}_t^Y) = \Pi_t(g,Y) = \frac{\tilde{\sigma}_t(g,Y)}{\tilde{\sigma}_t(1,Y)}, \tag{2.29}$$

*where $\tilde{\sigma}_t(g,Z)$ is defined by*

$$\tilde{\sigma}_t(g,Z) = \int_\Omega g(\omega)\exp\left(\sum_{i=1}^m \int_0^t h_s^i(X_s(\omega))\,dZ_s^i - \tfrac{1}{2}\int_0^t |h_s(X_s(\omega))|^2\,ds\right)d\Pi(\omega). \tag{2.30}$$

The stochastic integral appearing in (2.30) is to be interpreted as a stochastic integral on $(\Omega\times\Omega_0, \mathscr{A}\times\mathscr{A}_0, \Pi\times\Pi_0)$. Here, (2.30) defines $\tilde{\sigma}_t(g,Z)$ as a Wiener functional for $\Pi_0$-a.e. $Z$. Since $\Pi Y^{-1}$ is equivalent to $\Pi_0$, $\tilde{\sigma}_t(g,Y(\omega))$ is defined for $\Pi$-a.e. $\omega$.

REMARK 2.1. $\tilde{\sigma}_t(g,Y(\omega))$ can be described directly as follows.

$$\tilde{\sigma}_t(g,Y(\omega)) = \int_\Omega g(\omega')\exp\left(\sum_{i=1}^d \int_0^t h_s^i(X_s(\omega'))\,dY_s^i(\omega) - \tfrac{1}{2}\int_0^t |h_s(X_s(\omega'))|^2\,ds\right)d\Pi(\omega'), \tag{2.31}$$

where the stochastic integral in (2.31) is to be interpreted as a stochastic integral on $(\Omega\times\Omega, \mathscr{A}\otimes\mathscr{A}, \Pi\otimes\Pi)$. That (2.31) is equivalent to (2.30) follows from the fact that since $\Pi Y^{-1}$ is equivalent to $\Pi_0$, $\int f\,dZ$(under $\Pi_0$) $= \int f\,dZ$(under $\Pi Y^{-1}$) [see Meyer (1976)].

Now suppose in addition that $(X_t)$ is an $S$-valued Markov process with generator $\mathscr{L}$ [where $(S,\mathscr{S})$ is a measurable space]. Let $\mathscr{D}$ be the domain of $\mathscr{L}$. Let $\mathscr{I}(S,\mathscr{S})$ be the class of bounded measurable functions on $(S,\mathscr{S})$. Let $\mathscr{D}_0$ be the class of functions $f\in\mathscr{I}(S,\mathscr{S})$ s.t. $f_1$ defined by

$$f_1(s,x) = f(x)$$

belongs to $\mathscr{D}$ and let $(\mathscr{L}_t f)(x) = (\mathscr{L}f_1)(t,x)$. As before, let us write for $f\in\mathscr{I}(S,\mathscr{S})$

$$\sigma_t(f,Y) = \tilde{\sigma}_t(f(X_t),Y).$$

The Zakai equation (2.21) in this special case can be derived easily from the Bayes formula (2.29). For a proof, see Kallianpur–Karandikar (1983b). Ocone (1985) has shown that condition (2.32) can be replaced by the conditions $E_\Pi\int_0^T|h_s(X_s)|^2\,ds<\infty$ and $E_\Pi|h_s(X_s)|<\infty$ for $0\le s\le T$.

THEOREM 2.7. *Suppose that*

$$E_\Pi\left[\exp\left(\int_0^T |h_s(X_s)|^2\,ds\right)\right] < \infty. \tag{2.32}$$

*Then for all* $f \in \mathscr{D}_0$,

$$d\sigma_t(f,Y) = \sigma_t(\mathscr{L}_t f, Y)\,dt + \sum_{i=1}^{m} \sigma_t(h_t^i f, Y)\,dY_t^i. \tag{2.33}$$

When the process $(X_t)$ is an $\mathbf{R}^d$-valued diffusion process, the Zakai equation for the unnormalized conditional density $p_t(x,Y)$ (2.27) reduces to

$$dp_t(x,Y) = \mathscr{L}_t^* p_t(x,Y)\,dt + \sum_{i=1}^{m} h_t^i(x)p_t(x,Y)\,dY_t^i. \tag{2.34}$$

The existence of a solution to (2.34) and its identification as the unnormalized conditional density follow from Pardoux's result (Theorem 2.5 above).

A different approach to the solution of the filtering problem due to Davis (1980) is as follows. For each $Z \in C([0,T],\mathbf{R}^m)$, define a two parameter semigroup acting on $\mathscr{I}(S,\mathscr{S})$ by

$$\begin{aligned}(T_{s,t}^Z f)(x) = E_\Pi\Bigg( f(X_t)\exp\Bigg(\sum_{i=1}^{m}\Big\{-Z_t^i h_t^i(X_t) \\ + Z_s^i h_s^i(X_s) + \int_s^t h_u^i(X_u)\,dZ_u^i \\ - \tfrac{1}{2}\int_s^t (h_u^i(X_u))^2\,du\Big\}\Bigg)\Bigg| X_s = x\Bigg).\end{aligned} \tag{2.35}$$

It is easy to verify that

$$\sigma_t(f,Y) = \int_{R^d} T_{s,t}^Y\Bigg(\exp\Bigg(\sum_{i=1}^{m} Y_t^i h_t^i(\cdot)\Bigg)f\Bigg)(x)\,d\mu(x), \tag{2.36}$$

where $\mu$ is the distribution of $X_0$. If we assume that $h_t(X_t)$ is a continuous semimartingale, then by the integration by parts formula, we have

$$\begin{aligned}(T_{s,t}^Z g)(x) = E_\Pi\Bigg( f(X_t)\exp\Bigg\{-\int_s^t \sum_{i=1}^{m} Z_u^i\,dh_u^i(X_u) \\ - \tfrac{1}{2}\int_s^t |h_u(X_u)|^2\,du\Bigg\}\Bigg| X_s = x\Bigg).\end{aligned} \tag{2.37}$$

Observe that the path $Z$ appears as a parameter in (2.37). If $(X_t)$ is an $\mathbf{R}^d$-valued diffusion and $h$ is smooth, it is possible to compute the generator $\{A_s^Z\}$ of $T_{s,t}^Z$ for each $Z$. This, in view of (2.36), gives a "pathwise solution" to the filtering problem since, given an observation path Y, we can construct the semigroup $T_{s,t}^Y$ from its generator $\{A_s^Y\}$ and then compute $\sigma_t(f,Y)$ by (2.36). Davis (1980) has obtained an expression for $\{A_s^Z\}$ when $(X_t)$ is a Lévy process.

Another approach to the pathwise solution of the filtering problem will be discussed in Section 8.

**3. Some basic ideas of finitely additive white noise theory.** As explained in the Introduction, in our approach to filtering theory, the noise is defined by a Gaussian cylinder measure whereas the signal (or system) process is understood to be a stochastic process in the usual customary sense, i.e., a process, usually assumed to be Markov, defined on a countably additive probability space. For a rigorous treatment of a noise independent signal model we first need to define several important definitions and concepts upon which our work is based. The major thrust of these definitions is to create a theory general enough for the purpose of nonlinear prediction, filtering, and estimation theory. We introduce integration with respect to a quasicylinder probability measure (QCP) of suitable classes of cylinder functions, followed by a definition of absolute continuity of QCPs. The notions of a representation and lifting associated with a QCP are indispensable and, though very similar ideas are met with in the works of Segal and of Gross [Segal (1956), Gross (1962)], our problems are sufficiently different as to require a detailed and independent treatment. One of the most useful and important of our concepts is that of a quasicylindrical mapping (QCM), which has most of the desirable properties of a random variable in the countably additive theory. A QCM defined on a quasicylindrical probability space induces a QCP on the range space and an associated induced representation. Furthermore, it enables us to give a definition of conditional expectation more inclusive than the one to be found in an earlier paper of ours [Kallianpur and Karandikar (1983a)].

Let $H$ be a real separable infinite-dimensional Hilbert space with inner product $(\cdot,\cdot)$ and norm $|\cdot|$. Let $\mathscr{P} = \mathscr{P}(H)$ be the class of orthogonal projections on $H$ with finite-dimensional range. For $P \in \mathscr{P}$, let

$$\mathscr{C}_P = \{P^{-1}B\colon B \text{ a Borel set in range } P\}$$

and

$$\mathscr{C} = \mathscr{C}(H) = \bigcup_{P\in\mathscr{P}} \mathscr{C}_P.$$

DEFINITION. A cylinder measure $n$ on $(H,\mathscr{C})$ is a finitely additive probability measure on $(H,\mathscr{C})$ such that for all $P \in \mathscr{P}$ its restriction to $\mathscr{C}_P$ is countably additive.

Gross has defined $\int f\,dn$ for a certain class of functions. We will generalize Gross's definition to a space which is a "product" of a measurable space and a Hilbert space. Such a product space was used in Kallianpur-Karandikar (1983a) as a model for nonlinear filtering with white noise. There, we did not define integration on the product space for a sufficiently large class of functions, but defined conditional expectation by going over to the range.

Let $(\Omega,\mathscr{A})$ be a measurable space and let

$$E = \Omega \times H.$$

For $P \in \mathscr{P}$, let

$$\mathscr{E}_P = \mathscr{A} \otimes \mathscr{C}_P$$

and

$$\mathscr{E} = \bigcup_{P \in \mathscr{P}} \mathscr{E}_P.$$

Here, $\mathscr{A} \otimes \mathscr{C}_P$ denotes the product $\sigma$-field on $\Omega \times H = E$. Thus, $\mathscr{E}_P$ is a $\sigma$-field on $E$ for each $P$ and $\mathscr{E}$ is a field. We will denote $(E, \mathscr{E})$ by $(\Omega, \mathscr{A}) \odot (H, \mathscr{C})$. $(E, \mathscr{E})$ will be called a quasicylindrical measurable space.

DEFINITION. A quasicylinder probability measure $\beta$ on $(E, \mathscr{E})$ is a finitely additive positive measure on $(E, \mathscr{E})$ with $\beta(E) = 1$ such that for all $P \in \mathscr{P}$, $\beta_P$, its restriction to $\mathscr{E}_P$, is countably additive.

Dunford and Schwartz (1966) have given a definition of integration wrt finitely additive measures for a certain class of functions. We need to integrate a larger class of functions wrt $\beta$ using the property that $\beta_P$ is countably additive for each $P$. For this, we need the notion of a "representation" of a quasicylinder probability measure, which is defined below.

DEFINITION. A representation of a quasicylinder probability measure $\beta$ is a triplet $(\rho, L, \tilde{\Pi})$, where $\tilde{\Pi}$ is a countably additive probability measure on a measurable space $(\tilde{\Omega}, \tilde{\mathscr{A}})$, $\rho$ is a measurable mapping from $(\tilde{\Omega}, \tilde{\mathscr{A}})$ into $(\Omega, \mathscr{A})$, and $L$ is a mapping from $H$ into the space of real-valued Borel-measurable functions on $(\tilde{\Omega}, \tilde{\mathscr{A}})$ such that for all $A \in \mathscr{A}$ and for all $C \in \mathscr{C}$,

$$\beta(A \times C) = \tilde{\Pi}\{\tilde{\omega}: \rho(\tilde{\omega}) \in A, (L(h_1)(\tilde{\omega}), \dots, L(h_j)(\tilde{\omega})) \in B\}, \tag{3.1}$$

where $C \in \mathscr{C}$ is given by

$$C = \{h \in H: ((h h_1), \dots, (h, h_j)) \in B\} \tag{3.2}$$

for $h_i \in H$ and a Borel set $B$ in $\mathbf{R}^j$.

REMARK 3.1. If $(\rho, L, \tilde{\Pi})$ is a representation of a quasicylinder probability measure $\beta$, then it can be easily seen that

$$L(a_1 h_1 + a_2 h_2) = a_1 L(h_1) + a_2 L(h_2), \qquad \tilde{\Pi}\text{-a.e.} \tag{3.3}$$

for all $h_i \in H$, $a_i \in R$, $i = 1, 2$. Indeed, take $h_3 = a_1 h_1 + a_2 h_2$ and $B = \{(x_1, x_2, x_3) \in R^3: x_3 = a_1 x_1 + a_2 x_2\}$. Let $C$ be defined by (3.2) with $j = 3$ and let $A = \Omega$. Then, $\beta(A \times C) = \beta(E) = \tilde{\Pi}(\tilde{\Omega})$ and by (3.1),

$$\tilde{\Pi}(\tilde{\Omega}) = \tilde{\Pi}\{\tilde{\omega}: L(a_1 h_1 + a_2 h_2)(\tilde{\omega}) = a_1 L(h_1)(\tilde{\omega}) + a_2 L(h_2)(\tilde{\omega})\}.$$

Also, if *a priori*, it is known that $L$ satisfies (3.3) then in order to show that $(\rho, L, \tilde{\Pi})$ is a representation of $\beta$, it suffices to verify (3.1)–(3.2) for $j = 1$. To see this, first observe that (3.1) is equivalent to

$$\begin{aligned} &\int_E 1_A(\omega) \exp\left(i \sum_{k=1}^{j} a_k (h, h_k)\right) d\beta_P(\omega, h) \\ &\quad = \int_{\tilde{\Omega} \times \rho^{-1}(A)} 1(\tilde{\omega}) \exp\left(i \sum_{k=1}^{j} a_k L(h_k)(\tilde{\omega})\right) d\tilde{\Pi}(\tilde{\omega}) \end{aligned} \tag{3.4}$$

for all $A \in \mathscr{A}$, $h_k \in H$, $a_k \in \mathbf{R}$, $j \geq 1$, where $P$ is the orthogonal projection onto $\operatorname{span}\{h_1, \ldots, h_j\}$. [Observe that the integrand on the left-hand side of (3.4) is $\mathscr{E}_P$-measurable and since $\beta_P$ is countably additive on $\mathscr{E}_P$, the integral is well defined.] If $L$ satisfies (3.3), then clearly, (3.4) is equivalent to

$$\begin{aligned} &\int_E 1_A(\omega)\exp(i(h, h_0))\, d\beta_P(\omega, h) \\ &\qquad = \int_{\tilde\Omega} 1_{\rho^{-1}(A)}(\tilde\omega)\exp(iL(h_0)(\tilde\omega))\, d\tilde\Pi(\tilde\omega) \end{aligned} \tag{3.5}$$

for all $A \in \mathscr{A}$ and $h_0 \in H$, $P$ being the projection onto the linear span of $h_0$, which in turn is equivalent to (3.1)–(3.2) for $j = 1$.

Before we proceed, we remark that the following result, stated here without proof, guarantees the existence of a representation of $\beta$ for a large class of quasicylinder measures.

THEOREM 3.1. *Let $\beta$ be a quasicylinder probability measure on $(E, \mathscr{E})$. Assume that there exists a probability measure $\Pi$ on $(\Omega, \mathscr{A})$ and a mapping $\lambda\colon \Omega \times \mathscr{C} \to \mathbf{R}^+$ such that*

(i) *For all $C \in \mathscr{C}$, $\omega \to \lambda(\omega, C)$ is $\mathscr{A}$-measurable.*

(ii) *For all $\omega \in \Omega$, $C \to \lambda(\omega, C)$ is a cylinder probability measure on $(H, \mathscr{C})$.*

(iii) *For $A \in \mathscr{A}$, $C \in \mathscr{C}$,*

$$\beta(A \times C) = \int_\Omega 1_A(\omega)\lambda(\omega, C)\, d\Pi(\omega). \tag{3.6}$$

*Then a representation of $\beta$ exists.*

REMARK 3.2. Given $\lambda, \Pi$ satisfying (i), (ii) (in Theorem 3.1), the formula

$$\beta(F) = \int_\Omega \lambda(\omega, F^\omega)\, d\Pi(\omega), \qquad F \in \mathscr{E},$$

where $F^\omega = \{h \in H\colon (\omega, h) \in F\}$, defines a quasicylinder probability measure on $(E, \mathscr{E})$.

If $\mathscr{A}$ is countably generated, then it can be shown that a representation exists for any quasicylinder probability measure $\beta$ on $(\Omega, \mathscr{A}) \odot (H, \mathscr{C})$.

For the remaining part of this section, we will consider a fixed quasicylinder probability measure $\beta$ on $(E, \mathscr{E}) = (\Omega, \mathscr{A}) \odot (H, \mathscr{C})$. *We will assume throughout that a representation of $\beta$ exists and $(\rho, L, \tilde\Pi)$ will denote a (fixed) representation of $\beta$.*

Let $S$ be a complete separable metric space with a distance function $d$. Let $\mathscr{L}(\tilde\Omega, \tilde{\mathscr{A}}, \tilde\Pi; S)$ denote the space of $S$-valued $(\tilde{\mathscr{A}}, \mathscr{B}(S))$-measurable functions $X$ on $\tilde\Omega$, where two such mappings $X_1, X_2$ are equivalent if $X_1 = X_2$ a.e. $\tilde\Pi$. For $X_1, X_2 \in \mathscr{L}(\tilde\Omega, \tilde{\mathscr{A}}, \tilde\Pi; S)$, let $d^*(X_1, X_2)$ be defined by

$$d^*(X_1, X_2) = \int [d(X_1(\tilde\omega), X_2(\tilde\omega))] \wedge 1\, d\tilde\Pi(\tilde\omega).$$

Then, it is well known that $d^*$ is a metric on $\mathscr{L}(\hat{\Omega}, \hat{\mathscr{A}}, \hat{\Pi}; S)$ and $d^*(X_n, X) \to 0$ if and only if $X_n \to X$ in $\hat{\Pi}$ probability. Also, $\mathscr{L}(\hat{\Omega}, \hat{\mathscr{A}}, \hat{\Pi}; S)$ is a complete metric space under the metric $d^*$. This can be checked using the completeness of $S$ under $d$. We will denote $\mathscr{L}(\hat{\Omega}, \hat{\mathscr{A}}, \hat{\Pi}; \mathbf{R})$ by $\mathscr{L}(\hat{\Omega}, \hat{\mathscr{A}}, \hat{\Pi})$.

Definition. A function $f$: $E \to S$ is called a ($S$-valued) cylinder function if it can be written as

$$f(\omega, h) = f_1(\omega, (h, h_1), \ldots, (h, h_j)) \tag{3.7}$$

for some $h_1, \ldots, h_j \in H$ and some measurable function $f_1$: $(\Omega \times \mathbf{R}^j, \mathscr{A} \otimes \mathscr{B}(\mathbf{R}^j)) \to (S, \mathscr{B}(S))$.

It can be verified that $f$ is a cylinder function if and only if for some $P \in \mathscr{P}$,

$$f^{-1}(\mathscr{B}(S)) \subseteq \mathscr{E}_P. \tag{3.8}$$

For a cylinder function $f$ given by (3.7), define $R_\beta(f) \in \mathscr{L}(\hat{\Omega}, \hat{\mathscr{A}}, \hat{\Pi}; S)$ by

$$R_\beta(f) = f_1(\rho, L(h_1), \ldots, L(h_j)). \tag{3.9}$$

If $(\Omega', \mathscr{A}')$ is a measurable space, then $\Omega'$-valued cylinder functions $f$ and $R_\beta(f)$ for such an $f$ can be defined as above by replacing $(S, \mathscr{B}(S))$ by $(\Omega', \mathscr{A}')$.

Lemma 3.2. *Let $f$ be an $S$-valued cylinder function. Then*

(i) $R_\beta(f)$ *is unambiguously defined by* (3.7) *and* (3.9) *as an element of* $\mathscr{L}(\hat{\Omega}, \hat{\mathscr{A}}, \hat{\Pi}; S)$;

(ii) *if $P \in \mathscr{P}$ is such that* (3.8) *holds, then*

$$\hat{\Pi}[R_\beta(f)]^{-1} = \beta_P[f]^{-1}. \tag{3.10}$$

We now extend the mapping $R_\beta$ to a larger class of functions for which a useful theory of integration can be developed.

Define a partial order on $\mathscr{P}$ by $P_1 < P_2$ if range $P_1 \subseteq$ range $P_2$. It is easy to see that $(\mathscr{P}, <)$ is a directed set. (We also write $P_2 > P_1$).

Definition. Let $\mathscr{L}(E, \mathscr{E}, \beta; S)$ be the class of functions $f$ from $E$ into $S$ such that for all $P \in \mathscr{P}$, $f_P$ defined by

$$f_P(\omega, h) = f(\omega, Ph) \tag{3.11}$$

is $(\mathscr{E}_P, \mathscr{B}(S))$-measurable and the net $\{R_\beta(f_P): P \in \mathscr{P}\}$ is $d^*$-Cauchy (i.e., Cauchy in $\hat{\Pi}$ probability). For such an $f$, let

$$R_\beta(f) = \lim_{P \in \mathscr{P}} \text{ in } \hat{\Pi} \text{ probability } R_\beta(f_P). \tag{3.12}$$

The following lemma shows that the class $\mathscr{L}(E, \mathscr{E}, \beta; S)$ does not depend upon the choice of the representation.

LEMMA 3.3. *Let $(\rho', L', \Pi')$ be another representation of $\beta$ and let $R'_\beta$ be the map defined by* (3.9) *for cylinder functions. Let $f$: $E \to S$ be such that for all $P \in \mathscr{P}$, $f_P$ [defined by* (3.11)] *is $(\mathscr{E}_P, \mathscr{B}(S))$-measurable. Then*

(i) *$\{R_\beta(f_P)$: $P \in \mathscr{P}\}$ is Cauchy in $\tilde{\Pi}$ measure if and only if $\{R'_\beta(f_P)$: $P \in \mathscr{P}\}$ is Cauchy in $\Pi'$ measure.*

(ii) *If* (i) *holds, then denoting the limit of $R'_\beta(f_P)$ by $R'_\beta(f)$, we have*

$$\tilde{\Pi}\left[R_\beta(f)\right]^{-1} = \Pi'\left[R'_\beta(f)\right]^{-1}. \tag{3.13}$$

The proof is omitted.

Elements of $\mathscr{L}(E, \mathscr{E}, \beta; S)$ will be called $S$-valued $\beta$-measurable functions and the *mapping $R_\beta$ defined by* (3.12), *$R_\beta$: $\mathscr{L}(E, \mathscr{E}, \beta; S) \to \mathscr{L}(\tilde{\Omega}, \tilde{\mathscr{A}}, \tilde{\Pi}; S)$ will be called the $\beta$-lifting corresponding to $(\rho, L, \tilde{\Pi})$.*

Clearly, $\mathscr{L}(E, \mathscr{E}, \beta; S)$ contains $S$-valued cylinder functions. For, if $f$ satisfies (3.8) with $P \in \mathscr{P}$, then for all $P' > P$, $P' \in \mathscr{P}$, $f_{P'} \equiv f$ and $R_\beta(f_{P'}) = R_\beta(f)$. This also shows that (3.12) holds for cylinder functions and hence the definition (3.12) of $R_\beta(f)$ for a cylinder function $f$ does not contradict the previous definition (3.9).

An equivalent description of the class $\mathscr{L}(E, \mathscr{E}, \beta; S)$ which avoids the use of nets is given by the following result:

LEMMA 3.4. *Let $f$: $E \to S$ be such that for each $P \in \mathscr{P}$, $f_P$ defined by* (3.11) *is $(\mathscr{E}_P, \mathscr{B}(S))$-measurable. Then, $f$ belongs to $\mathscr{L}(E, \mathscr{E}, \beta; S)$ if and only if*

(3.14) *there exists $\{P_k\} \subseteq \mathscr{P}$, $P_k \uparrow I$, such that for all $\{P'_k\} \subseteq \mathscr{P}$, $P'_k > P_k$ for all $k \geq 1$, $\{R_\beta(f_{P'_k})\}$ is a $d^*$-Cauchy sequence.*

*Further, if* (3.14) *holds, then for all $\{P'_k\} \subseteq \mathscr{P}$, $P'_k > P_k$ $\forall k \geq 1$,*

$$R_\beta(f) = \lim_{k \to \infty} \text{ in } \tilde{\Pi}\text{-probability } R_\beta(f_{P'_k}). \tag{3.15}$$

PROOF. Suppose $f \in \mathscr{L}(E, \mathscr{E}, \beta; S)$. For each $k \geq 1$, get $\tilde{P}_k \in \mathscr{P}$ such that for all $P$, $P' \in \mathscr{P}$, $P > \tilde{P}_k$, $P' > \tilde{P}_k$

$$d^*\left(R_\beta(f_P), R_\beta(f_{P'})\right) < \frac{1}{k}. \tag{3.16}$$

Choose a sequence $\{P_k\} \subseteq \mathscr{P}$, $P_k \uparrow I$ such that $P_k > \tilde{P}_k$ for all $k$. Then (3.14) holds for this choice of $\{P_k\}$.

For the other part, suppose (3.14) holds. Completeness of $\mathscr{L}(\tilde{\Omega}, \tilde{\mathscr{A}}, \tilde{\Pi}; S)$ under the metric $d^*$ and the usual interlacing argument gives the existence of $Z \in \mathscr{L}(\tilde{\Omega}, \tilde{\mathscr{A}}, \tilde{\Pi}; S)$ such that for all $\{P'_k\} \subseteq \mathscr{P}$, $P'_k > P_k$ $\forall k \geq 1$,

$$R_\beta(f_{P'_k}) \to Z \quad \text{in } \tilde{\Pi} \text{ measure}. \tag{3.17}$$

We now show that (3.17) implies that the net $\{R_\beta(f_P)$: $P \in \mathscr{P}\}$ converges to $Z$ in $\tilde{\Pi}$ probability. Suppose $\{R_\beta(f_P)\}$ does not converge to $Z$. Then there exists an

$\varepsilon > 0$ such that for all $P \in \mathscr{P}$, there exists a $P' \in \mathscr{P}$, $P' > P$, with

$$d^*\big(R_\beta(f_{P'}), Z\big) \geq \varepsilon. \tag{3.18}$$

Now, for each $k \geq 1$, choose $P_k' \in P$, $P_k' > P_k$ such that (3.18) holds for $P' = P_k'$. Then $d^*(R_\beta(f_{P_k'}), Z) \geq \varepsilon$ for all $k$. This contradicts (3.17). Thus, $f \in \mathscr{L}(E, \mathscr{E}, \beta; S)$ and (3.15) holds. □

The next result lists some properties of the mapping $R_\beta$.

THEOREM 3.5.

(i) *Let $S = \mathbf{R}$. Then the space $\mathscr{L}(E, \mathscr{E}, \beta; \mathbf{R})$ is a linear space and is closed under pointwise multiplication of functions and the mapping $R_\beta$: $\mathscr{L}(E, \mathscr{E}, \beta; \mathbf{R}) \to \mathscr{L}(\tilde{\Omega}, \tilde{\mathscr{A}}, \tilde{\Pi}; \mathbf{R})$ is linear and multiplicative. Also, if $a \leq f \leq b$, then $\tilde{\Pi}(a \leq R_\beta(f) \leq b) = 1$.*

(ii) *For each $k$, let $S_k$ be a complete separable metric space and let $S = \times_{k=1}^{\infty} S_k$ be the product of $S_k$s with the product topology. $S$ itself is a complete separable metric space (under a suitable metric). Suppose $f_k \in \mathscr{L}(E, \mathscr{E}, \beta; S_k)$ for $k \geq 1$. Let $f$: $E \to S$ be defined by*

$$f(\omega, h) = \big(f_1(\omega, h), \ldots, f_k(\omega, h), \ldots\big), \qquad (\omega, h) \in E.$$

*Then $f \in \mathscr{L}(E, \mathscr{E}, \beta; S)$ and further*

$$R_\beta(f) = \big(R_\beta(f_1), \ldots, R_\beta(f_k), \ldots\big).$$

(iii) *Let $S_1, S_2$ be complete separable metric spaces and let $U \in \mathscr{B}(S_1)$. Let $g$ be a continuous function from $U$ into $S_2$. Suppose $f \in \mathscr{L}(E, \mathscr{E}, \beta; S_1)$ is such that $(\text{range} f) \subseteq U$ and $\tilde{\Pi}(R_\beta(f) \notin U) = 0$. Then $g \circ f \in \mathscr{L}(E, \mathscr{E}, \beta; S_2)$ and*

$$R_\beta(g \circ f) = g\big(R_\beta(f)\big). \tag{3.19}$$

*Also, if $f \in \mathscr{L}^*(E, \mathscr{E}, \beta; S_1)$ then $g \circ f \in \mathscr{L}^*(E, \mathscr{E}, \beta; S_2)$.*

PROOF. (i) and (ii) follow from the definition of the mapping $R_\beta$ and the properties of convergence in probability on countably additive measure spaces. For (iii), first observe that for $P \in \mathscr{P}$, $(g \circ f)_P$ [defined by (3.11)] is equal to $g \circ f_P$ and hence

$$\begin{aligned} R_\beta((g \circ f)_P) &= R_\beta(g \circ f_P) \\ &= g\big(R_\beta(f_P)\big). \end{aligned} \tag{3.20}$$

Now, $\tilde{\Pi}(R_\beta(f) \notin U) = 0$, continuity of $g$ from $U$ into $S_2$, and the fact that $R_\beta(f_P) \to R_\beta(f)$ in $\tilde{\Pi}$ probability implies that $R_\beta((g \circ f)_P) \to g(R_\beta(f))$ in $\tilde{\Pi}$ probability. This proves (iii). □

We will now define $\int f\, d\beta$ for a suitable class of functions $f$. The motivation for this definition is that the distribution of $R_\beta f$ under $\tilde{\Pi}$ can be considered as the "distribution of $f$ under $\beta$."

**Definition.** Let

$$\mathscr{L}^1(E,\mathscr{E},\beta) = \left\{ f \in \mathscr{L}(E,\mathscr{E},\beta;\mathbf{R}) : \int_{\hat{\Omega}} |R_\beta(f)|\, d\tilde{\Pi} < \infty \right\}$$

and for $f \in \mathscr{L}^1(E,\mathscr{E},\beta)$, define

$$\int_E f\, d\beta = \int_{\hat{\Omega}} R_\beta(f)\, d\tilde{\Pi}.$$

In view of Lemma 3.3, the class $\mathscr{L}^1(E,\mathscr{E},\beta)$ and the value of the integral $\int f\, d\beta$ for $f \in \mathscr{L}^1(E,\mathscr{E},\beta)$ do not depend on the particular choice of the representation $(\rho, L, \tilde{\Pi})$.

The definitions given above subsume the definition of integration wrt a cylinder measure $n$ on $(H,\mathscr{C})$. In fact, take $\Omega$ to be a singleton $\{\omega\}$ and let $\lambda(\omega,\cdot) \equiv n$. Then, we can identify $(E,\mathscr{E},\beta)$ with $(H,\mathscr{C},n)$ in the obvious way. Thus, all the notions defined on $(E,\mathscr{E},\beta)$ have a similar meaning on $(H,\mathscr{C},n)$. If $\beta$ is as above and $(\rho, L, \tilde{\Pi})$ is a representation of $\beta$, then since $\Omega$ is a singleton, $\rho$ is constant and we will call $(L,\tilde{\Pi})$ itself a representation of $\beta$ (or $n$).

We will now introduce subclasses of $\mathscr{L}(E,\mathscr{E},\beta;S)$ and $\mathscr{L}^1(E,\mathscr{E},\beta)$. The elements of these subclasses satisfy a stronger approximation property.

Let $\mathscr{L}^*(E,\mathscr{E},\beta;S)$ be the class of functions $f$ from $E$ to $S$ such that for all $P \in \mathscr{P}$, $f_P$ defined by (3.11) is $(\mathscr{E}_P, \mathscr{B}(S))$-measurable and for all sequences $\{P_k\} \subset \mathscr{P}$, converging strongly to the identity (written as $P_k \xrightarrow{s} I$), $R(f \circ P_k)$ is Cauchy in $\tilde{\Pi}$ probability. In view of the Lemma 3.4,

$$\mathscr{L}^*(E,\mathscr{E},\beta,S) \subseteq \mathscr{L}(E,\mathscr{E},\beta;S) \tag{3.21}$$

and for $f \in \mathscr{L}^*(E,\mathscr{E},\beta;S)$,

$$R_\beta(f \circ P_k) \to R_\beta(f) \quad \text{in } \tilde{\Pi}\text{-probability}$$

for all $P_k \xrightarrow{s} I$, $\{P_k\} \subseteq \mathscr{P}$. Let

$$\mathscr{L}^{1*}(E,\mathscr{E},\beta) = \left\{ f \in \mathscr{L}^*(E,\mathscr{E},\beta;\mathbf{R}) : \forall\ \{P_k\} \subseteq \mathscr{P},\ P_k \xrightarrow{s} I,\right.$$
$$\left. \int_{\hat{\Omega}} |R_\beta(f_{P_k}) - R_\beta(f)|\, d\tilde{\Pi} \to 0 \right\}.$$

Again, it is easy to see that

$$\mathscr{L}^{1*}(E,\mathscr{E},\beta) \subseteq \mathscr{L}^1(E,\mathscr{E},\beta). \tag{3.22}$$

The most commonly used definition of absolute continuity for finitely additive measures is the following: Let $\mu_1, \mu_2$ be finitely additive measures on $(D,\mathscr{D})$, where $\mathscr{D}$ is a field on $D$. Then $\mu_1$ is said to be absolutely continuous wrt $\mu_2$ if given $\varepsilon > 0$, there exists $\delta > 0$ s.t. for $A \in \mathscr{D}$, $\mu_2(A) < \delta$ implies $\mu_1(A) < \varepsilon$. However, this does not imply the existence of Radon–Nikodym derivative. We can only assert the following: Given any $\varepsilon > 0$, there exists a simple function $f_\varepsilon$ s.t. $|\mu_1(A) - \int_A f_\varepsilon\, d\mu_2| < \varepsilon$ for all $A \in \mathscr{D}$. This is known as Bochner's theorem and

$f_\varepsilon$ is known as an $\varepsilon$-derivative of $\mu_1$ wrt $\mu_2$. [See Dunford–Schwartz (1958).] The notion of $\varepsilon$-derivative is not suitable for statistical purposes, namely for defining likelihood ratios or conditional expectation.

Balakrishnan has used this $\varepsilon$-$\delta$ definition in the context of cylinder measures. Again, in his setup, absolute continuity does not imply existence of a Radon–Nikodym derivative.

Gross's notion of a Radon–Nikodym derivative for cylinder measures is also unsuitable for statistical purposes, for roughly speaking, his Radon–Nikodym derivative is a measurable function on the representation space.

We will use a stronger notion of absolute continuity—one in which the existence of Radon–Nikodym derivative is built into the definition.

Let $\beta, \beta_1$ be quasicylinder probability measures on $(E, \mathscr{E})$.

DEFINITION. $\beta_1$ is absolutely continuous wrt $\beta$ (written as $\beta_1 \ll \beta$) if there exists a nonnegative function $f \in \mathscr{L}^1(E, \mathscr{E}, \beta)$ such that for all $F \in \mathscr{E}$

$$\beta_1(F) = \int_E 1_F \cdot f \, d\beta. \tag{3.23}$$

Further, $f$ is defined to be the Radon–Nikodym derivative of $\beta_1$ wrt $\beta$ and will also be written as $d\beta_1/d\beta$.

The following theorem gives some of the properties of the Radon–Nikodym derivative in our setup.

THEOREM 3.6. *Let $\beta_1$ and $\beta_2$ be quasicylinder probability measures on $(E, \mathscr{E})$, such that $\beta_i \ll \beta$, $i = 1, 2$.*

(i) *If $a_1, a_2 \in \mathbf{R}^+$, then $(a_1\beta_1 + a_2\beta_2) \ll \beta$ and*

$$\frac{d(a_1\beta_1 + a_2\beta_2)}{d\beta} = a_1\frac{d\beta_1}{d\beta} + a_2\frac{d\beta_2}{d\beta}. \tag{3.24}$$

(ii) *If $g \in \mathscr{L}(E, \mathscr{E}, \beta; S)$, then $g \in \mathscr{L}(E, \mathscr{E}, \beta_1; S)$ and $R_{\beta_1}(g) = R_\beta(g)$. Further, if $g \in \mathscr{L}^*(E, \mathscr{E}, \beta; S)$, then $g \in \mathscr{L}^*(E, \mathscr{E}, \beta_1; S)$.*

(iii) *Suppose $g \in \mathscr{L}^*(E, \mathscr{E}, \beta; \mathbf{R})$. Then*

$$g \in \mathscr{L}^1(E, \mathscr{E}, \beta_1) \quad \textit{if and only if} \quad g \cdot (d\beta_1/d\beta) \in \mathscr{L}^1(E, \mathscr{E}, \beta). \tag{3.25}$$

*Further, if $g \in \mathscr{L}^1(E, \mathscr{E}, \beta_1)$, then*

$$\int_E g \cdot \frac{d\beta_1}{d\beta} \cdot d\beta = \int g \cdot d\beta_1. \tag{3.26}$$

(iv) *If $d\beta_1/d\beta > 0$ and $R_\beta(d\beta_1/d\beta) > 0$ a.s. $\tilde{\Pi}$, then $\beta \ll \beta_1$ and*

$$\frac{d\beta}{d\beta_1} = \left[\frac{d\beta_1}{d\beta}\right]^{-1}. \tag{3.27}$$

PROOF. (i) follows easily from the definition. For the remaining parts, let $X = R_\beta(d\beta_1/d\beta)$ and let $\Pi'$ be the measure on $(\tilde{\Omega}, \tilde{\mathscr{A}})$ defined by

$$\Pi'(\tilde{A}) = \int_{\tilde{A}} R_\beta\left(\frac{d\beta_1}{d\beta}\right) d\tilde{\Pi}, \qquad \text{for } \tilde{A} \in \tilde{\mathscr{A}}. \tag{3.28}$$

Now let $A \in \mathscr{A}$, and let $C \in \mathscr{C}$ be given by (3.2). Let

$$\tilde{A} = \left\{\tilde{\omega} \in \tilde{\Omega}: \rho(\tilde{\omega}) \in A, \left(L(h_1)(\tilde{\omega}), \ldots, L(h_j)(\tilde{\omega})\right) \in B\right\}.$$

[Recall that $(\rho, L, \tilde{\Pi})$ is a (fixed) representation of $\beta$.] Then,

$$1_{\tilde{A}} = R_\beta(1_{A \times C}) \tag{3.29}$$

and hence

$$\begin{aligned} \Pi'(\tilde{A}) &= \int_{\tilde{A}} R_\beta\left(\frac{d\beta_1}{d\beta}\right) d\tilde{\Pi} \\ &= \int_{\tilde{\Omega}} R_\beta(1_{A\times C}) \cdot R_\beta\left(\frac{d\beta_1}{d\beta}\right) \cdot d\tilde{\Pi} \\ &= \int_E 1_{A\times C} \cdot \frac{d\beta_1}{d\beta} \cdot d\beta \\ &= \beta_1(A \times C). \end{aligned} \tag{3.30}$$

Thus, $(\rho, L, \Pi')$ is a representation of $\beta_1$. Let $R_{\beta_1}$ be the $\beta_1$-lifting corresponding to $(\rho, L, \Pi')$. Then, it follows from the definitions of $R_\beta$, $R_{\beta_1}$ that for a cylinder function $f$,

$$R_\beta(f) = R_{\beta_1}(f). \tag{3.31}$$

Since $\Pi' \ll \tilde{\Pi}$, convergence in $\tilde{\Pi}$-measure for a sequence implies convergence in $\Pi'$-measure for the same sequence. This observation along with (3.31) and Lemma 3.4 implies (ii). For (iii), $R_{\beta_1}(g) = R_\beta(g)$ by (ii) and then $R_{\beta_1}(g) \in \mathscr{L}^1(\tilde{\Omega}, \tilde{\mathscr{A}}, \Pi')$ if and only if $R_{\beta_1}(g) \cdot X \in \mathscr{L}^1(\tilde{\Omega}, \tilde{\mathscr{A}}, \tilde{\Pi})$ which is the same as (3.25). Also

$$\int_{\tilde{\Omega}} R_{\beta_1}(g)\, d\Pi' = \int_{\tilde{\Omega}} R_\beta(g) \cdot R_\beta\left(\frac{d\beta_1}{d\beta}\right) \cdot d\tilde{\Pi},$$

which is the same as (3.26). For (iv), let $g_1 = [d\beta_1/d\beta]^{-1}$. Since $R_\beta(d\beta_1/d\beta) > 0$ a.s. $\tilde{\Pi}$, (iii) in Theorem 3.5 (with $S_1 = \mathbf{R}$ and $U = (0, \infty)$, $S_2 = (0, \infty)$ and $g(x) = 1/x$) implies that

$$\begin{aligned} R_\beta(g_1) &= \left[R_\beta\left(\frac{d\beta_1}{d\beta}\right)\right]^{-1} \\ &= X^{-1} \end{aligned}$$

Thus, for $F \in \mathscr{E}$, taking $f = 1_F \cdot g_1$ and applying (iii) above, we get

$$\int f \cdot \frac{d\beta_1}{d\beta} \cdot d\beta = \int f \cdot d\beta_1,$$

which is the same as

$$\int 1_F \, d\beta = \int 1_F g_1 \, d\beta_1.$$

Since $F \in \mathscr{E}$ is arbitrary, this proves (iv). □

We will later define conditional expectation on the quasicylindrical probability space $(E, \mathscr{E}, \beta)$ given a quasicylindrical mapping.

Let $(E', \mathscr{E}') = (\Omega', \mathscr{A}') \odot (H', \mathscr{C}')$ be another quasicylindrical measurable space, where $(\Omega', \mathscr{A}')$ is a measurable space, $H'$ is a real separable Hilbert space with inner product $(\cdot, \cdot)'$, and $\mathscr{C}' = \mathscr{C}(H')$. Let $\mathscr{P}' = \mathscr{P}(H')$ and for $P' \in \mathscr{P}'$, let

$$\mathscr{C}'_{P'} = \left\{ (P')^{-1} B \colon B \text{ a Borel set in range } P' \right\}$$

and as before, $\mathscr{E}'_{P'} = \mathscr{A}' \otimes \mathscr{C}'_{P'}$.

DEFINITION. A quasicylindrical mapping from $(E, \mathscr{E})$ into $(E', \mathscr{E}')$ is a mapping $\phi\colon E \to E'$ such that for all $P' \in \mathscr{P}'$, there exists a $P \in \mathscr{P}$ for which

$$\phi^{-1}(\mathscr{E}'_{P'}) \subseteq \mathscr{E}_P. \tag{3.32}$$

LEMMA 3.7. *Let $\phi$ be a quasicylindrical mapping from $(E, \mathscr{E}, \beta)$ into $(E', \mathscr{E}')$. Then, the set function $\beta'$ defined by*

$$\beta'(F') = \beta(\phi^{-1} F'), \qquad F' \in \mathscr{E}' \tag{3.33}$$

*defines a quasicylinder probability measure on $(E', \mathscr{E}')$.*

PROOF. Let $P' \in \mathscr{P}'$ and let $\beta'_{P'}$ be the restriction of $\beta'$ to $\mathscr{E}'_{P'}$. Get $P \in \mathscr{P}$ such that (3.32) holds. Then, clearly

$$\beta'_{P'} = \beta_P [\phi]^{-1} \tag{3.34}$$

and hence $\beta'_{P'}$ is countably additive on $\mathscr{E}'_{P'}$. □

The quasicylinder measure $\beta'$ is the measure induced by the quasicylindrical mapping $\phi$ and will also be denoted by $\beta[\phi]^{-1}$. Let $\phi = (\phi_1, \phi_2)$ where $\phi_1\colon E \to \Omega'$ and $\phi_2\colon E \to H'$. The following lemma, stated without proof, gives a representation $R_{\beta'}$ of $\beta'$ closely related to $R_\beta$.

LEMMA 3.8. *Let $h' \in H'$. Then $(\phi_2, h')'$ and $\phi_1$ are cylindrical functions. Let $\rho'$, $L'$ be defined by*

$$\rho' = R_\beta(\phi_1)$$

*and*

$$L'(h') = R_\beta[(\phi_2, h')'],$$

*$h' \in H'$. Then $(\rho', L', \tilde{\Pi})$ is a representation of $\beta'$. $(\rho', L', \tilde{\Pi})$ will be called the representation of $\beta'$ induced by the mapping $\phi$. Further, for any (S-valued)*

cylinder function $g$ on $(E', \mathscr{E}')$,

$$R_{\beta'}(g) = R_\beta(g \circ \phi). \tag{3.35}$$

REMARK 3.3. It is important to observe that (3.35) may not hold for all $g$ belonging to $\mathscr{L}(E', \mathscr{E}', \beta'; S)$. It seems so because the families $\{g_{P'}: P' \in \mathscr{P}'\}$ and $\{(g \circ \phi)_P: P \in \mathscr{P}\}$ are not comparable.

In view of this, we consider the following class. Let $\phi$ be a quasicylindrical mapping from $(E, \mathscr{E}, \beta)$ into $(E', \mathscr{E}')$. Let

$$\begin{aligned} \mathscr{U}(\phi) &= \mathscr{U}(E, \mathscr{E}, \beta; S, \phi) \\ &= \{g \in \mathscr{L}(E', \mathscr{E}', \beta'; S): g \circ \phi \in \mathscr{L}(E, \mathscr{E}, \beta; S) \\ &\qquad\qquad \text{and} \quad R_{\beta'}(g) = R_\beta(g \circ \phi)\}. \end{aligned}$$

As observed earlier in (3.35), $\mathscr{U}(\phi)$ contains cylinder functions.

We are now ready to define conditional expectation of a function $f \in \mathscr{L}^1(E, \mathscr{E}, \beta)$ with respect to a quasicylindrical mapping $\phi$.

DEFINITION. If there exists a function $g \in \mathscr{U}(\phi)$ and $g \circ \phi \in \mathscr{L}^1(E, \mathscr{E}, \beta)$ such that for all $F' \in \mathscr{E}'$,

$$\int_E f 1_{F'}(\phi)\, d\beta = \int_E g \circ \phi \cdot 1_{F'}(\phi)\, d\beta, \tag{3.36}$$

then we define $g \circ \phi$ to be the conditional expectation of $f$ given $\phi$ and write it as

$$E_\beta(f|\phi) = g \circ \phi. \tag{3.37}$$

REMARK 3.4. We have not asserted the existence of a function $g$ satisfying (3.36). Also, in view of the requirement $g \in \mathscr{U}(\phi)$ (3.36) is equivalent to

$$\int_E f 1_{F'}(\phi)\, d\beta = \int_{E'} g 1_{F'}\, d\beta', \qquad \text{for all } F' \in \mathscr{E}'. \tag{3.38}$$

In our earlier papers, we had used (3.38) as the definition of conditional expectation without requiring that $g \in \mathscr{U}(\phi)$.

The following theorem shows that the conditional expectation defined above has the usual properties.

THEOREM 3.9. *Let $\phi$ be a quasicylindrical mapping from $(E, \mathscr{E}, \beta)$ and let $\beta' = \beta\phi^{-1}$. Let $f, f_1, f_2 \in \mathscr{L}^1(E, \mathscr{E}, \beta)$ be such that $E_\beta(f|\phi)$ and $E_\beta(f_i|\phi)$, $i = 1, 2$ exist. Then*

(i) *For $a_1, a_2 \in \mathbf{R}$, $E_\beta(a_1 f_1 + a_2 f_2|\phi)$ exists and*

$$E(a_1 f_1 + a_2 f_2|\phi) = a_1 E_\beta(f_1|\phi) + a_2 E_\beta(f_2|\phi). \tag{3.39}$$

(ii) *Let $g \in \mathscr{U}(\phi)$ be such that $f \cdot g(\phi) \in \mathscr{L}^1(E, \mathscr{E}, \beta)$. Then $[E_\beta(f|\phi)] \cdot g(\phi) \in \mathscr{L}^1(E, \mathscr{E}, \beta)$ and*

$$\int_E f \cdot g(\phi)\, d\beta = \int_E E_\beta(f|\phi) \cdot g(\phi)\, d\beta. \tag{3.40}$$

(iii) *If* $(f)^2 \in \mathscr{L}^1(E, \mathscr{E}, \beta)$, *then* $[E_\beta(f|\phi)]^2 \in \mathscr{L}^1(E, \mathscr{E}, \beta)$.

(iv) *If* $(f)^2 \in \mathscr{L}^1(E, \mathscr{E}, \beta)$, *then*

$$\int_E \left[f - E_\beta(f|\phi)\right]^2 d\beta = \min_{g \in \mathscr{U}(\phi)} \int_E [f - g(\phi)]^2 d\beta. \tag{3.41}$$

PROOF. (i) Let $g_i = E_\beta(f_i|\phi)$, $i = 1, 2$. Then linearity of $R_\beta, R_{\beta'}$ [Theorem 3.5(i)] implies that $a_1 g_1 + a_2 g_2 \in \mathscr{U}(\phi)$. Also, for $F' \in \mathscr{E}'$,

$$\begin{aligned}\int_E 1_{F'}(\phi) \cdot [a_1 g_1(\phi) + a_2 g_2(\phi)]\, d\beta &= a_1 \int_E 1_{F'}(\phi) g_1(\phi)\, d\beta + a_2 \int_E 1_{F'}(\phi) g_2(\phi)\, d\beta \\ &= a_1 \int_E 1_{F'}(\phi) f_1\, d\beta + a_2 \int_E 1_{F'}(\phi) f_2\, d\beta \\ &= \int_E 1_{F'}(\phi) \cdot [a_1 f_1 + a_2 f_2]\, d\beta\end{aligned}$$

and hence (3.39) holds.

Let $\mathscr{D}_\phi$ be the $\sigma$ field on $\tilde{\Omega}$ generated by the family $\mathscr{F} = \{A \in \tilde{\mathscr{A}}: 1_A = R_\beta(1_{F'}(\phi))$ for some $F' \in \mathscr{E}'\}$ augmented by $\tilde{\Pi}$ null sets. Since $\mathscr{E}'$ is a field and the mapping $R_\beta$ is linear and multiplicative on $\mathscr{L}(E, \mathscr{E}, \beta; \mathbf{R})$, $\mathscr{F}$ is a field and hence from the definition of conditional expectation, for all $D \in \mathscr{D}_\phi$,

$$\int_D R_\beta(f)\, d\tilde{\Pi} = \int_D R_\beta\left[E_\beta(f|\phi)\right] d\tilde{\Pi}. \tag{3.42}$$

We will show first that $g \in \mathscr{U}(\phi)$ implies that $R_\beta(g(\phi))$ is $\mathscr{D}_\phi$-measurable.

If $g$ is a bounded real-valued cylinder function on $(E', \mathscr{E}')$, say $g^{-1}(\mathscr{E}'_{P'}) \subseteq \mathscr{B}(\mathbf{R}^1)$, then $g$ can be uniformly approximated by $\mathscr{E}'_{P'}$-measurable simple functions $g_k$. Since $R_{\beta'}(g_k) = R_\beta(g_k \circ \phi)$, and $g_k$ is a $\mathscr{E}'_{P'}$-measurable simple function, it follows that $R_{\beta'}(g_k)$ is $\mathscr{D}_\phi$-measurable. Now $g_k \to g$ uniformly and hence (see Theorem 3.5) $R_{\beta'}(g_k)$ converges to $R_{\beta'}(g)$ in $\tilde{\Pi}$-probability and hence $R_{\beta'}(g)$ is $\mathscr{D}_\phi$-measurable. Let $g$ be any cylinder function. Let $\tau_k: \mathbf{R} \to \mathbf{R}$ be defined by $\tau_k(x) = x$ if $-k \le x \le k$ and $\tau_k(x) = (\operatorname{sgn} x) \cdot k$ if $|x| \ge k$. Then

$$R_{\beta'}(\tau_k(g)) = \tau_k(R_{\beta'}(g))$$

and hence $R_{\beta'}(\tau_k(g))$ converges to $R_{\beta'}(g)$. Thus $R_{\beta'}(g)$ is $\mathscr{D}_\phi$-measurable.

Now, let $g \in \mathscr{U}(\phi)$ be arbitrary. Then, get $P'_k \in \mathscr{P}'$ such that $R_{\beta'}(g_{P'_k})$ converges to $R_{\beta'}(g)$ in $\tilde{\Pi}$-probability. Since $R_{\beta'}(g_{P'_k})$ is $\mathscr{D}_\phi$-measurable for all $k \ge 1$, it follows that $R_{\beta'}(g)$ is $\mathscr{D}_\phi$-measurable. Since $g \in \mathscr{U}(\phi)$, $R_\beta(g(\phi)) = R_{\beta'}(g)$. Thus, for any $g \in \mathscr{U}(\phi)$, $R_\beta(g(\phi))$ is $\mathscr{D}_\phi$-measurable. Now, (3.42) implies that

$$E_{\tilde{\Pi}}\left(R_\beta(f)|\mathscr{D}_\phi\right) = R_\beta\left(E_\beta(f|\phi)\right). \tag{3.43}$$

Now (ii), (iii), and (iv) follow from (3.43), the observation that $R_\beta(g(\phi))$ is $\mathscr{D}_\phi$-measurable for all $g \in \mathscr{U}(\phi)$ and the standard properties of conditional expectation in the countably additive theory. □

**4. The abstract statistical model and the Bayes formula.** We begin by recalling the definition of the canonical Gauss measure $m$ on $(H, \mathscr{C})$ [see Gross (1962)]. Fix $P \in \mathscr{P}$. Let dimension (range $P$) be $k$ and let $\{e_1, e_2, \ldots, e_k\}$ be an orthonormal (ON) basis for range $P$. For $C \in \mathscr{C}_P$ given by

$$C = \{h: (h, e_1), \ldots, (h, e_k) \in B\} \tag{4.1}$$

for $B \in \mathscr{B}(\mathbb{R}^k)$, define

$$m_P(C) = \left(\frac{1}{\sqrt{2\pi}}\right)^k \int_{\mathbb{R}^k} 1_B(\underline{x}) \exp\left(-\frac{1}{2}\sum_{j=1}^k x_i^2\right) d\underline{x}, \tag{4.2}$$

where $d\underline{x}$ denotes integration wrt the Lebesgue measure on $\mathbb{R}^k$. Using the rotational invariance of the normal distribution or Gauss measure on finite-dimensional Euclidean spaces, it can be checked that $m_P$ does not depend upon the choice of $\{e_1, \ldots, e_k\}$ and that the family $\{m_P: P \in \mathscr{P}\}$ is consistent. Let $m$ on $(H, \mathscr{C})$ be defined by $m = m_P$ on $\mathscr{C}_P$ for $P \in \mathscr{P}$. Then by its construction, $m$ is a (finitely additive) cylinder probability measure on $(H, \mathscr{C})$. $m$ is called the canonical Gauss measure on $(H, \mathscr{C})$. The measure $m$ is determined by its characteristic functional

$$\int_H \exp\{i(h, h_1)\}\, dm(h) = \exp\left(-\tfrac{1}{2}|h_1|^2\right), \qquad \text{for all } h_1 \in H. \tag{4.3}$$

*Let $(\Omega, \mathscr{A}, \Pi)$ be a countably additive probability space and let $\xi: (\Omega, \mathscr{A}, \Pi) \to (H, \mathscr{B}(H))$ be a measurable mapping.* Let $(E, \mathscr{E}) = (\Omega, \mathscr{A}) \odot (H, \mathscr{C})$. For $P \in \mathscr{P}$, let $\alpha_P$ be the product of $\Pi$ and $m_P$ on $\mathscr{E}_P = \mathscr{A} \otimes \mathscr{C}_P$. Then, it is easy to see that $\{\alpha_P: P \in \mathscr{P}\}$ is a consistent family and determines a QCP $\alpha$ on $(E, \mathscr{E})$ such that $\alpha = \alpha_P$ on $\mathscr{E}_P$. Clearly, the conditions of Theorem 3.1 are satisfied with $\lambda(\omega, \cdot) \equiv m$.

We shall be working henceforth with a particular representation of $\alpha$ constructed as follows: First let $(L_0, \Pi_1)$ be a representation of $m$ whose existence and special properties are assured by the following result.

THEOREM 4.1. *There exists a representation $(L_0, \Pi_1)$ where $\Pi_1$ is a probability measure on $(\Omega_1, \mathscr{A}_1)$ of $m$ s.t. the mapping $(\omega_1, h) \to L_0(h)(\omega_1)$ is measurable wrt $\mathscr{A}_1 \otimes \mathscr{B}(H)$. Further, if $\{e_j\}$ is any complete orthonormal basis (CONS) of $H$, then*

$$L_0(h) = \sum_{j=1}^{\infty} L_0(e_j)(h, e_j), \qquad \Pi_1\text{-a.s.} \tag{4.4}$$

*Now taking $(\tilde{\Omega}, \tilde{\mathscr{A}}, \tilde{\Pi}) = (\Omega, \mathscr{A}, \Pi) \otimes (\Omega_1, \mathscr{A}_1, \Pi_1)$, define $\rho(\tilde{\omega}) = \omega$ and $L(h)(\tilde{\omega}) = L_0(h)(\omega_1)$ for $\tilde{\omega} = (\omega, \omega_1)$. Then $(\rho, L, \tilde{\Pi})$ is the desired representation of $\alpha$. It is easy to see that* (4.4) *implies*

$$L(h)(\tilde{\omega}) = \sum_{j=1}^{\infty} L(e_j)(\tilde{\omega}) \cdot (h, e_j), \qquad \tilde{\Pi}\text{-a.s.} \tag{4.5}$$

*Let $e$ be the identity mapping on $H$. Considered as a mapping from $(H, \mathscr{C}, m)$ into $(H, \mathscr{C})$, $e$ will be called* Gaussian white noise.

*Abstract statistical model.* We will adopt the usual convention of regarding a function defined on either $H$ or $\Omega$ as defined on $E$ itself. With this convention, let $y$: $E \to H$ be defined by

$$y = \xi + e \tag{4.6}$$

so that for $(\omega, h) \in H$, $y(\omega, h) = \xi(\omega) + e(h) = \xi(\omega) + h$.

We shall call (4.6) the abstract statistical model. It is the prototype of the filtering model to be considered in Section 5. It is convenient to list below the problems that need to be resolved before a satisfactory theory can emerge:

(1) Is it possible to speak of the distribution of the "observation" $y$ in the abstract model (4.6)? In other words, define the induced measure $n = \alpha[Y]^{-1}$ or, more generally, the induced measure $\alpha[Qy]^{-1}$, where $Q$ is an arbitrary orthogonal projection on $H$. In view of the work of the previous section, this can be achieved by showing that $y$ (resp. $Qy$) is a QCM from $(E, \mathscr{E}, \alpha)$ into $(H, \mathscr{C})$ [resp. into $(H', \mathscr{C}')$ where $H' = QH$ and $\mathscr{C}' = \mathscr{C}(H')$].

(2) To show that $n$ is absolutely continuous wrt $m$ and to evaluate the Radon-Nikodym (R-N) derivative $dn/dm$.

(3) For an integrable function $g$ on $(\Omega, \mathscr{A}, \Pi)$ considered as a function on $(E, \mathscr{E}, \alpha)$ and for an orthogonal projection $Q$ on $H$, show that the conditional expectation $E_\alpha(g|Qy)$ exists and obtain a formula for it. The latter is the Bayes formula which is the basic statistical tool in our theory and is the finitely additive analogue of the formula obtained by Kallianpur and Striebel (1968) in the stochastic calculus treatment of the filtering problem.

Before we can settle these questions we need several preparatory results. Of these, Theorems 4.2 and 4.4 are concerned with mixtures of translates of the canonical Gauss measure $m$ on $(H, \mathscr{C})$ and their R-N derivatives wrt $m$.

Let $\mathscr{I}(H)$ be the class of real-valued functions $f$ on $H$ of the form

$$f(h) = \int_H \exp\left((h,k) - \tfrac{1}{2}|k|^2\right) d\nu(k), \qquad h \in H \tag{4.7}$$

for some $\nu \in \mathscr{M}_0(H)$ where $\mathscr{M}_0(H)$ is the class of countably additive probability measures on $(H, \mathscr{B}(H))$. Since $(h, k) - \frac{1}{2}|k|^2 \le |h|^2$, the integral in (4.7) exists for each $h \in H$.

THEOREM 4.2.

(i) $\mathscr{I}(H) \subseteq \mathscr{L}^{1*}(H, \mathscr{C}, m)$.

(ii) *If* $f \in \mathscr{I}(H)$ *is given by* (4.7), *then*

$$R_m(f) = \int_H \exp\left(L_0(k) - \tfrac{1}{2}|k|^2\right) d\nu(k). \tag{4.8}$$

PROOF. Fix $\{P_i\} \subseteq \mathscr{P}$, $P_i \xrightarrow{s} I$. Recall that $(L_0, \Pi_1)$ is a fixed representation of $m$, so that under $\Pi_1$, $L_0(k)$ is a normal random variable with mean zero and variance $|k|^2$. Also, if $k_1 \perp k_2$, then $L_0(k_1)$ and $L_0(k_2)$ are independent. Now, for a fixed $i$, if $\{e_1, e_2, \ldots, e_j\}$ is a basis of range $P_i$,

$$\begin{aligned} f_{P_i}(h) &= f(P_i h) \\ &= \int_H \exp\left((P_i h, k) - \tfrac{1}{2}|k|^2\right) d\nu(k) \\ &= \int \exp\left(\sum_{r=1}^{j} (h, e_r)(k, e_r) - \tfrac{1}{2}|k|^2\right) d\nu(k) \end{aligned}$$

and hence by the definition of the $m$-lifting $R_m$ for cylinder functions,

$$\begin{aligned} R_m(f_{P_i})(\omega_1) &= \int \exp\left(\sum_{r=1}^{j} L_0(e_r)(\omega_1)(k, e_r) - \tfrac{1}{2}|k|^2\right) d\nu(k) \\ &= \int \exp\left(L_0(P_i k)(\omega_1) - \tfrac{1}{2}|k|^2\right) d\nu(k), \qquad \Pi\text{-a.s.} \end{aligned} \tag{4.9}$$

as $P_i k = \sum_{r=1}^{j} e_r(k, e_r)$ and $L_0$ satisfies (3.3).

Let

$$V(\omega_1) = \int \exp\left(L_0(k)(\omega_1) - \tfrac{1}{2}|k|^2\right) d\nu(k). \tag{4.10}$$

Since $(k, \omega_1) \to L_0(k)(\omega_1)$ is $\mathscr{B}(H) \otimes \mathscr{A}_1$-measurable, the integral in (4.10) is well defined. To complete the proof, we will show that $R_m(f_{P_i}) \to V$ in $\mathscr{L}^1(\Omega_1, \mathscr{A}_1, \Pi_1)$.

Let

$$U_i = \int_{\Omega_1} \int_H \left| \exp\left(L_0(P_i k) - \tfrac{1}{2}|k|^2\right) - \exp\left(L_0(k) - \tfrac{1}{2}|k|^2\right) \right| d\nu(k)\, d\Pi_1. \tag{4.11}$$

Then clearly

$$\int_{\Omega_1} \left| R_m(f_{P_i}) - V \right| d\Pi_1 \leq U_i$$

and thus it suffices to show that $U_i \to 0$. Let $P_i^{\perp}$ be the orthogonal complement of $P_i$. Then since $k = P_i k + P_i^{\perp} k$, we have, using (3.3)

$$L_0(k) = L_0(P_i k) + L_0(P_i^{\perp} k).$$

This and Fubini's theorem imply that

$$U_i = \int_H \left[ \int_{\Omega_1} \exp\left(L_0(P_i k) - \tfrac{1}{2}|P_i k|^2\right) \exp\left(-\tfrac{1}{2}|P_i^{\perp} k|^2\right) \cdot \left|1 - \exp\left(L_0(P_i^{\perp} k)\right)\right| d\Pi_1 \right] d\nu.$$

As observed earlier, under $\Pi_1$, for fixed $k$, $L_0(P_i k)$ and $L_0(P_i^{\perp} k)$ are independent and further, $\int_{\Omega_1} \exp(L_0(P_i k) - \frac{1}{2}|P_i k|^2)\, d\Pi_1 = 1$. Thus,

$$U_i = \int_H \int_{\Omega_1} \exp\left(-\tfrac{1}{2}|P_i^{\perp} k|^2\right) \left|1 - \exp\left(L_0(P_i^{\perp} k)\right)\right| d\Pi_1\, d\nu. \tag{4.12}$$

For $\varepsilon > 0$, let

$$\delta(\varepsilon) = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} |1 - \exp(\varepsilon x)| \exp\left(-\frac{1}{2}|x|^2\right) dx.$$

Then, $\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$ and $\delta(\varepsilon) \leq 1 + e^{\varepsilon^2/2}$. Also, under $\Pi_1$, the distribution of $L_0(P_i^{\perp} k)$ is normal with mean zero and variance $|P_i^{\perp} k|^2$. Hence

$$\int_{\Omega_1} \left|1 - \exp\left(L_0\left(P_i^{\perp} k\right)\right)\right| d\Pi_1 = \delta\left(|P_i^{\perp} k|\right).$$

From (4.12) we then have

$$U_i = \int_H \exp\left(-\tfrac{1}{2}|P_i^{\perp} k|^2\right) \delta\left(|P_i^{\perp} k|\right) d\nu(k). \tag{4.13}$$

As $\delta(|P_i^{\perp} k|) \leq 1 + \exp(\frac{1}{2}|P_i^{\perp} k|^2)$, the integrand in (4.13) is dominated by 2. Since $P_i \xrightarrow{s} I$, $|P_i^{\perp} k| \to 0$ for all $k \in H$ the integrand in (4.13) goes to zero for all $k \in H$. Since $\nu$ is a finite measure, the dominated convergence theorem implies

$$U_i \to 0. \tag{4.14}$$

□

For $C \in \mathscr{C}$ and $h \in H$, let $C - h \subseteq H$ be defined by

$$C - h = \{h_0 \colon h_0 + h \in C\}.$$

If $C \in \mathscr{C}_P$ is given by $C = P^{-1}B$, $B \in \mathscr{B}(\text{range } P)$, then

$$\begin{aligned} C - h &= \{h_0 \colon h_0 + h \in C\} \\ &= \{h_0 \colon Ph_0 + Ph \in B\} \\ &= \left\{h_0 \colon h_0 \in P^{-1}(B - Ph)\right\} \end{aligned} \tag{4.15}$$

and hence

$$C - h \in \mathscr{C}_P. \tag{4.16}$$

Also, (4.15) implies that for $C \in \mathscr{C}_P$,

$$C - h = C - Ph. \tag{4.17}$$

LEMMA 4.3. *Let $h \in H$ and let $m'$ be defined by*

$$m'(C) = m(C - h), \qquad \text{for } C \in \mathscr{C}. \tag{4.18}$$

*Then $m'$ is a cylinder measure and $m' \ll m$ with*

$$\frac{dm'}{dm}(k) = \exp\left((h, k) - \tfrac{1}{2}|h|^2\right). \tag{4.19}$$

REMARK 4.1. From the properties of Gaussian measures on finite-dimensional linear spaces, it also follows that for $C \in \mathscr{C}_P$, the mapping

$$h' \to m_P(C - h')$$

from range $P$ into $\mathbf{R}$ is continuous and thus for all $C \in \mathscr{C}$, the mapping

$$h \to m(C - h)$$

is continuous from $H$ into $\mathbf{R}$. In particular, $h \to m(C - h)$ is $\mathscr{B}(H)$-measurable.

THEOREM 4.4. *Let $\mu \in \mathscr{M}(H)$.*

(i) *The set function $n$: $\mathscr{C} \to \mathbf{R}$ defined by*

$$m^{\mu}(C) = \int_H m(C - h)\,d\mu(h) \tag{4.20}$$

*is a cylinder measure on $(H, \mathscr{C})$.*

(ii) $m^{\mu} \ll m$ *and*

$$\frac{dm^{\mu}}{dm}(k) = \int_H \exp\left((h,k) - \tfrac{1}{2}|h|^2\right) d\mu(h). \tag{4.21}$$

(iii) $\mathscr{I}(H) \subseteq \mathscr{L}^*(H, \mathscr{C}, m^{\mu})$.

(iv) *There exists a representation $(L', \Pi')$ of $m^{\mu}$ [$\Pi'$ is a probability measure on $(\Omega', \mathscr{A}')$] such that*

$$(h, \omega') \to L'(h)(\omega') \quad \text{is } \mathscr{A}' \otimes \mathscr{B}(H)\text{-measurable}. \tag{4.22}$$

(v) *If $(L', \Pi')$ is any representation of $n$ satisfying (4.22), and $R'_{m^{\mu}}$ is the corresponding $m^{\mu}$-lifting then for $f \in \mathscr{I}(H)$ given by (4.8), we have*

$$R'_{m^{\mu}}(f)(\omega') = \int_H \exp\left(L'(k)(\omega') - \tfrac{1}{2}|k|^2\right) d\nu(k). \tag{4.23}$$

We now return to the abstract statistical model (4.6) and the questions raised at the beginning of this section.

Let $Q$ be an orthogonal projection on $H$. Let $H'$ be the range of $Q$. $H'$ itself is a Hilbert space with the inner product $(\cdot,\cdot)'$, which is the restriction of $(\cdot,\cdot)$ to $H' \times H'$. Let $\mathscr{C}' = \mathscr{C}(H')$ and $\mathscr{P}' = \mathscr{P}(H')$.

LEMMA 4.5. *$Q$: $H \to H'$ is a quasicylindrical mapping and the induced measure $m[Q]^{-1}$ is the canonical Gauss measure $m'$ on $H'$.*

The first two questions posed at the beginning of the section are answered by the following result:

THEOREM 4.6.

(i) *$Qy$ is a quasicylindrical mapping from $(E, \mathscr{E}, \alpha)$ into $(H', \mathscr{C}')$.*

(ii) *Let $n' = \alpha[Qy]^{-1}$. Then $n' \ll m'$ and*

$$\frac{dn'}{dm'}(h') = \int_\Omega \exp\left((h', Q\xi(\omega)) - \tfrac{1}{2}|Q\xi(\omega)|^2\right) d\Pi(\omega), \qquad h' \in H'. \tag{4.24}$$

*In particular, if $n = \alpha[y]^{-1}$, then $n \ll m$ and*

$$\frac{dn}{dm}(h) = \int_\Omega \exp\left((h, \xi(\omega)) - \tfrac{1}{2}|\xi(\omega)^2|\right) d\Pi(\omega), \qquad h \in H. \tag{4.25}$$

PROOF. Let $C_1 \in \mathscr{C}'_{P_1} \subseteq \mathscr{C}'$, for $P_1 \in \mathscr{P}'$ be given by

$$C_1 = \{h' \colon P_1 h' \in B\}. \tag{4.26}$$

$B \in \mathscr{B}(\text{range } P_1)$. Then $P = P_1 Q \in \mathscr{P}$ and

$$\begin{aligned} D &= \{(\omega, h): Qy(\omega, h) \in C_1\} \\ &= \{(\omega, h): Py(\omega, h) \in B\} \\ &= \{(\omega, h): P\xi(\omega) + Ph \in B\} \\ &\in \mathscr{E}_P. \end{aligned} \tag{4.27}$$

Thus, $Qy$ is a QCM from $(E, \mathscr{E}, \alpha)$ into $(H', \mathscr{C}')$. Now, recall that $\alpha_P$, which is the restriction of $\alpha$ to $\mathscr{E}_P$, is countably additive and is equal to $\Pi \otimes m_P$. Thus, using Fubini's theorem, we have from the definition of $n'$,

$$\begin{aligned} n'(C_1) &= \alpha(D) \\ &= \alpha_P(D) \\ &= \int_\Omega m_P(h: Ph + P\xi(\omega) \in B)\, d\Pi(\omega) \\ &= \int_\Omega m\big(h: Qh + Q\xi(\omega) \in P_1^{-1}B\big)\, d\Pi(\omega) \qquad \text{as } P = P_1 Q \\ &= \int_\Omega m'(h': h' + Q\xi(\omega) \in C_1)\, d\Pi(\omega) \\ &\qquad\qquad\qquad\qquad\qquad \text{as } m' = m[Q]^{-1}, C_1 = P_1^{-1}B \\ &= \int_\Omega m'(C_1 - Q\xi(\omega))\, d\Pi(\omega). \end{aligned} \tag{4.28}$$

Thus, if $\mu \in \mathscr{M}(H')$ is given by $\mu = \Pi[Q\xi]^{-1}$, then

$$n'(C_1) = \int_{H'} m'(C_1 - k)\, d\mu(k) \tag{4.29}$$

and now (4.24) follows from Theorem 4.4. Of course, (4.25) is a special case of (4.24). □

REMARK 4.2. Let $C_1 \in \mathscr{C}'$. Then

$$\begin{aligned} n(Q^{-1}C_1) &= \alpha\big((\omega, h): \xi(\omega) + h \in Q^{-1}C_1\big) \\ &= \alpha\big((\omega, h): Q\xi(\omega) + Qh \in C_1\big) \\ &= n'(C_1). \end{aligned} \tag{4.30}$$

Thus,

$$n' = n[Q]^{-1}. \tag{4.31}$$

LEMMA 4.7. *Let $g$: $H' \to S$ and let*

$$g_1(h) = g(Qh), \qquad h \in H. \tag{4.32}$$

*Then*

(i) $g_1 \in \mathscr{L}^*(H, \mathscr{C}, m; S) \Rightarrow g \in \mathscr{U}(Q) = \mathscr{U}(H, \mathscr{C}, m; S, Q)$

(ii) $g_1 \in \mathscr{L}^*(H, \mathscr{C}, n; S) \Rightarrow g \in \mathscr{U}(Q) = \mathscr{U}(H, \mathscr{C}, n; S, Q)$

(iii) $g_1 \in \mathscr{L}^*(H, \mathscr{C}, n; S)$ *and* $g_1 \in \mathscr{U}(y) = \mathscr{U}(E, \mathscr{E}, \alpha; S, y) \Rightarrow g \in \mathscr{U}(Qy) = \mathscr{U}(E, \mathscr{E}, \alpha; S, Qy)$.

THEOREM 4.8. *Let $g \in \mathscr{I}(H')$. Then*

$$g \in \mathscr{U}(Qy) = \mathscr{U}(E, \mathscr{E}, \alpha; Qy, \mathbf{R})$$

*and*

$$g(Qy) \in \mathscr{L}^*(E, \mathscr{E}, \alpha; \mathbf{R}).$$

We are now in a position to prove the main result, which answers the last question raised at the beginning of this section.

THEOREM 4.9 (Bayes formula). *Let $g$ be an integrable function on $(\Omega, \mathscr{A}, \Pi)$. Then, $E_\alpha(g|Qy)$ exists and*

$$E_\alpha(g|Qy) = \frac{\sigma_Q(g, Qy)}{\sigma_Q(l, Qy)}, \tag{4.33}$$

*where for $h' \in H'$,*

$$\sigma_Q(g, h') = \int_\Omega g(\omega)\exp\left((h', Q\xi(\omega)) - \tfrac{1}{2}|Q\xi(\omega)|^2\right) d\Pi(\omega) \tag{4.34}$$

*and*

$$\sigma_Q(1, h) = \int_\Omega \exp\left((h', Q\xi(\omega)) - \tfrac{1}{2}|Q\xi(\omega)|^2\right) d\Pi(\omega).$$

PROOF. Because of the linearity of $E_\alpha(\cdot|Qy)$ (Theorem 3.8) it suffices to prove (4.33) for positive $g$ such that $\int g\, d\Pi = 1$. For $C_1 \in \mathscr{C}'$, let

$$\phi_g(C_1) = \int_E g(\omega) 1_{C_1}(Qy(\omega, h))\, d\alpha(\omega, h). \tag{4.35}$$

We will first show that

$$\phi_g(C_1) = \int_\Omega g(\omega) m'(C_1 - Q\xi(\omega))\, d\Pi(\omega). \tag{4.36}$$

The proof of (4.36) follows the proof of (4.28) in Theorem 4.6. If $C_1$ is given by (4.26) for $P_1 \in \mathscr{P}'$ and $P = P_1 Q$, then

$$\begin{aligned}
\phi_g(C_1) &= \int_E g(\omega) 1_{C_1}(Qy(\omega, h))\, d\alpha(\omega, h) \\
&= \int_E g(\omega) 1_B(P\xi(\omega) + Ph)\, d\alpha_P(\omega, h) \\
&= \int_\Omega g(\omega)\left[m_P(h\colon Ph + P\xi(\omega) \in B)\right] d\Pi(\omega) \\
&= \int_\Omega g(\omega) m'(C_1 - Q\xi(\omega))\, d\Pi(\omega).
\end{aligned} \tag{4.37}$$

Hence, if $\mu' \in \mathscr{M}_0(H')$ is defined by

$$\mu'(A) = \int_\Omega g(\omega) 1_A(\xi(\omega))\, d\Pi(\omega), \qquad A \in \mathscr{B}(H), \tag{4.38}$$

then from (4.36)

$$\phi_g(C_1) = \int_{H'} m'(C_1 - k)\, d\mu'(k)$$

and thus by Theorem 4.4, $\phi_g$ is a cylinder probability measure on $(H', \mathscr{C}')$, $\phi_g \ll m'$, and

$$\begin{aligned} \frac{d\phi_g}{dm'}(h') &= \int_H \exp\left((h', k) - \tfrac{1}{2}|k|^2\right) d\mu'(k) \\ &= \int_\Omega g(\omega) \exp\left((h', Q\xi(\omega)) - \tfrac{1}{2}|Q\xi(\omega)|^2\right) d\Pi(\omega) \\ &= \sigma_Q(g, h'). \end{aligned} \tag{4.39}$$

Also, as observed in Theorem 4.6 [Equation (4.24)],

$$\frac{dn'}{dm'}(h) = \sigma_Q(1, h'). \tag{4.40}$$

Also, Theorem 4.2 implies that $R_{m'}(dn'/dm') > 0$ a.s.

Thus, by Theorem 3.6(iv), $\phi_g \ll n'$ and

$$\frac{d\phi_g}{dn'}(h') = \frac{\sigma_Q(g, h')}{\sigma_Q(1, h')} \tag{4.41}$$

and hence

$$\phi_g(C_1) = \int_{H'} 1_{C_1}(h') \frac{\sigma_Q(g, h')}{\sigma_Q(1, h')}\, dn'(h'). \tag{4.42}$$

By Theorem 4.8, $\sigma_Q(g, \cdot)$ and $\sigma_Q(1, \cdot)$ belong to $\mathscr{U}(Qy)$. Since $R_{m'}(\sigma_Q(1, \cdot) > 0$ a.s., $R_{n'}(\sigma_Q(1, \cdot) > 0$ a.s. Hence

$$\frac{\sigma_Q(g, \cdot)}{\sigma_Q(1, \cdot)} \in \mathscr{U}(Qy). \tag{4.43}$$

(4.42) and (4.43) imply that for $C_1 \in \mathscr{C}'$,

$$\phi_g(C_1) = \int_E 1_{C_1}(Qy) \frac{\sigma_Q(g, Qy)}{\sigma_Q(l, Qy)}\, d\alpha. \tag{4.44}$$

Now, (4.35), (4.43), and (4.44) imply that $E_\alpha(g|Qy) = [\sigma_Q(g, Qy)]/[\sigma_Q(1, Qy)]$. □

**5. Equations for the optimal filter: $\mathbb{R}^d$-valued Markov signal process.** In this section we formulate and solve the nonlinear filtering problem when the signal process $X = (X_t)$ is an $\mathbb{R}^d$-valued Markov process wrt a family $(\mathscr{A}_t)$ given on a countably additive probability space $(\Omega, \mathscr{A}, \Pi)$. It is assumed that $(X_t)$

admits a transition probability function and that the paths $X_t$ are progressively measurable wrt $(\mathscr{A}_t)$, but $(X_t)$ will not be restricted to be time homogeneous. Let $m \geq 1$ be an integer and let $h$: $[0,T] \times \mathbf{R}^d \to \mathbf{R}^m$ be Borel-measurable such that

$$\int_0^T |h_t(X_t)|^2 \, dt < \infty, \qquad \Pi\text{-a.s.} \tag{5.1}$$

Here and in what follows, $|\cdot|$ denotes the norm on $\mathbf{R}^m$ and, generally on any Euclidean space.

It is necessary first, to recast the abstract nonlinear filtering model and the Bayes formula in a form which explicitly brings into focus the role played by the parameter $t$. Let

$$H = \{\eta = (\eta^1, \eta^2, \dots \eta^m): [0,T] \to R^m, \, |\eta| \in L^2[0,T]\}.$$

Then $H$ is a real separable Hilbert space with the inner product

$$(\eta, \bar{\eta}) = \int_0^T \sum_{j=1}^m \eta_t^j \bar{\eta}_t^j \, dt \tag{5.2}$$

and norm $\|\eta\| = (\eta, \eta)^{1/2}$. Let $\xi_s(\omega) = h_s(X_s(\omega)), 0 \leq s \leq T$. In view of (5.1), for $\Pi$ almost all $\omega$, we have $\xi(\omega)$: $[0,T] \to \mathbf{R}^m$ with $|\xi(\omega)| \in L^2[0,T]$ so that $\xi(\omega) \in H$. Since no loss of generality is involved we will assume that these relations hold for all $\omega$. The progressive measurability of $(X_t)$ and the conditions on $h$ imply that the map $\omega \to \xi(\omega)$ is $(\mathscr{A}, \mathscr{B}(H))$-measurable, where $\mathscr{B}(H)$ is the $\sigma$-field of Borel sets in $H$.

Let $(E, \mathscr{E}, \alpha)$ be the quasicylindrical probability space constructed in Section 3. The abstract statistical model (4.6)

$$y = \xi + e$$

now takes the form

$$y_s = h_s(X_s) + e_s, \qquad 0 \leq s \leq T, \tag{5.3}$$

where $e = (e_t)$ is the Gaussian white noise independent of the signal $(X_t)$ and $y_t$ is the observation process.

Let $Q_t$ for $0 < t \leq T$ denote the orthogonal projection on $H$ with range

$$H_t = \left\{\eta \in H: \int_t^T |\eta_s|^2 \, ds = 0\right\}, \tag{5.4}$$

so that for $\eta \in H$,

$$\begin{aligned} (Q_t\eta)(s) &= \eta(s), && \text{for } 0 \leq s \leq t, \\ &= 0, && \text{for } t < s \leq T. \end{aligned}$$

Then, $Q_t y$ represents the observation $\{y_s: 0 \leq s \leq t\}$ over the time interval $[0,t]$. Applying the Bayes formula Theorem 4.9 for $Q = Q_t$, we obtain the conditional expectation formula for an integrable function $g$ on $(\Omega, \mathscr{A}, \Pi)$

$$E_\alpha(g|Q_t y) = \frac{\bar{\sigma}_t(g, y)}{\bar{\sigma}_t(1, y)}, \tag{5.5}$$

expectation formula for an integrable function $g$ on $(\Omega, \mathscr{A}, \Pi)$

$$\tilde{\sigma}_t(f, \eta) = \int_\Omega f(\omega)\exp\left((\eta, Q_t\xi(\omega)) - \tfrac{1}{2}\|Q_t\xi(\omega)\|^2\right) d\Pi(\omega)$$

(5.6)
$$= \int_\Omega f(\omega)\exp\left(\sum_{j=1}^m \int_0^t \eta_s^j h_s^j(X_s(\omega))\, ds - \tfrac{1}{2}\int_0^t |h_s(X_s(\omega))|^2\, ds\right) d\Pi(\omega)$$

For a Borel-measurable function $f$: $\mathbf{R}^d \to \mathbf{R}$, such that $f(X_t)$ is integrable, writing

$$\sigma_t(f, \eta) = \tilde{\sigma}_t(f(X_t), \eta) \tag{5.7}$$

we have

$$E_\alpha(f(X_t)|Q_t y) = \frac{\sigma_t(f, y)}{\sigma_t(1, y)}. \tag{5.8}$$

Thus, the family $\{\sigma_t(f, \eta)\}$ determines the conditional distribution of $X_t$ given $\{y_s: 0 \le s \le t\}$.

We shall now derive an evolution equation for $\{\sigma_t(f, \eta)\}$. This is analogous to the Zakai equation in the usual nonlinear filtering theory.

Let $\mathscr{L}$ be the extended generator of $X_t$ and let $\mathscr{D}$ be its domain. [See the Appendix for the definitions of $\mathscr{L}$ and $\mathscr{D}$. In the standard terminology, $\mathscr{L}$ is the generator of the $[0, \infty) \times \mathbf{R}^d$-valued time homogeneous Markov process $(t, X_t)$ and $\mathscr{D}$ is the domain of $\mathscr{L}$.] Let $\mathscr{D}_0$ be the class of functions $f$ from $\mathbf{R}^d$ into $\mathbf{R}$ such that $f_1$ defined by

$$f_1(s, x) = f(x), \qquad (s, x) \in [0, \infty) \times \mathbf{R}^d, \tag{5.9}$$

belongs to $\mathscr{D}$. For $f \in \mathscr{D}_0, 0 \le t < \infty$, define $\mathscr{L}_t$ by

$$(\mathscr{L}_t f)(x) = (\mathscr{L} f_1)(t, x), \tag{5.10}$$

where $f_1$ is given by (5.9). Observe that by Lemma A.1, for $f \in \mathscr{D}_0$,

$$\left\{f(X_t) - \int_0^t (\mathscr{L}_s f)(X_s)\, ds, \mathscr{A}_t\right\} \quad \text{is a martingale.} \tag{5.11}$$

The following theorem gives the white noise version of the Zakai equation for the unnormalized conditional expectation $\sigma_t(f, y)$.

THEOREM 5.1. *Suppose that*

$$E\int_0^T |h_s(X_s)|^2\, ds < \infty. \tag{5.12}$$

*Then, for all $y \in H$ and for all $f \in \mathscr{D}_0$*

$$\frac{d}{dt}\sigma_t(f, y) = \sigma_t(\mathscr{L}_t f, y) + \sum_{j=1}^m \sigma_t(h_t^j f, y) y_t^j - \tfrac{1}{2}\sigma_t(|h_t|^2 f, y), \tag{5.13}$$

*for a.e. $t$*

PROOF. Let $q_t(\eta, \omega)$ for $\eta \in H$, $\omega \in \Omega$, and $0 < t < \infty$ be defined by

$$q_t(\eta, \omega) = \exp\left( \sum_{j=1}^{m} \int_0^t \eta_s^j h_s^j(X_s(\omega))\, ds - \tfrac{1}{2} \int_0^t \left[ h_s^j(X_s(\omega)) \right]^2 ds \right). \tag{5.14}$$

Then, recall that [see (5.6), (5.7)]

$$\sigma_t(f, \eta) = \int_\Omega f(X_t(\omega)) q_t(\eta, \omega)\, d\Pi(\omega). \tag{5.15}$$

Since $(X_s)$ is progressively measurable, $q_t(\eta, \cdot)$ is $\mathscr{A}_t$-measurable and hence we have for an integrable $g$,

$$\tilde{\sigma}_t(g, \eta) = \tilde{\sigma}_t(E(g/\mathscr{A}_t), \eta). \tag{5.16}$$

Fix $f \in \mathscr{D}_0$ and $0 \le t \le T$. Let $g_t\colon \Omega \to R$ be defined by

$$g_t(\omega) = f(X_T(\omega)) - \int_t^T (\mathscr{L}_s f)(X_s(\omega))\, ds. \tag{5.17}$$

Then (5.11) implies that $E(g_t | \mathscr{A}_t) = f(X_t)$ and hence in view of (5.16), we have

$$\sigma_t(f, \eta) = \tilde{\sigma}_t(g_t, \eta). \tag{5.18}$$

Now, clearly, for all $\omega \in \Omega$, $\eta \in H$, $g_t, q_t$ are absolutely continuous functions of $t$ and for a.e. $t$

$$\frac{d}{dt} g_t(\omega) = (\mathscr{L}_t f)(X_t(\omega)), \tag{5.19}$$

$$\frac{d}{dt} q_t(\eta, \omega) = \left[ \sum_{j=1}^{m} h_t^j(X_t(\omega)) \eta_t^j - \tfrac{1}{2} \left| h_t(X_t(\omega)) \right|^2 \right] q_t(\eta, \omega). \tag{5.20}$$

Hence, $g_t q_t$ is also an absolutely continuous function of $t$ and for all $\omega \in \Omega$, $\eta \in H$, and $0 \le t \le T$, we have

$$\begin{aligned} g_t(\omega) q_t(\eta, \omega) = g_0(\omega) &+ \int_0^t (\mathscr{L}_s f)(X_s(\omega)) q_s(\eta, \omega)\, ds \\ &+ \sum_{j=1}^{m} \int_0^t g_s(\omega) h_s^j(X_s(\omega)) \eta_s^j q_s(\eta, \omega)\, ds \\ &- \tfrac{1}{2} \int_0^t g_s(\omega) \left| h_s(X_s(\omega)) \right|^2 q_s(\eta, \omega)\, ds. \end{aligned} \tag{5.21}$$

Recall that by the definition of $\mathscr{L}$ and $\mathscr{D}$, for $f \in \mathscr{D}_0$, $f, \mathscr{L}_t f$ are bounded (uniformly in $t$) and hence $g_t$ is also bounded (uniformly in $t$ for $0 \le t \le T$). Integrating both sides in (5.21) wrt $\Pi$ and using Fubini's theorem [which is justified as $\mathscr{L}_s f, g_s(\omega)$ are uniformly bounded and $h$ satisfies (5.12)] we get for $\eta \in H$ and $0 \le t \le T$,

$$\begin{aligned} \tilde{\sigma}_t(g_t, \eta) = \int_\Omega g_0(\omega)\, d\Pi(\omega) &+ \int_0^t \sigma_s(\mathscr{L}_s f, \eta)\, ds \\ &+ \sum_{j=1}^{m} \int_0^t \tilde{\sigma}_s(g_s h_s^j(X_s) \eta_s^j, \eta)\, ds - \tfrac{1}{2} \int_0^t \tilde{\sigma}_s\left( g_s |h_s(X_s)|^2, \eta \right) ds. \end{aligned} \tag{5.22}$$

Also, $E(g_t|\mathscr{A}_t) = f(X_t)$ and (5.16) also imply

(5.23) $$\tilde{\sigma}_s\big(g_s h_s^j(X_s)\eta_s^j, \eta\big) = \sigma_s\big(fh_s^j\eta_s^j, \eta\big)$$

and

(5.24) $$\tilde{\sigma}_s\big(g_s|h_s(X_s)|^2, \eta\big) = \sigma_s\big(f|h_s|^2, \eta\big).$$

From (5.18), (5.22), (5.23), and (5.24) it follows that

(5.25) $$\begin{aligned}\sigma_t(f, \eta) = Eg_0 + \int_0^t \sigma_s(\mathscr{L}_s f, \eta)\,ds + \sum_{j=1}^m \int_0^t \sigma_s\big(fh_s^j\eta_s^j, \eta\big)\,ds \\ - \tfrac{1}{2}\int_0^t \sigma_s\big(f|h_s|^2, \eta\big)\,ds.\end{aligned}$$

Now (5.13) follows by differentiating (5.25) and substituting $y$ for $\eta$. □

In the recent work on robust nonlinear filtering [e.g., Davis, (1979)] it is customary to use the Stratonovich form of the Zakai equation of the Itô theory and then to make the heuristic change from $dY_t$ to $y_t\,dt$. The resulting equation is precisely Equation (5.13). We shall comment on this in greater detail later in Section 10.

For $f$ such that $f(X_t)$ is integrable let us write

(5.26) $$\Pi_t(f, \eta) = \frac{\sigma_t(f, \eta)}{\sigma_t(l, \eta)}$$

so that $\Pi_t(f, y)$ is the conditional expectation of $f(X_t)$ given $Q_t y$. Equation (5.13) yields the following equation for $\Pi_t(f, y)$ which is the analogue of the Fujisaki–Kallianpur–Kunita (FKK) equation in our white noise setup.

**THEOREM 5.2.** *Let $h$ satisfy* (5.12). *Then for all $y \in H$,*

(5.27) $$\begin{aligned}\frac{d}{dt}\Pi_t(f, y) = \Pi_t(\mathscr{L}_t f, y) + \sum_{j=1}^m \Pi_t\big(fh_t^j y_t^j, y\big) - \tfrac{1}{2}\Pi_t\big(f|h_t|^2, y\big) \\ - \Pi_t(f, y)\left[\sum_{j=1}^m \Pi_t\big(h_t^j y_t^j, y\big) - \tfrac{1}{2}\Pi_t\big(|h_t|^2, y\big)\right] \quad \text{for a.e. } t\end{aligned}$$

**PROOF.** We have, using (5.26)

(5.28) $$\frac{d}{dt}\Pi_t(f, \eta) = \frac{[(d/dt)\sigma_t(f, \eta)]\sigma_t(l, \eta) - \sigma_t(f, \eta)(d/dt)\sigma_t(l, \eta)}{[\sigma_t(l, \eta)]^2}.$$

Now, (5.27) follows from (5.13), (5.26), and the fact that $\mathscr{L}_t 1 = 0$. □

For $0 \le t < \infty$ and $\eta \in H$, let $\Gamma_t(\eta, \cdot)$ and $F_t(\eta, \cdot)$ be measures on $\mathbf{R}^d$ defined by

(5.29) $$\Gamma_t(\eta, A) = \int_\Omega 1_A(X_t(\omega))\, q_t(\eta, \omega)\, d\Pi(\omega)$$

and

$$F_t(\eta, A) = F_t(\eta, A)[\Gamma_t(\eta, \mathbf{R}^m)]^{-1}, \qquad A \in \mathscr{B}(\mathbf{R}^d). \tag{5.30}$$

Then it follows by standard arguments that for $f$ such that $E|f(X_t)| < \infty$,

$$\sigma_t(f, \eta) = \int_{\mathbf{R}^m} f(x)\Gamma_t(\eta, dx) \tag{5.31}$$

and

$$E_\alpha(f(X_t)|Q_t y) = \Pi_t(f, y) = \int_{\mathbf{R}^d} f(x)F_t(y, dx). \tag{5.32}$$

Equations (5.13) and (5.27) are actually equations for the measures $\{\Gamma_t(\eta, \cdot)\}$ and $\{F_t(\eta, \cdot)\}$. Under stronger conditions on $h$, $\{\Gamma_t\}$, $\{F_t\}$ can be characterized as the unique solutions to (a slight modification of) equations (5.13) and (5.27). We will consider this problem in a more general (infinite-dimensional) context in Section 6. We specialize now to the case when $X$ is a diffusion process (see the Appendix for the definition) so that the paths $X_t$ are continuous. Moreover, $C_0^{1,2}([0, \infty) \times \mathbf{R}^d) \subseteq \mathscr{D}$ and for $f \in C_0^{1,2}([0, \infty) \times \mathbf{R}^d)$,

$$(\mathscr{L}f)(t, x) = \left(\frac{\partial}{\partial t} f\right)(t, x) + (\mathscr{L}_t f)(t, x), \tag{5.33}$$

where

$$(\mathscr{L}_t f)(t, x) = \tfrac{1}{2} \sum_{i, j=1}^{d} a_{ij}(t, x)\left(\frac{\partial^2 f}{\partial x^i \partial x^j}\right)(t, x) + \sum_{i=1}^{d} b_i(t, x)\left(\frac{\partial f}{\partial x^i}\right)(t, x). \tag{5.34}$$

The coefficients $a_{ij}$, $b_i$ are measurable functions from $[0, \infty) \times \mathbf{R}^d$ into $\mathbf{R}$, and the matrix $(a_{ij}(t, x))$ is symmetric and nonnegative definite for each $(t, x) \in [0, \infty) \times \mathbf{R}^d$.

The following result shows that if the process $(X_t)$ admits a density wrt the Lebesgue measure $\lambda$ on $\mathbf{R}^d$, then the measures $\Gamma_t(\eta, \cdot)$ also admit a density $p_t(x, \eta)$ wrt $\lambda$ (called the unnormalized conditional density) which is a (weak) solution of a second order parabolic differential equation.

THEOREM 5.3. *Suppose that for each $t$, the distribution of $X_t$ is absolutely continuous wrt $\lambda$. Then for all $\eta \in H$, the measure $\Gamma_t(\eta, \cdot)$ admits a density $p_t(x, \eta)$ wrt $\lambda$. Further, for each $y \in H$, $p_t(x, y)$ is a weak solution of the PDE:*

$$\frac{\partial}{\partial t} p_t(x, y) = \mathscr{L}_t^* p_t(x, y) + \left(\sum_{j=1}^{m} h_t^j(x) y_t^j - \tfrac{1}{2}|h_t(x)|^2\right) p_t(x, y). \tag{5.35}$$

REMARK 5.1. By a weak solution to (5.35) we mean that for all $f \in C_0^\infty(\mathbf{R}^d)$

$$\frac{\partial}{\partial t}\int_{\mathbf{R}^d} f(x)p_t(x,y)\,d\lambda = \int_{\mathbf{R}^d}[\mathscr{L}_t^* p_t(x,y)]\,f(x)\,dx$$
$$\text{(5.36)} \qquad + \int_{\mathbf{R}^d}\left(\sum_{j=1}^m h_t^j(x)y_t^j - \tfrac{1}{2}|h_t(x)|^2\right)p_t(x,y)f(x)\,dx.$$

PROOF. If $A \in B(\mathbf{R}^d)$ is such that $\lambda(A) = 0$, then by our hypothesis, $EI_A(X_t) = 0$ and hence clearly [see (5.29)], $\Gamma_t(\eta, A) = 0$ for all $\eta \in H$. Thus, $\Gamma_t(\eta,\cdot)$ is absolutely continuous wrt $\lambda$ and hence admits a $(t,x)$-measurable version of a density, say $p_t(x,\eta)$ wrt $\lambda$. Now (5.36) is the same as the equation (5.13). This completes the proof. □

We now turn to the following questions:

(1) Under what conditions on $a, b, h$ does the PDE (5.35) admit a unique classical solution (unique within a class to be specified)?

(2) Under what conditions is the unique solution obtained in (1) the unnormalized conditional density?

It should be noted that the second question is not trivial, for *a priori*, we can only say that the unnormalized conditional density is a weak solution of (5.35).

We present two results which show that the PDE (5.35) has a unique solution in a certain class, which is indeed the unnormalized conditional density. In both results, no assumption of boundedness of $h$ is made. The first of these assumes merely that $h$ is locally Hölder continuous and proves uniqueness for $y$ in a dense subset of $H$. The following lemma shows that given $p_t(x,y)$ for $y$ in a dense set in $H$, $\Gamma_t(x,y)$ can, in principle, be computed for all $y$ in $H$.

LEMMA 5.4. *If $\eta_j \to \eta$ in $H$, then $\Gamma_t(\eta_j,\cdot)$ converges to $\Gamma_t(\eta,\cdot)$ in the total variation norm.*

PROOF. It is easy to see that for each $\omega \in \Omega$, $q_t(\eta_j,\omega) \to q_t(\eta,\omega)$ and further, $q_t(\eta_j,\omega)$ is uniformly bounded by $\sup_j \exp(\frac{1}{2}\|\eta_j\|^2)$ finite as $\eta_j \to \eta$. This implies the required result. □

THEOREM 5.5. *Suppose that $h$: $[0,T]\times\mathbf{R}^d \to \mathbf{R}^m$ is a locally Hölder continuous function. Suppose that the diffusion and drift coefficients $a$ and $b$ of the signal process $(X_t)$ satisfy*

$$\text{(5.37)} \qquad \sum_{i,j=1}^d a_{ij}(t,x)z_iz_j \geq K_1\sum_{i=1}^d z_i^2, \qquad K_1 < \infty,$$

*for all $(t,x)$, for all $(z_1,\ldots,z_d) \in \mathbf{R}^d$;*

$$\text{(5.38)} \qquad a_{ij}, \quad \frac{\partial}{\partial x^i}a_{ij}, \quad \frac{\partial^2}{\partial x^i\,\partial x^j}a_{ij}, \quad b_i, \quad \frac{\partial}{\partial x^i}b_i$$

*are locally Hölder continuous functions satisfying the growth condition*

$$|g(t,x)| \leq K_2(1+|x|^2)^{1/2} \tag{5.39}$$

*for some $K_2 < \infty$. Suppose that the distribution of $X_0$ has a continuous density $\phi$ satisfying for some $\varepsilon > 0$, $K_3 < \infty$,*

$$|\phi(x)| \leq \exp\left(K_3(1+|x|^2)\right)^{\frac{1}{2}-\varepsilon}. \tag{5.40}$$

*Let $H_0 = \{y \in H: y_t \text{ is Hölder continuous}\}$. Then for all $y \in H_0$, the PDE*

$$\frac{\partial p_t(x,y)}{\partial t} = \mathscr{L}_t^* p_t(x,y) + \left(\sum_{j=1}^m h_t^j(x) y_t^j - \tfrac{1}{2}|h_t(x)|^2\right) p_t(x,y), \tag{5.41}$$

*with the initial condition*

$$p_0(x,y) = \phi(x),$$

*has a unique solution in the class $\mathscr{G}$, where $\mathscr{G}$ is the class of $C^{1,2}([0,T] \times \mathbf{R}^d)$ functions $g$ satisfying the growth condition*

$$|g(t,x)| \leq \exp\left(K_4(1+|x|^2)^{1/2}\right) \tag{5.42}$$

*for some $K_4 < \infty$.*

*Furthermore, the unique solution $p_t(x,y)$ (for $y \in H_0$) is the unnormalized conditional density of $X_t$ given $Q_t y$, i.e.,*

$$\Gamma_t(y,B) = \int_{R^d} 1_B(x) p_t(x,y)\, d\lambda(x) \tag{5.43}$$

*for all $y \in H_0$ and for all $B \in \mathscr{B}(\mathbf{R}^d)$, or equivalently,*

$$\sigma_t(f,y) = \int_{\mathbf{R}^d} f(x) p_t(x,y)\, d\lambda(x) \tag{5.43'}$$

*for every bounded, Borel function $f$ from $\mathbf{R}^d$ to $\mathbf{R}$.*

**Proof.** In Equation (5.41), $\mathscr{L}_t^*$ is the formal adjoint of $\mathscr{L}_t$, so that,

$$\begin{aligned}(\mathscr{L}_t^* g)(x) &= \tfrac{1}{2}\sum_{i,j=1}^d a_{ij}(t,x)\left(\frac{\partial^2 g}{\partial x^i \partial x^j}\right)(x) \\ &\quad + \tfrac{1}{2}\sum_{i=1}^d b_i^*(t,x)\left(\frac{\partial g}{\partial x^i}\right)(x) + c^*(t,x) g(t,x),\end{aligned} \tag{5.44}$$

where

$$b_i^*(t,x) = -b_i(t,x) + \sum_{j=1}^d \frac{\partial a_{ij}}{\partial x^j}(t,x) \tag{5.45}$$

and

$$c^*(t,x) = -\sum_{i=1}^d \frac{\partial b_i(t,x)}{\partial x^i} + \tfrac{1}{2}\sum_{i,j=1}^d \frac{\partial^2 a_{ij}(t,x)}{\partial x^i \partial x^j}. \tag{5.46}$$

It is easily seen that $a, b^*$ satisfy (5.38) and $c^*$ is a locally Hölder continuous function satisfying the growth condition (5.39).

Let $c(t, x, y) = \sum_{j=1}^{m} h_t^j(x) y_t^j - \frac{1}{2}|h_t(x)|^2$. Then, for $y \in H_0$ $c(t, x, y)$ is a locally Hölder continuous function and

(5.47) $$c(t, x, y) \le \tfrac{1}{2}|y_t|^2 \le K$$

for some $K < \infty$ as $y_t$ is continuous. For $y \in H_0$ fixed, let a differential operator $\bar{\mathscr{L}}$ be defined for $g \in C^{1,2}([0, T] \times \mathbf{R}^d)$ by

(5.48) $$(\bar{\mathscr{L}}g)(t, x) = (\mathscr{L}_t^* g)(t, x) + c(t, x, y) g(t, x) - \frac{\partial}{\partial t} g(t, x).$$

Taking $H(t, x) = \exp(K_5(1 + |x|^2)^{1/2} e^{\beta t})$, it can be checked that $\bar{\mathscr{L}}(H) \le 0$, $\bar{\mathscr{L}}^*(H^{-1}) \le 0$ for suitable $K_5, \beta$. [See Bodanko (1966) and Kallianpur-Karandikar (1984a).] Thus, by Theorems 1 and 3 in Besala (1979), $\bar{\mathscr{L}}$ admits a fundamental solution $G(t, x, s, z)$ and

(5.50) $$p_t(x, y) = \int_{\mathbf{R}^d} \phi(z) G(t, x, 0, z)\, d\lambda(z)$$

is a solution to $\bar{\mathscr{L}}u = 0$ with the initial condition $p_0(x, y) = \phi(x)$. That $p_t$ defined by (5.50) is the unique solution to (5.41) follows from the results of Bodanko (1966).

The formal adjoint $\bar{\mathscr{L}}^*$ of $\bar{\mathscr{L}}$ is easily seen to be

$$\begin{aligned}(\bar{\mathscr{L}}^* g)(t, x) &= (\mathscr{L}_t g)(t, x) + c(t, x, y) g(t, x) + \frac{\partial}{\partial t} g(t, x)\\ &= \mathscr{L} g(t, x) + c(t, x, y) g(t, x).\end{aligned}$$

Fix $f \in C_0^\infty(\mathbf{R}^d)$ and $0 < t_0 \le T$ and define $v(s, z), 0 \le s \le t_0, z \in \mathbf{R}^d$ by

(5.51) $$v(s, z) = \int_{R^d} f(x) G(t_0, x, s, z)\, d\lambda(z).$$

Then, by Theorem 2 in Besala (1979), $v$ satisfies $\bar{\mathscr{L}}^* v = 0$, i.e.,

(5.52) $$\mathscr{L} v(t, x) + c(t, x, y) v(t, x) = 0, \qquad 0 \le t < t_0,\ x \in \mathbf{R}^d$$

with the boundary condition

$$v(t_0, x) = f(x).$$

By the estimates on $G$ in Besala (1979), it follows that $v$ is bounded and hence by the Feynman–Kac formula, Theorem A.3 (the conditions of Theorem A.3 are satisfied as $v$ is bounded and $c$ is bounded above)

(5.53) $$v(0, X_0) = E_\Pi\left( f(X_{t_0}) \exp\left( \int_0^{t_0} c(t, X_t, y)\, dt \right) \Big| \sigma(X_0) \right).$$

Thus

(5.54) $$\begin{aligned}\int \phi(z) v(0, z)\, d\lambda(z) &= E_\Pi v(0, X_0)\\ &= E_\Pi\left( f(X_{t_0}) \exp\left( \int_0^{t_0} c(t, X_t, y)\, dt \right) \right)\\ &= \sigma_{t_0}(f, y).\end{aligned}$$

Using (5.40) and the estimates on $G$ given in Besala (1979) it follows that

$$\int_{\mathbf{R}^d}\int_{\mathbf{R}^d}|\phi(z)||f(x)|G(t_0,x,0,z)\,d\lambda(x)\,d\lambda(z) < \infty$$

and hence by Fubini's theorem, (5.50), (5.51), and (5.53), we have

$$\begin{aligned}\int_{\mathbf{R}^d} f(x)p_{t_0}(x,y)\,d\lambda(x) &= \int_{\mathbf{R}^d} f(x)\int_{\mathbf{R}^d}\phi(z)G(t_0,x,0,z)\,d\lambda(z)\,d\lambda(x)\\ &= \int_{\mathbf{R}^d}\phi(z)\left[\int_{\mathbf{R}^d} f(x)G(t_0,x,0,z)\,d\lambda(x)\right]d\lambda(z)\\ &= \int_{\mathbf{R}^d}\phi(z)v(0,z)\,d\lambda(z)\\ &= \sigma_{t_0}(f,y).\end{aligned} \tag{5.55}$$

Since (5.55) holds for all $f \in C_0^\infty(\mathbf{R}^d)$, it follows (by the usual arguments) that (5.55) also holds for all bounded measurable functions. Hence, taking $f = 1_H$, we get (5.43). This completes the proof. □

It is clear from Equation (5.41) that $[\partial p_t(x,y)]/\partial t$ cannot be continuous in $t$ for all $y \in H$. Hence there can be no classical solution of (5.41) for every $y \in H$. However, in our next result we show the existence of a unique solution of (5.41) (for every $y \in H$) in a slightly weaker sense.

For reasons of convenience, we shall postpone the proof of parts (i) and (ii) of the following theorem to Section 8. Here, we only prove (iii).

THEOREM 5.6. *Suppose that the diffusion and drift coefficients $a, b$ of the signal process $(X_t)$ satisfy (5.37) and (5.38). Further suppose that the density $\phi$ of $X_0$ satisfies (5.40). For some suitable $\alpha > 0$ let*

$$E\exp\left(\alpha|X_0|^2\right) < \infty,$$

*and for all $i, j, k$, suppose*

$$a_{ij} \quad \text{is bounded;} \tag{5.56}$$

$$h^k, \quad \frac{\partial h^k}{\partial x^i}, \quad \frac{\partial^2 h^k}{\partial x^i\,\partial x^j}, \quad \frac{\partial h^k}{\partial t} \tag{5.57}$$

*are locally Hölder continuous in $(t,x)$ and*

(5.58)

$$h^k, \quad \frac{\partial h^k}{\partial t}, \quad \sum_j a_{ij}\frac{\partial h^k}{\partial x^i}, \quad \sum_{i,j} a_{ij}\left(\frac{\partial^2 h^k}{\partial x^i\partial x^j} + \frac{\partial h^k}{\partial x^i}\cdot\frac{\partial h^k}{\partial x^j}\right), \quad \sum_i b_i\frac{\partial h^k}{\partial x^i}$$

*satisfy the growth condition (5.39). Then*

(i) *for every $y \in H$, there is a unique $p_t(x,y)$ such that (5.41) is satisfied for a.e. $t$ and*

$$p_t(x,y)\cdot\exp\left[-\sum_{i=1}^m h_t^i(x)\int_0^t y_s^i\,ds\right] \in \mathcal{G}. \tag{5.59}$$

(ii) *The mapping $y \to p_t(x, y)$ is continuous in the following sense: if $y_n \to y$ in $H$, then $p(\cdot, y_n)$ converges to $p_\cdot(\cdot, y)$ uniformly on compact subsets of $[0, T] \times \mathbf{R}^d$.*

(iii) *For all $y \in H$, the unique solution $p_t(x, y)$ of (5.41) is the unnormalized conditional density, i.e., (5.43′) holds for all $y \in H$ and for all bounded measurable functions $f$ from $\mathbf{R}^d$ to $\mathbf{R}$.*

PROOF. We will prove (i) and (ii) later. For (iii), observe that for $y \in H_0$, as proved in Theorem 5.4, for all $B \in \mathscr{B}(\mathbf{R}^d)$,

$$\Gamma_t(y, B) = \int_{\mathbf{R}^d} 1_B(x) p_t(x, y)\, d\lambda(x). \tag{5.60}$$

For $B \in \mathscr{B}(\mathbf{R}^d)$ bounded, by Lemma 5.4 and part (ii) above, both sides of (5.60) are continuous functions of $y$ and hence the denseness of $H_0$ in $H$ implies that (5.60) holds for all $y \in H$. The restriction that $B \in \mathscr{B}(\mathbf{R}^d)$ is bounded can be removed by the usual arguments. □

REMARK 5.2. Note that $h(x) = x^3$ (the cubic sensor) satisfies the conditions of Theorem 5.5.

**6. Equations for the optimal filter: Markov signal with general state space.** Our purpose in this section is to study the finitely additive white noise theory in a more general framework so as to include applications to signal and observation processes taking values in infinite-dimensional Hilbert spaces. The chief difficulty here is the lack of a conditional density since there is no Lebesgue measure (or any natural measure) in Hilbert space. The partial differential equations of Section 5 are now replaced by finitely additive analogues of measure-valued equations of FKK and Zakai and of a type of equations introduced by Kunita (1971) and studied also by Szpirglas (1978). The equivalence and uniqueness of solutions of these equations are established in the four principal results of this section. The proofs, of which only an outline is given, are based on auxiliary results on equations governing measures induced by multiplicative transformations of Markov processes. The full details are to be found in Kallianpur–Karandikar (1984b).

Let $\mathscr{S}$ be a Polish space with its Borel $\sigma$-field $\mathscr{B}(S)$ and let $\{X_t\}_{0 \le t \le T}$, defined on a countably additive probability space $(\Omega, \mathscr{A}, \Pi)$, be an $\mathscr{S}$-valued Markov process wrt a family $(\mathscr{A}_t)$ of sub-$\sigma$ fields of $\mathscr{A}$. We will assume that the paths of $\{X_t\}$ are progressively measurable and that the process $\{X_t\}$ admits a transition probability function. Let $\{V_t^s\}$ be the two parameter semigroup associated with $\{X_t\}$ acting on $\mathscr{I}(S, \mathscr{B}(S))$ and let $\{\hat{T}_t\}$ be the one parameter semigroup associated with $\hat{X}_t = (t, X_t)$ acting on $\mathscr{I}(\hat{S}, \mathscr{B}(\hat{S}))$, $\hat{S} = [0, \infty) \times S$.

Let $\mathscr{L}$ be the extended generator of $\{T_t\}$ and let $\mathscr{D}$ be the domain of $\mathscr{L}$. (See the Appendix.)

Let $\mathscr{K}$ be a real separable Hilbert space with the inner product $(\cdot, \cdot)_{\mathscr{K}}$ and norm $\|\cdot\|_{\mathscr{K}}$. Let $0 < T < \infty$ and let $h$ be a measurable function from $[0, T] \times$

$S \to \mathscr{X}$ such that

$$\int_0^T \|h_s(X(\omega))\|_{\mathscr{X}}^2 \, ds < \infty, \qquad \Pi\text{-a.s. } \omega. \tag{6.1}$$

Let

$$H = L^2([0,T], \mathscr{X}) = \left\{\eta: [0,T] \to \mathscr{X}: \int_0^T \|\eta(s)\|_{\mathscr{X}}^2 \, ds < \infty\right\}. \tag{6.2}$$

$H$ is itself a real separable Hilbert space with the inner product

$$(\eta_1, \eta_2) = \int_0^T (\eta_1(s), \eta_2(s))_{\mathscr{X}} \, ds.$$

Let $e = (e_t)$ be the Gaussian white noise on $(H, \mathscr{C}, m)$, where $m$ is the Gaussian measure on $(H, \mathscr{C})$. Let $\xi$: $(\Omega, \mathscr{A}, \Pi) \to H$ be defined by

$$\begin{aligned} \xi_s(\omega) &= h_s(X_s(\omega)), && \text{if (6.1) holds, } 0 \le s \le T, \\ &= 0, && \text{otherwise.} \end{aligned} \tag{6.3}$$

The measurability of $h$ and the assumption that paths of $(X_t)$ are progressively measurable implies that $\xi$ is Borel-measurable. Let $(E, \mathscr{E}) = (\Omega, \mathscr{A}) \odot (H, \mathscr{C})$ and let $\alpha$ be the product of $\Pi$ and $m$ constructed in Section 4. The abstract statistical model (4.6) with $H, \xi$ given by (6.2), (6.3) on $(E, \mathscr{E}, \alpha)$

$$y = \xi + e \tag{6.4}$$

now becomes

$$y_s = h_s(X_s) + e_s, \qquad 0 \le s \le T. \tag{6.4'}$$

In this section, (6.4') is our model of nonlinear filtering. For $0 \le s < t \le T$, let

$$H_t^s = \left\{\eta \in H: \int_0^s \|\eta(u)\|_{\mathscr{X}}^2 \, du + \int_t^T \|\eta(u)\|_{\mathscr{X}}^2 \, du = 0\right\}. \tag{6.5}$$

It is easily seen that $H_t^s$ is a closed linear subspace of $H$. Let $Q_t^s$ be the orthogonal projection onto $H_t^s$. For the sake of convenience, we will denote $H_t^0$ and $Q_t^0$ by $H_t$ and $Q_t$, respectively.

For $B \in \mathscr{B}(S)$, $0 \le t \le T$, and $\eta \in H$ let $\Gamma_t(\eta)(B)$, $F_t(\eta)(B)$ be defined by

$$\Gamma_t(\eta)(B) = \int_\Omega 1_B(X_t(\omega)) \exp\left[(\eta, Q_t\xi(\omega)) - \tfrac{1}{2}\|Q_t\xi(\omega)\|^2\right] d\Pi(\omega) \tag{6.6}$$

and

$$F_t(\eta)(B) = [\Gamma_t(\eta)(S)]^{-1} \Gamma_t(\eta)(B). \tag{6.7}$$

Since $(\eta, Q_t\xi(\omega)) - \frac{1}{2}\|Q_t\xi(\omega)\|^2 \le \frac{1}{2}\|\eta\|^2$, it is easy to see that for all $\eta \in H$, $\Gamma_t(\eta)$ is a countably additive finite measure on $(S, \mathscr{B}(S))$ and $F_t(\eta)$ is a countably additive probability measure on $(S, \mathscr{B}(S))$. Let $\mathscr{M}(S)$ be the class of countably additive finite measures on $(S, \mathscr{B}(S))$. Then, $\Gamma_t(\eta), F_t(\eta) \in \mathscr{M}(S)$.

The Bayes formula (Theorem 4.9) implies that for all bounded Borel-measurable functions $f$ on $S$,

$$E_\alpha(f(X_t) | Q_t y) = \int_S f(x) F_t(y)(dx). \tag{6.8}$$

Thus, $F_t(y)$ is the conditional distribution of $X_t$ given $Q_t y$ and in view of (6.7), $\Gamma_t(y)$ will be called the unnormalized conditional distribution of $X_t$ given $Q_t y$.

The next result shows that $\Gamma_t(y)$, $F_t(y)$ are $\mathscr{M}(S)$-valued random variables on $(E, \mathscr{E}, \alpha)$. Let $n = \alpha[y]^{-1}$

THEOREM 6.1. *For* $0 \le t \le T$,

$$\Gamma_t, F_t \in \mathscr{L}^*(H, \mathscr{C}, n; \mathscr{M}(S)), \tag{6.9}$$

$$\Gamma_t(y), F_t(y) \in \mathscr{L}^*(E, \mathscr{E}, \alpha; \mathscr{M}(S)), \tag{6.10}$$

*and*

$$\Gamma_t, F_t \in \mathscr{U}(y) = \mathscr{U}(E, \mathscr{E}, \alpha; \mathscr{M}(S), y). \tag{6.11}$$

PROOF. Recall that $\mathscr{M}(S)$ is itself a Polish space with a metric $d_0$ given by

$$d_0(\mu_1, \mu_2) = \sum_{j=1}^{\infty} \frac{1}{2^j}\left\{\left|\int f_j\, d\mu_1 - \int f_j\, d\mu_2\right| \wedge 1\right\},$$

where $f_1 \equiv 1$ and $\{f_j\colon j \ge 1\}$ is a countable subset of $C_b(S)$ dense in the uniform norm topology. [See Stroock–Varadhan (1979).] Consider the map $g_1\colon \mathscr{M}(S) \to \mathbf{R}^\infty$ defined by

$$g_1(\mu) = \{\langle f_j, \mu\rangle\}_{j\ge 1},$$

where $\langle f, \mu\rangle = \int f\, d\mu$. Let $U$ be the range of $g_1$. Then the form of the metric $d_0$ implies that $g_1$ is a homeomorphism onto $U$. Let $g\colon U \to \mathscr{M}(S)$ be the inverse of $g_1$, which is continuous. Now, in view of Theorem 3.5, to prove the assertions for $\{\Gamma_t\}$, it suffices to prove that for all $j \ge 1$,

$$\langle f_j, \Gamma_t\rangle \in \mathscr{L}^*(H, \mathscr{C}, n; \mathbf{R}), \tag{6.12}$$

$$\langle f_j, \Gamma_t(y)\rangle \in \mathscr{L}^*(E, \mathscr{E}, \alpha; \mathbf{R}), \tag{6.13}$$

$$\langle f_j, \Gamma_t\rangle \in \mathscr{U}(y) = \mathscr{U}(E, \mathscr{E}, \alpha; \mathbf{R}, y), \tag{6.14}$$

and

$$\left\{R_n(\langle f_j, \Gamma_t\rangle)\right\}_{j\ge 1} \in U, \qquad \text{a.s.}, \tag{6.15}$$

where $R_n$ is an $n$-lifting. Now, it is easy to see that $\langle f_j, \Gamma_t\rangle \in \mathscr{I}(H)$ and hence (6.12), (6.13), and (6.14) follow from Theorems 3.5, 4.4, and 4.8.

Let $L_1$ be a measurable representative of $n$, $L_1\colon H \to \mathscr{L}(\Omega_1, \mathscr{A}_1, \Pi_1)$ say, and let $R_n$ be the corresponding $n$-lifting. For a fixed $t \in [0, T]$, $\omega_1 \in \Omega$, let $\Lambda_t(\omega_1) \in \mathscr{M}(S)$ be defined by

(6.16)

$$\Lambda_t(\omega_1)(B) = \int_\Omega 1_B(X_t(\omega))\exp\left(L_1(Q_t\xi(\omega))(\omega_1) - \tfrac{1}{2}\|Q_t\xi(\omega)\|^2\right) d\Pi(\omega).$$

Then, using Theorem 4.4, it can be checked that

(6.17) $$R_n(\langle f_j, \Gamma_t\rangle) = \langle f_j, \Lambda_t(\omega_1)\rangle, \qquad j \geq 1,$$

so that (6.15) holds. To prove the assertions for $\{F_t\}$, observe that (6.17) implies that

(6.18) $$R_n(\langle 1, \Gamma_t(\cdot)\rangle) > 0, \qquad \Pi_1\text{-a.s.}$$

Now, the assertions (6.12), (6.13), and (6.14) for $F_t$ follow from the corresponding results for $\Gamma_t$ (proved above), (6.18), and Theorem 3.5.

In Kallianpur–Karandikar (1984b), it has been shown that the measures $\{\Gamma_t\}$ and $\{F_t\}$ satisfy equations analogous to Zakai and FKK equations, and can be characterized as the unique solution to these equations. We will state these results and outline the proof. We assume that the process $\{X_t\}$ satisfies the condition A.13 (see the Appendix) and that $h$ satisfies

(6.19) $$\|h_s(x)\|_K \leq q(s), \qquad \forall\, x \in S$$

for a measurable function $q$ on $[0, T]$ satisfying

(6.20) $$\int_0^T q^2(s)\,ds < \infty.$$

THEOREM 6.2. *For all $y \in H$, $\{\Gamma_t(y)\}$ satisfies*

(6.21) $$\begin{aligned}\langle f(t,\cdot), \Gamma_t(y)\rangle &= \langle f(0,\cdot), \Gamma_0(y)\rangle + \int_0^t \langle (\mathscr{L}f)(s,\cdot), \Gamma_s(y)\rangle\,ds \\ &\quad + \int_0^t \Big\langle \big\{(h_s(\cdot), y_s)_{\mathscr{K}} - \tfrac{1}{2}\|h_s(\cdot)\|_{\mathscr{K}}^2\big\} f(s,\cdot), \Gamma_s(y)\Big\rangle\,ds,\end{aligned}$$

*for all $f \in \mathscr{D}$.*

*Further, $\Gamma_t(y)$ is the unique solution of* (6.21) *in the class of measures $\{K_t\}$ on $(S, \mathscr{B}(S))$ satisfying*

(6.22) *for all $A \in \mathscr{B}(S)$, $K_t(A)$ is a bounded measurable function of $t$ and $K_0(A) = E_{\Pi}1_A(X_0)$.*

THEOREM 6.3. *For all $y \in H$, $F_t(y)$ satisfies*

(6.23) $$\begin{aligned}\langle f(t,\cdot), F_t(y)\rangle &= \langle f(0,\cdot), F_0(y)\rangle + \int_0^t \langle (\mathscr{L}f)(s,\cdot), F_s(y)\rangle\,ds \\ &\quad + \int_0^t \Big\langle \big\{(h_s(\cdot), y_s)_{\mathscr{K}} - \tfrac{1}{2}\|h_s(\cdot)\|_{\mathscr{K}}^2\big\} f(s,\cdot), F_s(y)\Big\rangle\,ds \\ &\quad - \int_0^t \Big\langle \big\{(h_s(\cdot), y_s)_{\mathscr{K}} - \tfrac{1}{2}\|h_s(\cdot)\|_{\mathscr{K}}^2\big\}, F_s(y)\Big\rangle \\ &\quad \cdot \langle f(s,\cdot), F_s(y)\rangle\,ds\end{aligned}$$

*for all $f \in \mathscr{D}$.*

*Further, $F_t(y)$ is the unique solution of* (6.23) *in the class of measures $\{K_t\}$ on $(S, \mathscr{B}(S))$ satisfying* (6.22).

THEOREM 6.4.

(a) *For all* $y \in H$, $\{\Gamma_t(y)\}$ *is the unique solution of*

$$\begin{aligned}\langle f, \Gamma_t(y)\rangle &= \langle V_t^0 f, \Gamma_0(y)\rangle \\ &\quad + \int_0^t \Big\langle \big\{(h_s(\cdot), y_s)_{\mathscr{K}} - \tfrac{1}{2}\|h_s(\cdot)\|_{\mathscr{K}}^2\big\} V_t^s f(\cdot), \Gamma_s(y)\Big\rangle\, ds \end{aligned} \tag{6.24}$$

*for all* $f \in \mathscr{I}(S, \mathscr{B}(S))$, *in the class of* $\{K_t\} \subseteq \mathscr{M}(S)$ *satisfying* (6.22).

(b) *Define* $\Gamma_t^j(y)$ *inductively as follows. For* $B \in \mathscr{B}(S)$, *let*

$$\Gamma_t^0(y)(B) = E_\Pi 1_B(X_t)$$

*and for* $j \geq 0$, *let* $\langle f, \Gamma_t^{j+1}(y)\rangle$ *be the right-hand side of* (6.24), *with* $\Gamma_t^j(y)$ *instead of* $\Gamma_t(y)$.

*Then* $\Gamma_t^j(y)$ *converges to* $\Gamma_t(y)$ *uniformly in* $t$, *in total variation norm on* $\mathscr{M}(S)$.

THEOREM 6.5.

(a) *For all* $y \in H$, $\{F_t(y)\}$ *is the unique solution of*

$$\begin{aligned}\langle f, F_t(y)\rangle &= \langle V_t^0 f, F_0(y)\rangle \\ &\quad + \int_0^t \Big\langle \big\{(h_s(\cdot), y_s)_{\mathscr{K}} - \tfrac{1}{2}\|h_s(\cdot)\|_{\mathscr{K}}^2\big\} V_t^s f, F_s(y)\Big\rangle\, ds \\ &\quad - \int_0^t \Big\langle \big\{(h_s(\cdot), y_s)_{\mathscr{K}} - \tfrac{1}{2}\|h_s(\cdot)\|_{\mathscr{K}}^2\big\}, F_s(y)\Big\rangle \\ &\qquad \cdot \langle V_t^s f, F_s(y)\rangle\, ds \end{aligned} \tag{6.25}$$

*for all* $f \in \mathscr{I}(S, \mathscr{B}(S))$ *in the class of* $\{K_t\} \subseteq \mathscr{M}(S)$ *satisfying* (6.22).

(b) *Define* $F_t^j(y)$ *inductively, by*

$$F_t^0(y)(B) = E1_B(X_t), \qquad B \in \mathscr{B}(S) \tag{6.26}$$

*and for* $j \geq 0$, $\langle f, F_t^{j+1}(y)\rangle$ *by the right-hand side of* (6.25) *with* $F_t^j(y)$ *instead of* $F_t(y)$.

*Then* $F_t^j(y)$ *converges to* $F_t(y)$ *uniformly in* $t$, *in total variation norm on* $\mathscr{M}(S)$.

*Outline of proofs of Theorems 6.2–6.5.* All the assertions in these theorems are proved for each $y \in H$, fixed. Let

$$g_s(t, x) = (h_s(x), y_s)_{\mathscr{K}} - \tfrac{1}{2}\|h_s(x)\|_{\mathscr{K}}^2. \tag{6.27}$$

Then, by the definition of $\Gamma_t(y)$, it follows that for all $B \in \mathscr{B}(S)$,

$$\Gamma_t(y)(B) = \int_\Omega 1_{(\{t\}\times B)}(\hat{X}_t) \exp\Big(\int_0^t g_s(\hat{X}_s)\, ds\Big)\, d\Pi, \tag{6.28}$$

where $\hat{X}_t = (t, X_t)$.

Also the conditions (6.19), (6.20) on $h$ imply that

$$|g_s(t,x)| \le A(s), \tag{6.29}$$

where

$$A(s) = q_s^2 + \|y_s\|_{\mathscr{K}}^2$$

and $A$ satisfies

$$\int_0^T A(s)\,ds < \infty. \tag{6.30}$$

First, using the representation (6.28) and some elementary properties of the semigroup $\{V_t^*\}$, it is shown that $\Gamma_t(y)$ satisfies (6.24) and uniqueness is proved by a Gronwall-type inequality. Part (b) of Theorem 6.4 is proved by obtaining a recursive estimate of the total variation of $\Gamma_t(y) - \Gamma_t'(y)$. Theorem 6.2 is deduced from Theorem 6.4 by showing that the equations (6.21) and (6.24) are equivalent, i.e., $\{K_t\}$ satisfying (6.22) is a solution of (6.21) if and only if it satisfies (6.24). The fact that $F_t(y)$ satisfies (6.23) follows from (6.7) and (6.21). Then it is shown that the equations (6.23) and (6.25) are equivalent in the sense described above, so that $F_t(y)$ satisfies (6.25). The uniqueness assertion in Theorem 6.5 is proved by a Gronwall-type inequality. This and the equivalence of equations (6.23) and (6.25) imply the uniqueness part of Theorem 6.3. The proof of Theorem 6.5(b) is similar to that of Theorem 6.4(b). □

In the stochastic calculus theory of nonlinear filtering, the optimal filter has been shown to be a measure-valued Markov process [see, e.g., Kunita (1971)]. The same question arises in our theory. However, the optimal nonlinear filter $F_t(y)$ is a measure-valued process defined on a finitely additive probability space and no definition of the Markov property in this context is available to us in the existing literature. With a suitable definition of this property we have been able to prove that $F_t(y)$ and $\Gamma_t(y)$ are Markov processes on $(E, \mathscr{E}, \alpha)$. The Markov property also holds if these processes are regarded as defined on $(H, \mathscr{C}, n)$. The proofs will appear in a forthcoming monograph. Also see Kallianpur and Karandikar (1984c).

**7. Likelihood ratios.** The theory in Sections 3 and 4 can be used to derive likelihood ratios in the finitely additive framework. Their applications to detection problems and to parameter estimation will be postponed for future consideration. Earlier work related to the material of this section will be found in Balakrishnan (1982).

In order to allow infinite-dimensional processes as well as random fields as signals in statistical problems, in the abstract statistical model (4.6) we shall take

$$H = L^2([0,T], \mathscr{K}), \tag{7.1}$$

where $\mathscr{K}$ is a separable Hilbert space with inner product $(\ ,\ )_{\mathscr{K}}$ and norm $\|\cdot\|_{\mathscr{K}}$. $H$ is a Hilbert space with inner product

$$(\eta, \eta') = \int_0^T (\eta_s, \eta_s')_{\mathscr{K}}\,ds. \tag{7.2}$$

Let $(X_t)$ be an $S$-valued process on a countably additive probability space $(\Omega, \mathscr{A}, \Pi)$, where $S$ is a Polish space. Assume that the map $(\omega, t) \to X_t(\omega)$ is $\mathscr{A} \times \mathscr{B}_{[0,T]}$-measurable. Let $h$: $[0,T] \times S \to \mathscr{X}$ be a measurable function such that

$$\int_0^T \|h_s(X_s)\|_{\mathscr{X}}^2 \, ds < \infty, \qquad \Pi\text{-a.s.} \tag{7.3}$$

Take

$$\xi(\omega) = h_s(X_s(\omega)), \qquad 0 \le s \le T. \tag{7.4}$$

Then $\xi = (\xi_s)$ is an $H$-valued random variable on $(\Omega, \mathscr{A}, \Pi)$. Let $e = (e_s)$ be the Gaussian white noise on $(H, \mathscr{C}, m)$, where $m$ is the canonical Gaussian measure on $(H, \mathscr{C})$. Let $(E, \mathscr{E}, \alpha) = (\Omega, \mathscr{A}, \Pi) \odot (H, \mathscr{C}, m)$. Now the abstract statistical model (4.6)

$$y = \xi + e$$

considered in Section 4 takes the form

$$y_t = h_t(X_t) + e_t, \qquad 0 \le t \le T. \tag{7.5}$$

By Theorem 4.6, $y$ is a quasicylindrical mapping from $(E, \mathscr{E}, \alpha)$ into $(H, \mathscr{C})$, the induced measure $n = \alpha[y]^{-1}$ is absolutely continuous wrt $m$ and for $\eta \in H$,

$$\begin{aligned} \frac{dn}{dm}(\eta) &= \int_\Omega \exp\left((\eta, \xi(\omega)) - \tfrac{1}{2}\|\xi(\omega)\|^2\right) d\Pi(\omega) \\ &= \int_\Omega \exp\left(\int_0^T (\eta_s, \xi_s(\omega))_{\mathscr{X}} \, ds - \tfrac{1}{2}\int_0^T \|\xi_s(\omega)\|_{\mathscr{X}}^2 \, ds\right) d\Pi(\omega). \end{aligned} \tag{7.6}$$

For $0 \le t \le T$, let $H_t$ be the closed linear subspace of $H$ given by

$$H_t = \left\{\eta \in H \colon \int_t^T \|\eta_s\|_{\mathscr{X}}^2 \, ds = 0\right\}$$

and let $Q_t$ be the orthogonal projection onto $H_t$. Again by Theorem 4.6, $Q_t y$ is a quasicylindrical mapping from $(E, \mathscr{E}, \alpha)$ into $(H_t, \mathscr{C}(H_t) = \mathscr{C}_t)$ and if $n_t = \alpha[Q_t y]^{-1}$, then $n_t \ll m_t$ and for $\eta \in H_t$,

$$\frac{dn_t}{dm_t}(\eta) = \int_\Omega \exp\left(\int_0^t (\eta_s, \xi_s(\omega))_{\mathscr{X}}^2 \, ds - \tfrac{1}{2}\int_0^t \|\xi_s(\omega)\|_{\mathscr{X}}^2 \, ds\right) d\Pi(\omega), \tag{7.7}$$

where $m_t$ is the canonical Gaussian measure on $H_t$.

For $0 \le t \le T$, let $\rho_t$: $H \to \mathbf{R}$ be defined by

$$\rho_t(\eta) = \frac{dn_t}{dm_t}(Q_t \eta), \qquad \eta \in H. \tag{7.8}$$

It is easy to check that

$$\rho_t \in \mathscr{I}(H), \qquad 0 \le t \le T. \tag{7.9}$$

PROPOSITION 7.1. *$(\rho_t)_{0 \le t \le T}$ is a "martingale" on $(H, \mathscr{C}, m)$, i.e.,*

$$\rho_t = E_m(\rho_T | Q_t), \qquad 0 \le t \le T. \tag{7.10}$$

PROOF. First, (7.9), Theorem 4.4, and Lemma 4.7 imply that

$$\frac{dn_t}{dm_t} \in \mathscr{Y}(Q_t). \tag{7.11}$$

Now, if $C \in \mathscr{C}_t$ and $C_1 = Q_t^{-1}C$, then

$$\begin{aligned}
\int_H 1_{C_1}(\eta)\rho_t(\eta)\,dm(\eta) &= \int_{H_t} 1_C(\eta)\frac{dn_t}{dm_t}(\eta)\,dm_t(\eta) \qquad \text{[by (7.11)]}\\
&= n_t(C)\\
&= \alpha(Q_t y \in C)\\
&= \alpha(y \in C_1)\\
&= n(C_1)\\
&= \int_H 1_{C_1}(\eta)\rho_T(\eta)\,dm(\eta).
\end{aligned} \tag{7.12}$$

The relations (7.11) and (7.12) imply (7.10). □

In analogy with the usual terminology, we define $\rho_t(y)$ to be the likelihood ratio for the model (7.5) for observations over the interval $[0, t]$.

By the definition of $\rho_t$, (7.8), and the expression (7.7), it follows that

$$\rho_t(y) = \int_\Omega q_t(\omega, y)\,d\Pi(\omega), \tag{7.13}$$

where

$$q_t(\omega, y) = \exp\left(\int_0^t (h_s(X_s(\omega)), y_s)_{\mathscr{K}}\,ds - \tfrac{1}{2}\int_0^t \|h_s(X_s(\omega))\|_{\mathscr{K}}^2\,ds\right). \tag{7.14}$$

We will now obtain an alternative expression for the likelihood $\rho_t(y)$.

Let us recall that by the Bayes formula (Theorem 4.9) for a function $f\colon S \to \mathbf{R}$ with $E_\Pi |f(X_t)| < \infty$,

$$\begin{aligned}
\Pi_s(f) &= E_s(f(X_t)|Q_t y)\\
&= \frac{1}{\rho_t(y)} \cdot \int_\Omega f(X_t(\omega))q_t(\omega, y)\,d\Pi(\omega).
\end{aligned} \tag{7.15}$$

For a measurable function $g\colon S \to \mathscr{K}$ s.t. $E_\Pi \|g(X_t(\omega))\|_{\mathscr{K}}^2 < \infty$, we define

$$\Pi_s(g) = \frac{1}{\rho_t(y)} \cdot \int_\Omega g(X_t(\omega))q_t(\omega, y)\,d\Pi(\omega), \tag{7.16}$$

where the integral in (7.16) is a Bochner–Pettis integral. From the standard properties of the latter it follows that for any $k \in \mathscr{K}$,

$$(\Pi_s(g), k)_{\mathscr{K}} = \Pi_s((g, k)_{\mathscr{K}}),$$

so that for any $k \in \mathscr{K}$, we have

$$(\Pi_s(g), k)_{\mathscr{K}} = \mathbf{E}_\alpha((g(X_t), k)_{\mathscr{K}} | Q_t y). \tag{7.17}$$

Hence we define $\Pi_s(g)$ given by (7.16) as the conditional expectation of $g(X_t)$ given $Q_t y$ on $(E, \mathscr{E}, \alpha)$.

Formally differentiating, we get

$$\begin{aligned}
\frac{d}{dt}\log \rho_t(y) &= \frac{1}{\rho_t(y)} \frac{d}{dt}\rho_t(y) \\
&= \frac{1}{\rho_t(y)} \frac{d}{dt}\int_\Omega q_t(\omega, y)\, d\Pi(\omega) \\
&= \frac{1}{\rho_t(y)} \int_\Omega \frac{d}{dt} q_t(\omega, y)\, d\Pi(\omega) \\
&= \frac{1}{\rho_t(y)} \int_\Omega \left[ (h_t(X_t(\omega), y_t)_{\mathscr{K}}) - \tfrac{1}{2}\| h_t(X_t(\omega))\|^2_{\mathscr{K}} \right] q_t(\omega, y)\, d\Pi(\omega) \\
&= (\Pi_t(h_t), y_t)_{\mathscr{K}} - \tfrac{1}{2}\Pi_t(\|h_t\|^2_{\mathscr{K}}).
\end{aligned} \tag{7.18}$$

and hence as $\rho_0(y) = 1$, we get

$$\rho_t(y) = \exp \int_0^t \left[ (\Pi_s(h_s), y_s)_{\mathscr{K}} - \tfrac{1}{2}\Pi_t(\|h_s\|^2_{\mathscr{K}}) \right] ds. \tag{7.19}$$

We will now prove that (7.19) is true.

THEOREM 7.2. *Assume that*

$$E_\Pi \int_0^T \| h_t(X_t)\|^2_{\mathscr{K}}\, dt < \infty. \tag{7.20}$$

*Then*

$$\rho_t(y) = \exp\left[ \int_0^t (\Pi_s(h_s), y_s)_{\mathscr{K}}\, ds - \tfrac{1}{2}\int_0^t \Pi_s(\|h_s\|^2_k)\, ds \right]. \tag{7.19$'$}$$

PROOF. We need to justify the interchange of the differentiation wrt $t$ and the integral wrt $\omega$ in (7.18). For this, first observe that clearly, $q_t(\omega, y)$ is absolutely continuous for all $\omega$, $y$ and hence

$$\begin{aligned}
q_t(\omega, y) &= q_0(\omega, y) + \int_0^t \left[ \frac{d}{ds} q_s(\omega, y) \right] ds \\
&= 1 + \int_0^t \left[ (h_s(X_s(\omega), y_s)_{\mathscr{K}}) - \tfrac{1}{2}\| h_s(X_s(\omega))\|^2_{\mathscr{K}} \right] \\
&\qquad \cdot q_s(\omega, y)\, ds.
\end{aligned} \tag{7.21}$$

Integrating both sides of (7.21) wrt $d\Pi(\omega)$ and using Fubini's theorem for interchanging the integrals on the right-hand side, we get

$$\rho_t(y) = \int_\Omega q_t(\omega, y)\, d\Pi(\omega)$$

(7.22)
$$= 1 + \int_0^t \int_\Omega \left[ (h_s(X_s(\omega), y_s)_{\mathscr{X}} - \tfrac{1}{2}\|h_s(\omega)\|^2_{\mathscr{X}} \right] \cdot q_s(\omega, y)\, d\Pi(\omega)\, ds.$$

Now, (7.22) implies that $\rho_t(y)$ is absolutely continuous and that (7.19) holds for a.e. $t$ in $[0, T]$. □

*Multiparameter white noise.* In the previous discussion, if we take

(7.23)
$$\mathscr{X} = L^2(D, \mathscr{B}_D, \lambda, \mathbf{R}),$$

where $D$ is a Borel subset of a Euclidean space, $\mathscr{B}_D$ is the Borel $\sigma$-field on $D$, and $\lambda$ is the Lebesgue measure, then the white noise on $H = L^2([0, T], \mathscr{X})$, namely,

$$e = (e_t(x) : 0 \le t \le T, x \in D)$$

can be regarded as white noise indexed by $(t, x) \in [0, T] \times D$ as for $0 \le s_1 \le s_2 \le T$, $A \in \mathscr{B}_D$ we have

(7.24)
$$\int_{s_1}^{s_2} \left[ \int_A e_t(x)\, d\lambda(x) \right] dt = \int_{s_1}^{s_2} (e_t, 1_A)_{\mathscr{X}}\, dt = \left( e, 1_{(s_1, s_2)}(\cdot) \cdot 1_A(\cdot) \right).$$

Hence the distribution of the expression on the left-hand side in (7.24) is normal with mean zero and variance $\lambda(A) \cdot (s_2 - s_1)$. This fact can also be seen by computing its characteristic functional. With the above choice of $\mathscr{X}$, the model (7.5) can be written as

(7.25)
$$y_t(x) = h_t(X_t, x) + e_t(x), \qquad (t, x) \in [0, T] \times D,$$

where, for $t$ in $[0, T]$ and $\sigma \in S$, $h_t(\sigma, \cdot) \in \mathscr{X}$. It is easy to verify, using (7.17) that for $\eta \in H$,

(7.26)
$$(\Pi_s(h_s), \eta_s)_{\mathscr{X}} = \int_D \Pi_s(h_s(\cdot, x))\eta_s(x)\, d\lambda(x).$$

From Theorem 7.2 it now follows that the likelihood ratio for the model (7.25) for observations over $[0, t] \times D$ is given by the formula

(7.27)
$$\rho_t(y) = \exp\left[ \int_0^t \int_D \Pi_s(h_s(\cdot, x)) y_s(x)\, d\lambda(x)\, ds - \frac{1}{2} \int_0^t \int_D \left( \Pi_s(h_s(\cdot, x)) \right)^2 d\lambda(x)\, ds \right].$$

*Likelihood ratios involving random fields.* In somewhat greater generality we now consider a version of the model (7.5) in which the signal process (as well as noise) are multiparameter processes taking values in a possibly infinite-dimensional separable Hilbert space. For simplicity we consider only the case of two parameters.

Let $\mathbf{T} = [0, T] \times [0, T]$. We will denote a generic element $(t_1, t_2)$ of $\mathbf{T}$ by $\mathbf{t}$. Let $\mathscr{X}$ be a separable Hilbert space and let

$$H = \left\{\eta: \mathbf{T} \to \mathscr{X}: \int_0^T \int_0^T \|\eta_{t_1, t_2}\|_{\mathscr{X}}^2 \, dt_1 \, dt_2 < \infty\right\} \tag{7.28}$$

with inner product

$$(\eta, \eta') = \int_0^T \int_0^T (\eta_{t_1, t_2}, \eta'_{t_1, t_2})_{\mathscr{X}} \, dt_1 \, dt_2. \tag{7.29}$$

Then $H$ is a Hilbert space with this inner product.

Let the signal process $\{X_{\mathbf{t}}: \mathbf{t} \in \mathbf{T}\}$ be an $S$-valued random field, defined on a countably additive probability space $(\Omega, \mathscr{A}, \Pi)$ where $S$ is a Polish space. Let

$$h: \mathbf{T} \times S \to \mathscr{X} \tag{7.30}$$

be a measurable function such that

$$\int_0^T \int_0^T \left\| h_{t_1, t_2}\big(X_{t_1, t_2}(\omega)\big)\right\|_{\mathscr{X}}^2 dt_1 \, dt_2 < \infty \qquad \Pi\text{-a.s.} \tag{7.31}$$

Define

$$\xi_{\mathbf{t}}(\omega) = h_{\mathbf{t}}(X_{\mathbf{t}}(\omega)); \qquad \mathbf{t} \in \mathbf{T}.$$

Then $\xi = (\xi_{\mathbf{t}})$ is an $H$-valued random variable on $(\Omega, \mathscr{A}, \Pi)$. Let $e = (e_{\mathbf{t}})$ be the Gaussian white noise on $(H, \mathscr{C}, m)$, where $m$ is the canonical Gaussian measure on $(H, \mathscr{C})$. Let $(E, \mathscr{E}, \alpha) = (\Omega, \mathscr{A}, \Pi) \odot (H, \mathscr{C}, m)$. Now, the abstract statistical model (4.6) takes the form

$$y_{\mathbf{t}} = h_{\mathbf{t}}(X_{\mathbf{t}}) + e_{\mathbf{t}}, \qquad \mathbf{t} \in \mathbf{T}. \tag{7.32}$$

For $\mathbf{t} \in \mathbf{T}$, $\mathbf{t} = (t_1, t_2)$, let

$$H_{\mathbf{t}} = \left\{\eta \in H: \int_0^T \int_0^T \|\eta_{\mathbf{s}}\|_{\mathscr{X}}^2 \, ds_1 \, ds_2 = \int_0^{t_2} \int_0^{t_1} \|\eta_{\mathbf{s}}\|_{\mathscr{X}}^2 \, ds_1 \, ds_2\right\}$$

and let $Q_{\mathbf{t}}$ be the orthogonal projection onto $H_{\mathbf{t}}$. Then it is easy to see that for $\mathbf{t} = (t_1, t_2)$

$$\begin{aligned} (Q_{\mathbf{t}}\eta)_{\mathbf{s}} &= \eta_{\mathbf{s}}, \qquad \text{if } s_1 \le t_1,\ s_2 \le t_2;\ \mathbf{s} = (s_1, s_2), \\ &= 0, \qquad \text{otherwise.} \end{aligned} \tag{7.33}$$

Thus, $Q_{\mathbf{t}} y$ represents the observations $\{y_{\mathbf{s}}: s_1 \le t_1,\ s_2 \le t_2\}$. Let $\eta_{\mathbf{t}} = \alpha[Q_{\mathbf{t}} y]^{-1}$ and let $m_{\mathbf{t}}$ be the canonical Gaussian measure on $H_{\mathbf{t}}$. Then by Theorem 4.6, it follows that

$$\frac{dn_{\mathbf{t}}}{dm_{\mathbf{t}}}(\eta) = \int_\Omega q_{\mathbf{t}}(\omega, \eta) \, d\Pi(\omega), \qquad \eta \in H_{\mathbf{t}}, \tag{7.34}$$

where $q_{\mathbf{t}}(\omega, \eta)$ is given by

$$q_{\mathbf{t}}(\omega,\eta) = \exp\left[\int_0^{t_2}\int_0^{t_1}\left[\left(\left(h_{s_1,s_2}\left(X_{s_1,s_2}(\omega)\right), \eta_{s_1,s_2}\right)_{\mathscr{K}}\right.\right.\right.$$
(7.35)
$$\left.\left. - \tfrac{1}{2}\left\|h_{s_1,s_2}\left(X_{s_1,s_2}(\omega)\right)\right\|^2_{\mathscr{K}}\right) ds_1\, ds_2\right],$$

for $\mathbf{t} = (t_1, t_2) \in \mathbf{T}, \omega \in \Omega, \eta \in H_{\mathbf{t}}$. Let

(7.36) $$\rho_{\mathbf{t}}(\eta) = \frac{dn_{\mathbf{t}}}{dm_{\mathbf{t}}}(Q_{\mathbf{t}}\eta), \qquad \eta \in H.$$

Then, as in the previous section, we will call $\rho_{\mathbf{t}}(y)$ the likelihood ratio for the model (7.32) for observations over $[0, t_1] \times [0, t_2]$, $\mathbf{t} = (t_1, t_2)$.

For a function $g$: $S \to \mathbf{R}$ such that $E|g(X_{\mathbf{t}})| < \infty$, let

(7.37) $$\Pi_{\mathbf{t}}(g) = E_{\alpha}(g(X_{\mathbf{t}})|Q_{\mathbf{t}}y).$$

Then, by the Bayes formula, Theorem 4.9,

(7.38) $$\Pi_{\mathbf{t}}(g) = \frac{\int_{\Omega} g(X_{\mathbf{t}}(\omega))q_{\mathbf{t}}(\omega, y)\, d\Pi(\omega)}{\rho_{\mathbf{t}}(y)}.$$

We now obtain an expression for $\rho_{\mathbf{t}}(y)$ analogous to (7.19). For simplicity in notation, we give the expression for the case $\mathscr{K} = \mathbf{R}$.

THEOREM 7.3. *Assume that* $\mathscr{K} = \mathbf{R}$ *and*

(7.39) $$E_{\Pi}\int_0^T\int_0^T\left|h_{t_1,t_2}\left(X_{t_1,t_2}\right)\right|^4 dt\, dt_2 < \infty.$$

*Then*

$$\rho_{t_1,t_2}(y) = \exp\left(\int_0^{t_1}\int_0^{t_2}\{A(s_1,s_2,y) + B(s_1,s_2,y)\right.$$
(7.40)
$$\left. + C(s_1,s_2,y)\}\, ds_1\, ds_2\right),$$

*where*

(7.41) $$A(s_1,s_2,y) = \Pi_{s_1,s_2}\left(h_{s_1,s_2}\right)y_{s_1,s_2} - \tfrac{1}{2}\Pi_{s_1,s_2}\left(\left[h_{s_1,s_2}\right]^2\right)$$

$$B(s_1,s_2,y) = \int_0^{s_1}\int_0^{s_2}\left[-\Pi_{s_1,s_2}\left(h_{s_1,u_2}h_{u_1,s_2}\right)y_{s_1,u_2}y_{u_1,s_2}\right.$$
(7.42)
$$- \tfrac{1}{2}\Pi_{s_1,s_2}\left(h_{s_1,u_2}h^2_{u_1,s_2}\right)y_{s_1,u_2}$$
$$- \tfrac{1}{2}\Pi_{s_1,s_2}\left(h^2_{s_1,u_2}h_{u_1,s_2}\right)y_{u_1,s_2}$$
$$\left. + \tfrac{1}{4}\Pi_{s_1,s_2}\left(h^2_{s_1,u_2}h^2_{u_1,s_2}\right)\right] du_1\, du_2,$$

*and*

$$C(s_1, s_2, y) = \int_0^{s_1}\int_0^{s_2}\left[\Pi_{s_1,s_2}(h_{s_1,u_2})y_{s_1,u_2} - \tfrac{1}{2}\Pi_{s_1,s_2}(h^2_{s_1,u_2})\right] \cdot\left[\Pi_{s_1,s_2}(h_{u_1,s_2})y_{u_1,s_2} - \tfrac{1}{2}\Pi_{s_1,s_2}(h^2_{u_1,s_2})\right] du_1\, du_2. \tag{7.43}$$

**Proof.** Let us denote, for a fixed $y \in H$,

$$F(t_1, t_2) = \log(\rho_{t_1,t_2}(y)). \tag{7.44}$$

Proceeding as in the proof of Theorem 7.2, it can be proved that:

(i) for fixed $t_1$, $F(t_1, t_2)$ is absolutely continuous in $t_2$;

(ii) for a.e. $t_2$ (fixed), $[\partial F(t_1, t_2)]/\partial t_2$ is absolutely continuous in $t_1$;

(iii) $[\partial F(0, t_2)]/\partial t_2 = 0$;

(iv) $[\partial^2 F(t_1, t_2)]/(\partial t_1\, \partial t_2) = A(t_1, t_2) + B(t_1, t_2) + C(t_1, t_2)$, where $A$, $B$, and $C$ are given by (7.41), (7.42), and (7.43) for $y \in H$ fixed.

Then,

$$\frac{\partial F(t_1, t_2)}{\partial t_2} = \frac{\partial F(0, t_2)}{\partial t_2} + \int_0^{t_1} \frac{\partial^2 F(s_1, t_2)}{\partial s_1\, \partial t_2}\, ds_1 = \int_0^{t_1} \frac{\partial^2 F(s_1, t_2)}{\partial s_1\, \partial t_2}\, ds_1 \tag{7.45}$$

and

$$F(t_1, t_2) = F(t_1, 0) + \int_0^{t_2} \frac{\partial F(t_1, s_2)}{\partial s_2}\, ds_2 = \int_0^{t_2} \frac{\partial F(t_1, s_2)}{\partial s_2}\, ds_2 = \int_0^{t_2}\int_0^{t_1} \frac{\partial^2 F(s_1, s_2)}{\partial s_1\, \partial s_2}\, ds_1\, ds_2. \tag{7.46}$$

Now, (7.40) follows from (7.46) and (iv) above. □

Further simplification of the formula (7.40) requires an extension of the Bayes formula of Theorem 4.9 to a class of functions depending on both $\omega$ and $y$. These and other related problems concerning random fields will be investigated in our later work.

**8. Robustness of the white noise theory and its consistency with the martingale-theoretic approach.** In this section we prove a series of results whose main purpose is to show that the white noise theory of filtering is consistent with the more familiar theory based on martingale calculus and Itô stochastic differential equations. We begin by showing in Theorems 8.1 and 8.2 that the conditional expectation $E(f(X_t)|\mathscr{F}_t^Y)$ is the $\alpha$-lifting of the conditional expectation $\Pi_t(f, y)$ of the white noise setup and, in fact, that it can be obtained as the limit in probability of a suitable sequence of white noise conditional expectations. (In the spirit of Theorem 8.2 a similar approximation result for a wider class of Wiener functionals is given in Theorem 8.4.) We next consider the unnormalized conditional density $p_t(x, y)$, which has been obtained as the unique solution of the finitely additive version of the Zakai equation. Roughly speaking what we show is that, given any path $Y$ in $C([0,T], \mathbf{R}^d)$ and $Y^n$ a sequence in $\mathscr{H}$ approximating $Y$ in the uniform topology, the sequence of corresponding $p_t(x, \dot{Y}^n)$ converges to a unique limit $p_t'(x, Y)$ [independent of the sequence $(Y^n)$] which is a version of the unnormalized conditional density of the conventional Zakai equation. Moreover, the extension thus obtained is robust in the sense in which that term has been used by recent writers [Clark (1978), Davis (1979), and Pardoux (1982)]. Theorem 8.9 and the remark following it are devoted to making the above statements precise. Corollary 8.10 yields a special result involving polygonal path approximations that might be of independent interest. Incidentally, the white noise filtering theory possesses an "internal" robustness property which is described in Lemma 5.4.

In the course of obtaining Theorem 8.9 we also derive in Theorem 8.8 a pathwise solution to the conventional Zakai equation essentially extending a result of Pardoux (1982) for the case of unbounded $h$.

We will first study the relationship between the finitely additive white noise model considered in Section 5 [given by (5.3)] and the usual countably additive model discussed in Section 2 in the signal-noise independent case [given by (1.2)] for a given $\mathbf{R}^d$-valued Markov process $(X_t)$ on a countably additive probability space $(\Omega, \mathscr{A}, \Pi)$.

Let us recall that the model

$$y_s = h_s(X_s) + e_s \tag{8.1}$$

is given on the quasicylindrical probability space

$$(E, \mathscr{E}, \alpha) = (\Omega, \mathscr{A}, \Pi) \odot (H, \mathscr{C}, m). \tag{8.2}$$

Without loss of generality, we can assume that the model (1.2) in the signal-noise independent case is given by

$$Y_t = \int_0^t h_s(X_s)\, ds + Z_t \tag{8.3}$$

on $(\tilde{\Omega}, \tilde{\mathscr{A}}, \tilde{\Pi}) = (\Omega, \mathscr{A}, \Pi) \otimes (\Omega_0, \mathscr{A}_0, \Pi_0)$ where

$$\Omega_0 = C_0([0,T], R^m) \tag{8.4}$$

is the space of continuous functions $\omega_0$: $[0,T] \to \mathbf{R}^m$ s.t. $\omega_0(0) = 0$ with the topology of uniform convergence, and as in Section 2, $\Pi_0$ is the Wiener measure on $(\Omega_0, \mathscr{A}_0)$ and $\{Z_t\}$ are the coordinate maps on $\Omega_0$ given by

$$Z_t(\omega_0) = \omega_0(t) \tag{8.5}$$

so that $\{Z_t\}$ is a Brownian motion under $\Pi_0$.

We will continue to use the notation established in Sections 2 and 5 with the following exception.

We will denote the unnormalized conditional expectation of $f(X_t)$ given $\mathscr{F}_t^Y$ by $\hat{\sigma}_t(f, Y)$ [instead of $\sigma_t(f, Y)$, which was used in Section 2] and the conditional expectation itself will be denoted by $\hat{\Pi}_t(f, Y)$. In this setup, $\hat{\sigma}_t(f, Y)$ is given by

$$\begin{aligned}\hat{\sigma}_t(f, Y(\tilde{\omega})) = \int_\Omega f(X_s(\omega))\exp\Bigg(\sum_{i=1}^m \int_0^t h_s^i(X_s(\omega))\, dY_s^i(\tilde{\omega})\\ - \tfrac{1}{2}\int_0^t |h_s(X_s(\omega))|^2\, ds\Bigg)\, d\Pi(\omega),\end{aligned} \tag{8.6}$$

where the stochastic integral in (8.6) is on the product space $(\Omega \times \tilde{\Omega}, \mathscr{A} \otimes \tilde{\mathscr{A}}, \Pi \otimes \tilde{\Pi})$, so that $\{X_s(\omega)\}, \{Y_s(\tilde{\omega})\}$ are independent under $\Pi \otimes \tilde{\Pi}$. This follows from Remark 2.1.

Let $L_0$: $H \to \mathscr{L}(\Omega_0, \mathscr{A}_0, \Pi_0)$ be defined by

$$L_0(\eta) = \sum_{i=1}^m \int_0^T \eta_s^i\, dZ_s^i, \qquad \eta \in H, \tag{8.7}$$

where $\int_0^T \eta_s^i\, dZ_s^i$ denotes the Wiener integral. It is easy to see that $(L_0, \Pi_0)$ is a representation of $m$. In view of Theorem 4.1, we can and do choose a version of $L_0$ such that $(\omega_0, \eta) \to L_0(\eta)(\omega_0)$ is $\mathscr{A}_0 \times \mathscr{B}(H)$-measurable.

For $\tilde{\omega} = (\omega, \omega_0) \in \tilde{\Omega}$, let

$$L(\eta)(\tilde{\omega}) = L_0(\eta)(\omega_0) \tag{8.8}$$

and

$$\rho(\tilde{\omega}) = \omega. \tag{8.9}$$

Then $(\rho, L, \tilde{\Pi})$ is a representation of $\alpha$ (as observed at the beginning of Section 4). Let $R_\alpha$ be the $\alpha$-lifting corresponding to $(\rho, L, \tilde{\Pi})$.

THEOREM 8.1. *For all $f$: $\mathbf{R}^m \to \mathbf{R}$ with $E|f(X_t)| < \infty$, $0 \le t \le T$,*

$$R_\alpha(\Pi_t(f, y)) = \tilde{\Pi}_t(f, Y), \tag{8.10}$$

*i.e.*,

$$R_\alpha(E_\alpha(f(X_t)|Q_t y)) = E_{\tilde{\Pi}}(f(X_t)|\mathscr{F}_t^Y). \tag{8.10$'$}$$

PROOF. Because of linearity of $\Pi_t(f, y), \hat{\Pi}_t(f, Y)$ in $f$, it suffices to prove (8.10′) for positive $f$. Let $L_1$ be the representation of $n = \alpha[y]^{-1}$ induced by

$(\rho, L, \tilde{\Pi})$ and $R_n$ be the corresponding $n$-lifting (see Lemma 3.8). Then

$$\begin{aligned}
L_1(\eta)(\omega) &= R_n((\eta, y))(\tilde{\omega}) \\
&= R_n((\eta, \xi))(\tilde{\omega}) + R_n((\eta, e))(\tilde{\omega}) \\
&= (\eta, \xi(\tilde{\omega})) + L(\eta)(\tilde{\omega}) \\
&= (\eta, \xi(\omega)) + L_0(\eta)(\omega_0), \\
&= \sum_{i=1}^{m} \int_0^t \eta_s^i h_s^i(X_s(\omega))\, ds + \sum_{i=1}^{m} \int_0^t \eta_s^i\, dZ_s^i(\omega_0), \qquad \tilde{\omega} = (\omega, \omega_0).
\end{aligned} \tag{8.11}$$

Now, the Wiener integral can also be thought of as the Itô integral (of a nonrandom function) and hence from (8.11)

$$L_1(\eta)(\tilde{\omega}) = \left( \sum_{i=1}^{m} \int_0^T \eta_s^i\, dY_s^i \right)(\tilde{\omega}). \tag{8.12}$$

Recall that

$$\sigma_t(f, \eta) = \int_\Omega f(X_t(\omega)) \exp\left( (Q_t \xi(\omega), \eta) - \tfrac{1}{2} \| Q_t \xi(\omega) \|^2 \right) d\Pi(\omega) \tag{8.13}$$

so that

$$\sigma_t(f, \eta) = \int_H \exp\left( (\eta', \eta) - \tfrac{1}{2} \|\eta'\|^2 \right) d\nu(\eta'), \tag{8.14}$$

where for $B \in \mathscr{B}(H)$,

$$\nu(B) = \int f(X_t(\omega)) 1_B(Q_t \xi(\omega))\, d\Pi(\omega). \tag{8.15}$$

Thus, $\sigma_t(f, \cdot) \in \mathscr{I}(H)$ and hence, by Theorems 4.4 and 4.6,

$$\begin{aligned}
R_n(\sigma_t(f, \cdot)) &= \int_H \exp\left( L_1(\eta') - \tfrac{1}{2} \|\eta'\|^2 \right) d\nu(\eta') \\
&= \int_H \exp\left( \sum_{i=1}^{m} \int_0^T \eta_s^i\, dY_s^i - \tfrac{1}{2} \int_0^T |\eta_s|^2\, ds \right) d\nu(\eta).
\end{aligned} \tag{8.16}$$

Since the support of $\nu$ is $H_t$, we can rewrite (8.16) as

$$R_n(\sigma_t(f, \cdot)) = \int_{H_t} \exp\left( \sum_{i=1}^{m} \int_0^t \eta_s^i\, dY_s^i - \tfrac{1}{2} \int_0^t |\eta_s|^2\, ds \right) d\nu_1(\eta), \tag{8.17}$$

where for $B \in \mathscr{B}(H_t)$,

$$\nu_1(B) = \int_\Omega f(X_t(\omega)) 1_B(Q_t \xi(\omega))\, d\Pi(\omega).$$

It can be seen that the right-hand sides of (8.17) and (8.6) are identical and hence we get

$$R_n(\sigma_t(f, \cdot)) = \hat{\sigma}_t(f, Y). \tag{8.18}$$

Now, by Theorem 4.8 (with $Q = I$ and $H' = H$), we have

$$\sigma_t(f, \cdot) \in \mathscr{U}(y) = \mathscr{U}(E, \mathscr{E}, \alpha; \mathbf{R}, y) \tag{8.19}$$

so that

$$R_\alpha(\sigma_t(f, y)) = R_n(\sigma_t(f, \cdot)). \tag{8.20}$$

From (8.18) and (8.20) we get

$$R_\alpha(\sigma_t(f, y)) = \hat{\sigma}_t(f, Y). \tag{8.21}$$

Now, (8.21) for $f = 1$, implies

$$R_\alpha(\sigma_t(1, y)) = \hat{\sigma}_t(1, Y) \tag{8.22}$$

and thus

$$\begin{aligned} R_\alpha(\Pi_t(f, y)) &= \frac{R_\alpha(\sigma_t(f, y))}{R_\alpha(\sigma_t(1, y))} \\ &= \frac{\hat{\sigma}_t(f, Y)}{\hat{\sigma}_t(1, Y)} \\ &= \hat{\Pi}(f, Y). \end{aligned} \tag{8.23}$$

□

The previous theorem shows that the "functional" $\hat{\Pi}_t(f, Y)$ can be obtained from the "functional" $\Pi_t(f, y)$ via an $\alpha$-lifting. The next theorem will give a better picture of this relationship.

Let $\mathscr{H}$ be the reproducing kernel Hilbert space of the Wiener space $(\Omega_0, \mathscr{A}_0, \Pi_0)$. It is well known that

$$\mathscr{H} = \left\{\eta: [0, T] \to \mathbf{R}^m: \eta(0) = 0 \text{ and } \dot{\eta} = \frac{d\eta}{dt} \in H\right\}. \tag{8.24}$$

Let $i$ be the identity mapping from $\mathscr{H}$ into $\Omega_0$, which is continuous. The adjoint map $i^*$ maps the dual $\Omega_0^*$ of $\Omega_0$ into $\mathscr{H}^* = \mathscr{H}$. It is well known that $i^*(\Omega_0^*) = \mathscr{H}_1$, where

$$\mathscr{H}_1 = \{\eta \in H: \dot{\eta} \text{ is of bounded variation}\} \tag{8.25}$$

and the duality map $((\cdot, \cdot))$ from $\mathscr{H}_1 \times \Omega_0 \to \mathbf{R}$ is given by

$$((\eta, \omega_0)) = \sum_{i=1}^m \left[\dot{\eta}_T^i \omega_0^i(T) - \int_0^T \omega_0^i(t)\, d\dot{\eta}_t^i\right], \tag{8.26}$$

where we take the right continuous with left limits version of $\dot{\eta}_t$, continuous at $T$. Such a version exists as $\eta \in \mathscr{H}_1$. The integral $\int_0^T \omega_0^i(t)\, d\dot{\eta}_t^i$ is to be taken as a Riemann–Stieltjes integral. Observe that any $\hat{P} \in \mathscr{P}(\mathscr{H})$ with (range $\hat{P}$) $\subseteq \mathscr{H}_1$ can be extended to $\Omega_0$ as follows. Let $\{\hat{\phi}_j: 1 \le j \le k\} \subseteq \mathscr{H}_1$ be a basis for range $\hat{P}$. Then

$$\hat{P}\omega_0 = \sum_{j=1}^k ((\hat{\phi}_j, \omega_0))\hat{\phi}_j. \tag{8.27}$$

It should be observed that $\hat{P}\omega_0$ is defined pointwise and no limiting procedure is involved in its definition.

Let $\tau\colon \mathscr{H} \to H$ be defined by

$$\tau(\eta) = \hat{\eta}. \tag{8.28}$$

Then, clearly, $\tau$ is a Hilbert space isomorphism (indeed, the norm on $\mathscr{H}$ is defined so that $\tau$ is an isomorphism).

With these notations, we can state the next theorem.

THEOREM 8.2. *Let $\{\hat{P}_j\} \subseteq \mathscr{P}(\mathscr{H})$ be a sequence such that*

$$\left(\text{range}\,\hat{P}_j\right) \subseteq \mathscr{H}_1 \qquad \text{for all } j \geq 1 \tag{8.29}$$

*and*

$$\hat{P}_j \xrightarrow{s} I_{\mathscr{H}} \quad \text{as } j \to \infty. \tag{8.30}$$

*Then for all $f\colon \mathbf{R}^d \to \mathbf{R}$ s.t. $E|f(X_t)| < \infty$,*

$$\Pi_t\big(f, \tau(\hat{P}_j Y(\tilde{\omega}))\big) \to \hat{\Pi}_t(f, Y(\tilde{\omega})) \quad \text{in } \tilde{\Pi} \text{ probability}. \tag{8.31}$$

PROOF. As in the previous theorem, it is enough to prove the assertion for positive $f$. Let $P_j \in \mathscr{P} = \mathscr{P}(H)$ be defined by

$$P_j = \tau \hat{P}_j \tau^{-1}. \tag{8.32}$$

We will first prove that for all $j \geq 1$,

$$R_n\big(\sigma_t(f, P_j\eta)\big) = \sigma_t\big(f, \tau(\hat{P}_j Y(\tilde{\omega}))\big), \qquad \tilde{\Pi}\text{-a.s.} \tag{8.33}$$

For this fix $j \geq 1$, and let $\phi_1, \phi_2, \ldots, \phi_k$ be a basis of range $P_j$ and let $\hat{\phi}_i = \tau^{-1}(\phi_i)$. Then $\hat{\phi}_1, \ldots, \hat{\phi}_k$ is a basis of range $\hat{P}_j$. Now

$$\begin{aligned} \sigma_t(f, P_j\eta) &= \int f(X_t(\omega'))\exp\Big((P_j\eta, Q_t\xi(\omega')) - \tfrac{1}{2}\|Q_t\xi(\omega')\|^2\Big)\, d\Pi(\omega') \\ &= \int f(X_t(\omega'))\exp\left(\sum_{i=1}^{k} (\phi_i, \eta)(\phi_i, Q_t\xi(\omega')) - \tfrac{1}{2}\|Q_t\xi(\omega')\|^2\right) d\Pi(\omega') \end{aligned} \tag{8.34}$$

and hence

$$\begin{aligned} R_n\big(\sigma_t(f, P_j\eta)\big) = \int f(X_t(\omega')) & \\ \cdot\exp\Bigg(\sum_{i=1}^{k} L_1(\phi_i)(\phi_i, Q_t\xi(\omega')) & \\ - \tfrac{1}{2}\|Q_t\xi(\omega')\|^2\Bigg)\, d\Pi(\omega'). & \end{aligned} \tag{8.35}$$

Since $\hat{\phi}_i \in (\text{range}\,\hat{P}_j) \subseteq \mathscr{H}_1$, using (8.12), (8.26), and the integration by parts formula, it can be checked that

$$L_1(\phi_i) = ((\hat{\phi}_i, Y)), \qquad \tilde{\Pi}\text{-a.s.} \tag{8.36}$$

Hence

$$\begin{aligned}\sum_{i=1}^{k} L_1(\phi_i)(\phi_i, Q_t\xi(\omega')) &= \sum_{i=1}^{k} ((\hat{\phi}_i, Y(\tilde{\omega})))(\tau^{-1}\phi_i, \tau^{-1}Q_t\xi(\omega'))\\ &= \sum_{i=1}^{k} ((\hat{\phi}_i, Y(\tilde{\omega})))(\hat{\phi}_i, \tau^{-1}Q_t\xi(\omega')) \qquad (8.37)\\ &= \left(\hat{P}_j Y(\tilde{\omega}), \tau^{-1}Q_t\xi(\omega')\right) \qquad \tilde{\Pi}\text{-a.s.}\\ &= \left(\tau\left[\hat{P}_j Y(\tilde{\omega})\right], Q_t\xi(\omega')\right), \qquad \tilde{\Pi}\text{-a.s.}\end{aligned}$$

Now (8.35) and (8.37) imply (8.33). Proceeding as in Theorem 8.1 we get

$$\begin{aligned}R_n\left(\Pi_t(f, P_j\eta)\right) &= \frac{R_n\left(\sigma_t(f, P_j\eta)\right)}{R_n\left(\sigma_t(1, P_j\eta)\right)}\\ &= \frac{\hat{\sigma}_t\left(f, \tau\left(\hat{P}_j Y(\tilde{\omega})\right)\right)}{\hat{\sigma}_t\left(1, \tau\left(\hat{P}_j Y(\tilde{\omega})\right)\right)} \qquad (8.38)\\ &= \hat{\Pi}_t\left(f, \tau\left(\hat{P}_j Y(\tilde{\omega})\right)\right), \qquad \tilde{\Pi}\text{-a.s.}\end{aligned}$$

Now the required assertion follows from (8.38), Theorem 8.1, and the fact that $\Pi_t(f, \eta) \in \mathscr{U}(y) = \mathscr{U}(E, \mathscr{E}, \alpha; \mathbf{R}, y)$ (in view of Theorem 4.8). □

COROLLARY 8.3. *For each $k \geq 1$, let $\{0 = t_0^k < t_1^k < \cdots < t_{m_k}^k = T\}$ be a partition of $[0, T]$ such that*

$$\lim_{k\to\infty} \sup_{1\leq j\leq m_k} |t_j^k - t_{j-1}^k| = 0. \tag{8.39}$$

*For each $\tilde{\omega} \in \tilde{\Omega}$, let $Y_k(\tilde{\omega})$ be defined by*

$$Y_k(\tilde{\omega})(s) = Y(\tilde{\omega})\left(t_{j-1}^k\right) + \frac{Y(\tilde{\omega})\left(t_j^k\right) - Y(\tilde{\omega})\left(t_{j-1}^k\right)}{t_j^k - t_{j-1}^k}\left(s - t_{j-1}^k\right) \tag{8.40}$$

*for $t_{j-1}^k \leq s \leq t_j^k, 1 \leq j \leq m_k$. Then for all $f\colon \mathbf{R}^d \to \mathbf{R}$ such that $E|f(X_t)| < \infty$,*

$$\Pi_t\left(f, \dot{Y}_k(\tilde{\omega})\right) \to \hat{\Pi}_t\left(f, Y(\tilde{\omega})\right) \quad \text{in } \tilde{\Pi} \text{ probability}, \tag{8.41}$$

*where $\dot{Y}_k(\tilde{\omega}) = (d/ds)Y_k(\tilde{\omega})$. Observe that here, $Y_k$ is the usual polygonal approximation to $Y$ and $\dot{Y}_k$ is an approximation of the nonexistent "derivative $\dot{Y}$" of $Y$.*

This result can be deduced from Theorem 8.2 by identifying $Y_k$ with $\hat{P}_k Y$ for an appropriate sequence $\{\hat{P}_k\}$ satisfying the conditions of Theorem 8.2.

Proceeding as in the proof of the previous result, we can also derive the following result on approximation of Wiener functionals.

THEOREM 8.4. *Let $\mu \in \mathscr{M}_0(\mathscr{H})$ and let $\hat{f}, \hat{g}$ be the Wiener functionals defined by*

$$\hat{f}(\omega_0) = \int_{\mathscr{H}} \exp\left(\sum_{i=1}^{m} \int_0^T \dot{\eta}_s^i \, dZ_s^i(\omega_0) - \tfrac{1}{2}\int_0^T |\dot{\eta}_s|^2 \, ds\right) d\mu(\eta) \tag{8.42}$$

*and*

$$\hat{g}(\omega_0) = \int_{\mathscr{H}} \exp\left( i \sum_{j=1}^{m} \int_0^T \dot{\eta}_s^j \, dZ_s^j(\omega_0) \right) d\mu(\eta), \qquad \Pi_0\text{-}a.s. \tag{8.43}$$

*where* $\int_0^T \dot{\eta}_s^j \, dZ_s^j$ *is a Wiener integral.*

*For* $\theta \in \mathscr{H}$, *let* $f, g$ *be functionals on* $\mathscr{H}$ *defined by*

$$f(\theta) = \int_{\mathscr{H}} \exp\left( \sum_{j=1}^{m} \int_0^T \dot{\eta}_s^j \dot{\theta}_s^j \, ds - \tfrac{1}{2} \int_0^T |\dot{\eta}_s|^2 \, ds \right) d\mu(\eta). \tag{8.44}$$

*and*

$$g(\theta) = \int_{\mathscr{H}} \exp\left( i \sum_{j=1}^{m} \int_0^T \dot{\eta}_s^j \dot{\theta}_s^j \, ds \right) d\mu(\eta). \tag{8.45}$$

*Then for all sequences* $(\hat{P}_j) \subseteq \mathscr{P}(\mathscr{H})$ *such that* $\hat{P}_j \xrightarrow{s} I$ *and* (range $\hat{P}_j$) $\subseteq \mathscr{H}_1$,

$$f(\hat{P}_j Z) \to \hat{f} \quad in \quad \mathscr{L}^1(\Omega_0, \mathscr{A}_0, \Pi_0) \tag{8.46}$$

*and*

$$g(\hat{P}_j Z) \to \hat{g} \quad in \quad \mathscr{L}^1(\Omega_0, \mathscr{A}_0, \Pi_0) \tag{8.47}$$

*as* $j \to \infty$.

The next corollary follows from Theorem 8.4 by choosing $\hat{P}_j$ as given in Corollary 8.3.

COROLLARY 8.5. *Let* $f, g, \hat{f}, \hat{g}$ *be as in Theorem* 8.4. *Let* $\{0 = t_0^k < t_1^k < \cdots < t_{m_k}^k = T\}$, $k \geq 1$ *be a sequence of partitions of* $[0, T]$ *satisfying* (8.39). *Let* $Z_k(\omega_0), \omega_0 \in \Omega_0$ *be defined by*

$$Z_k(\omega_0)(s) = \omega_0(t_{j-1}^k) + \frac{\omega_0(t_j^k) - \omega_0(t_{j-1}^k)}{t_j^k - t_{j-1}^k}(s - t_{j-1}^k), \tag{8.48}$$

*for* $t_{j-1}^k \leq s \leq t_j^k, 1 \leq j \leq m_k$. *Then*

$$f(Z_k) \to \hat{f} \quad in \quad \mathscr{L}^1(\Omega_0, \mathscr{A}_0, \Pi_0)$$

*and*

$$g(Z_k) \to \hat{g} \quad in \quad \mathscr{L}^1(\Omega_0, \mathscr{A}_0, \Pi_0).$$

In Theorem 8.2 we showed that the functional $\tilde{\Pi}_t(f, Y)$ can be obtained from the functional $\Pi_t(f, y)$. However, the convergence in (8.31) is convergence in $\tilde{\Pi}$ probability and hence given an observation path $Y$, Theorem 8.2 does not show us how to obtain $\tilde{\Pi}_t(f, Y)$ for this path $Y$ from $\{\Pi_t(f, y): y \in H\}$. We will now show that under additional conditions on the signal process $(X_t)$ and $h$, the unnormalized conditional density $\hat{p}_t(x, y)$ of $X_t$ can be obtained as the continuous extension of $\{p_t(x, \tau Y): Y \in \mathscr{H}\}$, where $\tau$ is the isomorphism between $\mathscr{H}$ and $H$ given by (8.28). As a consequence, $\tilde{\Pi}_t(f, Y)$ for a given $Y$ can be obtained as the limit of $\Pi_t(f, \dot{Y}_k)$, where $Y_k \in \mathscr{H}$ is any sequence converging to $Y$ in uniform norm.

Throughout the rest of this section, we assume that the process $(X_t)$ and the function $h$ satisfy the conditions of Theorem 5.6 [i.e., (5.37), (5.38), (5.40), (5.56), (5.57), (5.58), and (5.39) hold].

Let $\Omega^* = \{\omega_0 \in \Omega_0: t \to \omega_0(t) \text{ is Hölder continuous}\}$. Recall that $Z_t$ is the coordinate process on $\Omega_0$ defined by (8.5). We will use $Z(\omega_0)$ to denote the path $Z_t(\omega_0)$ [so that $Z(\omega_0) = \omega_0$] for notational convenience. For each $Z \in \Omega^*$, consider the PDE

$$\begin{aligned}\frac{\partial \Psi_t}{\partial t}(x,Z) &= \tfrac{1}{2}\sum_{i,j=1}^{d} a_{ij}(t,x)\frac{\partial^2 \Psi_t(x,Z)}{\partial x^i \partial x^j} \\ &\quad + \sum_{i=1}^{d} \hat{b}_i(t,x,Z)\frac{\partial \Psi_t(x,Z)}{\partial x^i} + \hat{c}(t,x,Z)\Psi_t(x,Z),\end{aligned} \tag{8.49}$$

where

$$\hat{b}_i(\cdot,\cdot,Z) = -b_i + \sum_{j=1}^{d}\frac{\partial a_{ij}}{\partial x^j} + \tfrac{1}{2}\sum_{j=1}^{d} a_{ij}\frac{\partial}{\partial x^j}\left(\sum_{k=1}^{m} h_t^k Z_t^k\right) \tag{8.50}$$

and

$$\begin{aligned}\hat{c}(\cdot,\cdot,Z) &= -\tfrac{1}{2}|h_t|^2 - \sum_{j=1}^{d}\frac{\partial b_i}{\partial x_i} + \tfrac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2 a_{ij}}{\partial x^i \partial x^j} \\ &\quad + \tfrac{1}{2}\sum_{i,j=1}^{d} a_{ij}\left[\frac{\partial^2}{\partial x^i \partial x^j}\left(\sum_{k=1}^{m} h_t^k Z_t^k\right) + \frac{\partial}{\partial x_i}\left(\sum_{k=1}^{m} h_t^k Z_t^k\right) \cdot \frac{\partial}{\partial x_j}\left(\sum_{k=1}^{m} h_t^k Z_t^k\right)\right] \\ &\quad + \sum_{i=1}^{d}\left[\left(-b_i + \sum_{j=1}^{d}\frac{\partial a_{ij}}{\partial x^j}\right)\frac{\partial}{\partial x^i}\left(\sum_{k=1}^{m} h_t^k Z_t^k\right)\right] - \sum_{k=1}^{m}\frac{\partial h_t^k}{\partial t} Z_t^k.\end{aligned} \tag{8.51}$$

Equation (8.49) is known as the robust form of Zakai's equation and is formally the equation satisfied by

$$\Psi_t(x,Z) = \hat{p}_t(x,Z)\exp\left(-\sum_{k=1}^{m} h_t^k(x) Z_t^k\right),$$

where $\hat{p}_t(x,Z)$ is a solution of SPDE (2.33).

Our first result is on the existence, uniqueness, and continuous dependence (on $Z$) of the solution to the PDE (8.49).

THEOREM 8.6. *Suppose the conditions of Theorem* 5.6 *hold. Then*

(i) *For all $Z \in \Omega^*$, the PDE* (8.49) *with the initial condition $\Psi_0(x,Z) = \phi(x)$ has a unique solution $\Psi_t(x,Z)$ in the class $\mathscr{G}$. [See Theorem* 5.5 *for definition of $\mathscr{G}$.]*

(ii) *The mapping $Z \to \Psi\cdot(\cdot,Z)$ from $\Omega^*$ into $C([0,T]\times \mathbf{R}^d)$ (equipped with the topology of uniform convergence on compacta) is continuous.*

We are now in a position to prove parts (i) and (ii) of Theorem 5.6, which we had not proved earlier.

PROOF OF THEOREM 5.6. For $\eta \in H$, let

$$p_t(x,\eta) = \Psi_t(x,Z)\exp\left(\sum_{i=1}^{m} h_t^i(x)Z_t^i\right), \tag{8.52}$$

where

$$Z_t^i = \int_0^t \eta_s^i\, ds. \tag{8.53}$$

Here, $\Psi_t(x,Z)$ is the unique solution to (8.49) in the class $\mathscr{G}$. Using the product rule, it is easy to check that $p_t(x,\eta)$ defined by (8.52) satisfies (5.41) for a.e. $t$. Clearly also (5.59) holds. If $p'$ satisfies (5.59) and satisfies (5.41) for a.e. $t$ and if $\Psi'$ is defined by

$$\Psi_t'(x,Z) = p_t'(x,\eta)\exp\left(-\sum_{i=1}^{m} h_t^i(x)Z_t^i\right), \tag{8.54}$$

where $Z,\eta$ are related by (8.53), then it follows that $\Psi'$ is a solution of (8.49) and belongs to the class $\mathscr{G}$ and hence, in view of Theorem 8.6, $\Psi' = \Psi$ and also $p' = p$. This proves part (i) of Theorem 5.6. Part (ii) follows from the definition of $p$ given in (8.52) and Theorem 8.6(ii). □

THEOREM 8.7. *Fix $f$ to be a bounded continuous function from $\mathbf{R}^d$ into $\mathbf{R}$ and let $0 < t < T$. Recall that*

$$\begin{aligned}\hat{\sigma}_t(f,Z) = \int & f(X_t(\omega)) \\ & \cdot\exp\left(\sum_{i=1}^{m}\int_0^t h_s^i(X_s(\omega))\, dZ_s^i - \tfrac{1}{2}\int_0^t |h_s(X_s(\omega))|^2\, ds\right) d\Pi(\omega),\end{aligned}$$

*where the stochastic integral $\int_0^t h_s^i(X_s(\omega))\, dZ_s^i$ is the Itô integral on the product space $(\Omega,\mathscr{A},\Pi)\otimes(\Omega_0,\mathscr{A}_0,\Pi_0)$, $\Pi_0$ being the Wiener measure on $\Omega_0$. Then there exists a version $\sigma_t'(f,Z)$ defined for all $Z \in \Omega_0$, of $\hat{\sigma}_t(f,Z)$, i.e.*

$$\sigma_t'(f,Z) = \hat{\sigma}_t(f,Z), \qquad \Pi_0\ a.e.\ (Z), \tag{8.55}$$

*such that the mapping $Z \to \sigma_t'(f,Z)$ is continuous from $\Omega_0$ into $\mathbf{R}$ (wrt the topology of uniform convergence on $\Omega_0$).*

For the proof we refer the reader to Kallianpur–Karandikar (1983b, Proposition 8.4). The desired version is given by

$$\begin{aligned}\sigma_t'(f,Z) = \int_\Omega f(X_t(\omega))\exp\Bigg(\sum_{i=1}^{m}\left(h_t^i(X_t(\omega))Z_t^i - \int_0^t Z_s^i\, dh_s(X_s(\omega))\right) \\ - \tfrac{1}{2}\int_0^t |h_s(X_s(\omega))|^2\, ds\Bigg)\, d\Pi(\omega).\end{aligned}$$

We are now ready to prove the two main results of this section. The first one is on the existence of a "pathwise version" of the unnormalized conditional density for the model (8.3).

THEOREM 8.8. *For $Z \in \Omega^*$ let $\Psi_t(x, Z)$ be the unique solution to* (8.49) *with the initial condition $\Psi_0(x, Z) = \phi(x)$ (see Theorem* 8.6) *and let $\hat{p}_t(x, Z)$ be defined by*

$$\hat{p}_t(x, Z) = \Psi_t(x, Z)\exp\left(\sum_{k=1}^{m} h_t^k(x)Z_t^k\right), \qquad Z \in \Omega^*. \tag{8.56}$$

*Note that since $Z(\omega_0) \in \Omega^*$ $\Pi_0$-a.e., $Y(\tilde{\omega})$ given by* (8.3) *belongs to $\Omega^*$ $\tilde{\Pi}$-a.s. Then $\{\hat{p}_t(x, Y(\tilde{\omega}))\colon \tilde{\omega} \in \tilde{\Omega}\}$ is a version of the unnormalized conditional density of $X_t$ given $\mathscr{F}_t^Y$ [where $Y$ is given by* (8.3)], *i.e., for $f \in C_b(\mathbb{R}^d)$,*

$$E_{\tilde{\Pi}}\left( f(X_t)|\mathscr{F}_t^Y\right) = \frac{\int f(x)\hat{p}_t(x, Y)\,dx}{\int \hat{p}_t(x, Y)\,dx}, \qquad \tilde{\Pi}\text{-a.e.} \tag{8.57}$$

PROOF. Using the integration by parts formula, it follows that for $\eta \in H$, $0 \le t \le T$

(8.58)
$$\int_0^t h_s(X_s(\omega))\eta_s\,ds = h_t(X_t(\omega))Z_t - \int_0^t Z_s\,dh_s(X_s(\omega)), \qquad \Pi\text{-a.e. } \omega,$$

where

$$Z_t = \int_0^t \eta_s\,ds, \qquad 0 \le t \le T. \tag{8.59}$$

Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be a bounded continuous function. From (8.58) and the definitions of $\sigma_t'(f, Z)$, $\sigma_t(f, \eta)$, it follows that if $\eta \in H$, $Z \in \Omega_0$ are related by (8.59), then

$$\sigma_t'(f, Z) = \sigma_t(f, \eta). \tag{8.60}$$

Also, from (8.52) and (8.56), it follows that if $\eta, Z$ are related by (8.59), then we have

$$\hat{p}_t(x, Z) = p_t(x, \eta), \qquad 0 \le t \le T, x \in \mathbb{R}^d. \tag{8.61}$$

Now, (8.60), (8.61), and Theorem 5.5 imply that for all $Z \in \mathscr{H}$,

$$\sigma_t'(f, Z) = \int f(x)\hat{p}_t(x, Z)\,dx. \tag{8.62}$$

The mapping $Z \to \hat{p}_t(x, Z)$ is continuous in the topology of uniform convergence on compacta [by Theorem 8.6 and (8.56)] and also the mapping $Z \to \sigma_t'(f, Z)$ is continuous by Theorem 8.7. Thus, as $\mathscr{H}$ is dense in $\Omega_0$, the identity (8.62) continues to hold for all $Z \in \Omega^*$.

Now, as the measure $\tilde{\Pi}[Y]^{-1}$ is equivalent to $\Pi_0$ (on $\Omega_0$) the relation (8.55) implies that

$$\begin{aligned} \hat{\sigma}_t(f, Y) &= \sigma_t'(f, Y) \qquad \tilde{\Pi}\text{-a.e.} \\ &= \int f(x)\hat{p}_t(x, Y)\,dx, \qquad \tilde{\Pi}\text{-a.e.} \end{aligned} \tag{8.63}$$

since $Y \in \Omega^*$ a.e. $\tilde{\Pi}$. This and the Bayes formula, Theorem 4.9, prove (8.57) for all $f \in C_b(\mathbf{R}^d)$. □

Pardoux (1979) has obtained the above result under conditions that are somewhat different from ours and include boundedness assumptions on $h$, $\partial h/\partial t$, $\partial h/\partial x_j$, and $\partial^2 h/(\partial x_i\,\partial x_j)$. More recently, Pardoux (1982) has shown it under less restrictive assumptions that permit $h$ to have linear growth. It may be noted, however, that our proof essentially uses the Bayes formula and does not rely on stochastic differential equations.

The problem of robust filtering in the unbounded case has also been considered in Baras, Blankenship, and Mitter (1981) and Baras, Blankenship, and Hopkins (1983). They show under certain conditions that Equation (8.49) admits a unique solution $\overline{\Psi}$ but do not identify $p_t$ defined by (8.56) to be the unnormalized conditional density—a point of considerable importance from the standpoint of filtering theory.

The next result, whose proof is really contained in the proof of the previous theorem, shows that (under certain conditions) the unnormalized conditional density of $X_t$ for the model (8.3) is a continuous extension of the unnormalized conditional density of $X_t$ for the model (8.1).

THEOREM 8.9. *Suppose that the conditions of Theorem* 5.6 *are satisfied. Let* $p_t(x, \eta)$ *be the unnormalized conditional density of* $X_t$ *given by Theorem* 5.6. *Then*

(i) *There exists a continuous function*

$$\hat{p}_\cdot(\cdot, Z)\colon \Omega^* \to C([0, T] \times \mathbf{R}^d) \tag{8.64}$$

*such that for all* $\eta \in H$,

$$\hat{p}_t(x, Z) = p_t(x, \eta), \qquad 0 \le t \le T, x \in \mathbf{R}^d, \tag{8.65}$$

*where* $Z \in \mathscr{H}$ *is given by*

$$Z_t = \int_0^t \eta_s\,ds. \tag{8.66}$$

(ii) $\hat{p}_t(x, Y)$ *is a version of the unnormalized conditional density of* $X_t$ *given* $\mathscr{F}_t^Y$.

PROOF. Indeed, we have already shown that the choice of $\hat{p}$ given by (8.56) satisfies all the requirements. □

REMARK 8.1. If the conditions of Theorems 8.8 and 8.9 are strengthened by the assumption of boundedness in place of linear growth in (5.38) and (5.58) then these theorems remain true with $\Omega^*$ replaced by $\Omega_0$ in their statement. For $Z \in \Omega_0$, $\tilde{c}(t,x,Z)$ may not be locally Hölder continuous in $(t,x)$ so that we cannot appeal to the results of Besala (1979) on parabolic PDEs with unbounded coefficients. However, for fixed $t$, $0 \le t \le T$, $Z \in \Omega_0$, $\tilde{c}(t,x,Z)$ is locally Hölder continuous in $x$ and then instead, we can use Theorem 12 of Chapter 1 in Friedman (1964) to obtain a solution $\Psi_t(x,Z)$ of (8.49) for all $Z \in \Omega_0$.

For the rest of the proof, the same arguments given earlier for the case $Z \in \Omega^*$ yield the required results. Then we have $\hat{p}_t(x,Y)$ for every $Y \in \Omega_0$.

REMARK 8.2. Theorems 8.8 and 8.9 (and their versions with $\Omega^*$ replaced by $\Omega_0$ under the stronger conditions as outlined in Remark 8.1) show that the robustness results of the stochastic calculus approach to nonlinear filtering can be obtained from the finitely additive white noise theory.

COROLLARY 8.10. *Suppose that the conditions of Theorem* 5.6 *are satisfied. Let* $\{Y_k\}$ *be the polygonal approximation to* $Y$ *given by* (8.40) *where the sequence of partitions* $\{t_i^k\}$ *satisfies* (8.39). *Then* $Y_k \to Y$ *uniformly and hence by Theorem* 8.9,

$$p.(\cdot,\dot{Y}_k) \to \hat{p}.(\cdot,Y), \qquad \tilde{\Pi}\text{-}a.e. \tag{8.67}$$

*In particular, for all continuous* $f\colon \mathbb{R}^d \to \mathbb{R}$ *with compact support,*

$$\sigma_t(f,\dot{Y}_k) \to \hat{\sigma}_t(f,Y), \qquad \tilde{\Pi}\text{-}a.e. \tag{8.68}$$

*It follows from this and* (8.23) *that for such an* $f$,

$$\Pi_t(f,\dot{Y}_k) \to \hat{\Pi}_t(f,Y), \qquad \tilde{\Pi}\text{-}a.e. \tag{8.69}$$

**9. Directions of further work.** The theory worked out in this article is by no means complete. We are aware of many questions that could be asked to which we do not now have an answer. We conclude by listing some of the problems that can be treated (and on which work is already in progress) using our approach although a further development of the theory of Sections 3 and 4 might be necessary to ensure complete success:

(1) Finitely additive white noise theory of prediction and smoothing.

(2) Robustness in the general (infinite-dimensional) measure-valued case.

(3) Innovations problem for the nonlinear filtering problem. [Some progress on this problem has been made in the recent unpublished work of Seo and Mazumdar (1984).]

(4) Applications of the theory to statistical parameter estimation and detection problems and to random fields.

(5) Further analysis of the optimal filter as a functional of the observation $y$.

(6) Robustness of the filtering solution in a class of finitely additive models in which the noise deviates from Gaussian white noise.

(7) Discrete approximations to the filtering problem. (For discretization of Itô-type SDEs continuing work is reported by Pardoux and his colleagues [Pardoux and Talay (1984)].)

(8) Extension of the white noise theory to the more general model in which signal and noise need not be independent.

A very important and interesting problem is the study of the optimal filter as a function of the observations. Lemma 5.4 (which has a very simple proof) gives some information on this question. M. Chaleyat-Maurel, in a paper to appear in the *J. Funct. Anal.*, has shown that (for the signal and noise dependent case, under the assumption that the coefficients are $C^\infty$) the optimal filter in the stochastic calculus theory is a $C^\infty$ function of the observation, in the sense of Malliavin. In other words, it possesses directional derivatives of any order in directions belonging to the Sobolev space $H^1$. It is natural to expect analogous results to hold in the setup of the white noise theory.

We would like, in conclusion, to observe that the theory outlined in this paper has not been developed in any polemical spirit, to be set in opposition to the conventional theory that has so many brilliant achievements to its credit. From one point of view, the white noise theory may be regarded as another "language" in which to formulate and solve the nonlinear filtering problem (at least, in the signal and noise independent case). The results of Section 8 show that it is consistent with the "language" of the conventional theory. However, it is not clear at this stage whether the white noise approach might not lead to robustness results that have no counterpart in the conventional theory.

We thank one of the reviewers for his insightful comments and also for providing the reference to the Chaleyat-Maurel paper.

## Appendix

*Markov process and its extended generator.* Let $(\Omega, \mathscr{A}, \Pi)$ be a countably additive probability space and let $(\mathscr{A}_t)_{0 \le t < \infty}$ be an increasing family of sub-$\sigma$ fields of $\mathscr{A}$ such that $\mathscr{A}_0$ contains all $\Pi$-null sets in $\mathscr{A}$. For a measurable space $(S_1, \mathscr{S}_1)$, let $\mathscr{J}(S_1, \mathscr{S}_1)$ be the class of real-valued bounded measurable functions on $(S_1, \mathscr{S}_1)$. Let $(S, \mathscr{S})$ be a measurable space. Recall that an $S$-valued $\{\mathscr{A}_t\}$-adapted process $\{X_t\}$ is a family of mappings from $\Omega$ into $S$ such that for all $t$,

$$X_t^{-1}B \in \mathscr{A}_t, \qquad \text{for all } B \in \mathscr{S}. \tag{A.1}$$

Definition. Let $\{X_t\}$ be an $S$-valued $\{\mathscr{A}_t\}$-adapted process. Say that $\{X_t\}$ is Markov wrt the family $\{\mathscr{A}_t\}$ if for all $i \ge 1, 0 \le s \le t_1 < t_2 < \cdots t_i < \infty$, $f \in \mathscr{J}(S^i, \mathscr{S}^i)$, where $(S^i, \mathscr{S}^i)$ is the $i$-fold product of $(S, \mathscr{S})$,

$$E_\Pi\big(f(X_{t_1}, X_{t_2}, \dots, X_{t_i}) | \mathscr{A}_s\big) = g(X_s), \qquad \Pi\text{-a.s.} \tag{A.2}$$

for some $g \in \mathscr{J}(S, \mathscr{S})$.

The equation (A.2) is known as the Markov property and is usually given in many equivalent forms. One of the most commonly used versions is

$$(A.3)\quad E_\Pi\big(f(X_{t_1},\dots,X_{t_i})|\mathscr{A}_s\big)=E_\Pi\big(f(X_{t_1},\dots,X_{t_i})|\sigma(X_s)\big),\qquad \Pi\text{-a.s.},$$

which is easily seen to be equivalent to (A.2). [Here, $\sigma(X_s)$ denotes the smallest $\sigma$ field on $\Omega$ wrt which $X_s$ is measurable.]

We will assume that the process $\{X_t\}$ is $\mathscr{A}_t$-progressively-measurable, i.e., for all $t$, the mapping $(s,\omega)\to X_s(\omega)$ from $[0,t]\times\Omega$ is measurable wrt $\mathscr{B}([0,t])\otimes\mathscr{A}_t$.

We also assume that the Markov process $\{X_t\}$ admits a transition probability function $P(\cdot,\cdot,\cdot,\cdot)$, i.e., there exists a function

$$P\colon\{(s,x,t,B)\colon 0\le s\le t<\infty, x\in S, B\in\mathscr{S}\}\to\mathbf{R}$$

such that

(i) for all $0\le s\le t, x\in S$, $P(s,x,t,\cdot)$ is a probability measure on $(S,\mathscr{S})$;

(ii) for all $t<\infty, B\in\mathscr{S}$, the mapping $(s,x)\to P(s,x,t,B)$ is $\mathscr{B}([0,t])\otimes\mathscr{S}$-measurable;

(iii) for all $0\le s\le t_1\le t_2, x\in S, B\in\mathscr{S}$,

$$(A.4)\qquad P(s,x,t_2,B)=\int_S P(s,x,t_1,dz)P(t_1,z,t_2,B);$$

(iv) for all $0\le s\le t, B\in\mathscr{S}$,

$$(A.5)\qquad E_\Pi(1_B(X_t)|\mathscr{A}_s)=P(s,X_s,t,B),\qquad \Pi\text{-a.s.}$$

It should be observed that (A.2) implies the existence of a function $P$ satisfying (A.5), but in general it may not be possible to choose a version of $P$ satisfying the conditions (i), (ii), and (iii) above.

The Markov process $\{X_t\}$ is called time homogeneous if

$$P(s,x,t,B)=P(0,x,t-s,B)$$

for all $0\le s\le t<\infty, x\in S, B\in\mathscr{S}$. In what follows, we do not assume that $\{X_t\}$ is time homogeneous.

We want to remark that a Markov process on a finite interval $[0,T]$ can be defined exactly as before: We require that (A.2) holds for $0\le s\le t_1\le t_2\le\cdots\le t_i\le T$. In this case we say that $\{X_t\}$ admits a transition probability function $P(\cdot,\cdot,\cdot,\cdot)$ if there exists a function $P\colon\{s,x,t,B)\colon 0\le s\le t\le T, x\in S, B\in\mathscr{S}\}\to\mathbf{R}$ satisfying (i)–(iv) above, with $t,t_1,t_2\le T$.

Given an $S$-valued Markov process $\{X_t\}_{0\le t\le T}$ wrt $\{\mathscr{A}_t\}$, which admits a transition probability function $P$, let us define

$$\mathscr{A}_t=\mathscr{A}_T,\qquad t\ge T,$$
$$X_t=X_T,\qquad t\ge T$$

and

$$P(s,x,t,B)=P(s,x,T,B),\qquad \text{for all } t\ge T, x, B \text{ if } s<T,$$
$$=1_B(x),\qquad \text{for all } s\ge T, t\ge T, x, B.$$

Then it is easy to check that $(X_t)_{0 \le t < \infty}$ is an $S$-valued Markov process wrt the family $\{\mathscr{A}_t\}$ and that $P(\cdot, \cdot, \cdot, \cdot)$ is its transition probability function.

Thus, all the definitions and results which we will give later for a Markov process on $[0, \infty)$ also apply to a Markov process on $[0, T]$.

We now associate a "two parameter" semigroup with the Markov process $\{X_t\}$ as follows.

For $f \in \mathscr{J}(S, \mathscr{S}), 0 \le s \le t$, let

$$(V_t^s f)(x) = \int_S f(\zeta) P(s, x, t, d\zeta). \tag{A.6}$$

In view of our assumptions on $P$, it follows that $V_t^s f \in \mathscr{J}(S, \mathscr{S})$ and that for $f \in \mathscr{J}(S, \mathscr{S}), 0 \le s \le t_1 \le t_2$

$$V_{t_2}^s f = V_{t_2}^{t_1}(V_{t_1}^s f). \tag{A.7}$$

The semigroup $V_t^s$ is related to the Markov process $\{X_t\}$ via the following property, which can be checked easily using (A.5): For all $f \in \mathscr{J}(S, \mathscr{S}), 0 \le s \le t$,

$$E_\Pi(f(X_t)|\mathscr{A}_s) = (V_t^s f)(X_s). \quad \text{a.s. } \Pi. \tag{A.8}$$

We now want to define the notion of a generator of the Markov process $\{X_t\}$. We have not assumed $\{X_t\}$ to be time homogeneous and hence we first associate a one parameter semigroup $(T_t)$ with the process $\{X_t\}$ as follows: Let $\hat{S} = [0, \infty) \times S$ and let $\hat{\mathscr{S}} = \mathscr{B}([0, \infty)) \otimes \mathscr{S}$. For $f \in \mathscr{J}(\hat{S}, \hat{\mathscr{S}}), 0 \le t < \infty$, let

$$(T_t f)(s, x) = \int_S f(s + t, \zeta) P(s, x, s + t, d\zeta). \tag{A.9}$$

It is easy to see that

$$(T_t f)(s, x) = [V_{s+t}^s f(s + t, \cdot)](x). \tag{A.10}$$

From the property (iii) of $P$ or from (A.7), (A.10) it follows that $T_t$ is a one parameter semigroup, i.e.,

$$T_{t_2}(T_{t_1} f) = T_{t_1 + t_2} f. \tag{A.11}$$

The semigroup $T_t$ is related to the Markov process $\{X_t\}$ via the property

$$E_\Pi(f(s + t, X_{s+t})|\mathscr{A}_s) = (T_t f)(s, X_s), \quad \Pi\text{-a.s.}, \tag{A.12}$$

which follows from (A.8), (A.10) or directly from (A.5). The relation (A.12) implies (the well known fact) that the process $\hat{X}_t = (t, X_t)$ is a time homogeneous $S$-valued Markov process wrt the family $(\mathscr{A}_t)$, and that $T_t$ is the semigroup associated with $\hat{X}_t$ in the usual terminology [see Dynkin (1964)].

For $f_i, f \in \mathscr{J}(\hat{S}, \hat{\mathscr{S}})$, say that $f_i \to f$ weakly if $f_i \to f$ pointwise and $\{f_i\}$ is uniformly bounded.

Let

$$\mathscr{J}_0 = \{f \in \mathscr{J}(\hat{S}, \hat{\mathscr{S}}): T_t f \to f \text{ weakly as } t \downarrow 0\}.$$

We assume that $\mathscr{J}_0$ is dense in $\mathscr{J}(\hat{S}, \hat{\mathscr{S}})$, i.e.,

$$\text{for all} \quad f \in \mathscr{J}(\hat{S}, \hat{\mathscr{S}}), \exists f_i \in \mathscr{J}_0, \text{s.t. } f_i \to f \text{ weakly}. \tag{A.13}$$

This is a technical condition on $\{X_t\}$ and means that $(T_t f)$ is continuous at 0 for a rich class of functions $f$.

Let $\mathscr{D}$ be the class of functions $f \in \mathscr{J}_0$ such that there exists $g \in \mathscr{J}_0$ satisfying for all $t$,

$$(T_t f)(s,x) = f(s,x) + \int_0^t (T_u g)(s,x)\,du, \qquad (s,x) \in \hat{S}. \tag{A.14}$$

Since in (A.14), $g$ is required to belong to $\mathscr{J}_0$, it follows that $(T_u g)(s,x)$ is a continuous function of $u$. Thus (A.14) implies that

$$\frac{T_t f - f}{t} \to g \quad \text{weakly as } t \downarrow 0 \tag{A.15}$$

and thus (A.14) determines $g$ uniquely.

For $f \in \mathscr{D}$, let

$$\mathscr{L} f = g, \tag{A.16}$$

where $g$ is related to $f$ by (A.14). $\mathscr{L}$ will be called the *extended generator* of the Markov process $X_t$ and $\mathscr{D}$ is the domain of $\mathscr{L}$.

The following lemma gives an important property of the generator $\mathscr{L}$.

LEMMA A.1. *Let $f \in \mathscr{D}$. Then*

$$\left\{ f(\hat{X}_t) - \int_0^t (\mathscr{L} f)(\hat{X}_u)\,du, \mathscr{A}_t \right\} \quad \textit{is a martingale.} \tag{A.17}$$

PROOF. To prove (A.17), it suffices to show that for all $s, t$,

$$E_\Pi\left( f(\hat{X}_{t+s}) - \int_s^{s+t} (\mathscr{L} f)(\hat{X}_u)\,du \,|\, \mathscr{A}_s \right) = f(\hat{X}_s), \qquad \Pi\text{-a.s.} \tag{A.18}$$

From the definition of $\mathscr{L}$, we have

$$(T_t f)(\hat{X}_s) = f(\hat{X}_s) + \int_0^t [T_u(\mathscr{L} f)](\hat{X}_s)\,du, \tag{A.19}$$

which in view of (A.12) implies

$$\begin{aligned} E_\Pi\left( f(\hat{X}_{t+s} | \mathscr{A}_s \right) &= f(\hat{X}_s) + \int_0^t E_\Pi\left( (\mathscr{L} f)(\hat{X}_{s+u}) | \mathscr{A}_s \right) du \\ &= f(\hat{X}_s) + E_\Pi\left( \int_0^t (\mathscr{L} f)(\hat{X}_{s+u})\,du \,|\, \mathscr{A}_s \right). \end{aligned} \tag{A.20}$$

The interchange of conditional expectation and integral in the last step is justified because of the assumption that the paths of $X_t$ are progressively measurable.

(A.18) now follows immediately. □

REMARK. Using the condition (A.13), it can be proved that the class $\mathscr{D}$ is large enough. Indeed, if (A.13) holds, then $\mathscr{D}$ is a "measure determining" class. This fact is useful in the uniqueness assertions in Section 6.

*d-dimensional diffusion processes.* An $\mathbf{R}^d$-valued Markov process $\{X_t\}$ (wrt the family $\mathscr{A}_t = \sigma(X_s: s \le t) \vee \{\Pi \text{ null sets}\}$ on a countably additive probability space $(\Omega, \mathscr{A}, \Pi)$ having a generator $\mathscr{L}$ with domain $\mathscr{D}$ is called a diffusion process if

(i) $\Pi\{\omega: t \to X_t(\omega) \text{ is continuous}\} = 1$ and

(ii) $C_0^{1,2}([0,\infty) \times \mathbf{R}^d) \subseteq \mathscr{D}$ with

$$(\mathscr{L}f)(t,x) = \frac{\partial f}{\partial t}(t,x) + \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(t,x)\frac{\partial^2 f}{\partial x_i \partial x_j}(t,x) + \sum_{i=1}^{d} b_i(t,x)\frac{\partial f}{\partial x_i}(t,x) \tag{A.21}$$

for all $f \in C_0^{1,2}([0,\infty) \times \mathbf{R}^d)$, where $a_{ij}, b_i$ are measurable functions and the matrix $((a_{ij}(t,x))$ is symmetric nonnegative definite for all $(t,x)$.

The matrix-valued function $a$ is called the diffusion coefficient and $b$ is called the drift coefficient of the process $\{X_t\}$.

The question of the existence of a process $\{X_t\}$ with given diffusion and drift coefficients $a, b$ and when $a, b$ characterize the law of $\{X_t\}$ has been treated exhaustively by Stroock–Varadhan (1979).

The following moment estimate was obtained in Kallianpur (1980) under stronger conditions.

THEOREM A.2. *Let $\{X_t\}$ be an $\mathbf{R}^d$-valued diffusion on $(\Omega, \mathscr{A}, \Pi)$ with diffusion coefficient a and drift coefficient b satisfying, for each $T > 0$,*

$$|a_{ij}(t,x)| \le C_T \tag{A.22}$$

*and*

$$|b_i(t,x)| \le C_T(1 + |x|^2) \tag{A.23}$$

*for all $0 \in t \le T, x \in \mathbf{R}^d, i, j \ge 1$ and a suitable constant $C_T < \infty$. Also suppose for some $C_1 > 0$,*

$$E_\Pi \exp(C_1 |X_0|^2) < \infty. \tag{A.24}$$

*Then, for each $T > 0$, there exist constants $C_2 > 0, C_3 < \infty$ [depending on the constants $C_T, C_1$ appearing in conditions (A.22), (A.23), (A.24)] such that*

$$E_\Pi \exp\Big(C_2 \sup_{0 \le t \le T} |X_t|^2\Big) \le C_3. \tag{A.25}$$

OUTLINE OF PROOF. When the process $\{X_t\}$ is given as a solution to an Itô stochastic differential equation, the existence of $C_2, C_3$ satisfying (A.25) is proved in Kallianpur (1980, Theorem 5.7.2). First, an application of Lemma A.1 gives that

$$V^i(t) = X_t^i - \int_0^t b_i(s, x_s)\, ds$$

is a martingale and its quadratic variation is given by

$$\int_0^t a_{ii}(s, x_s)\, ds.$$

Using Burkholder's inequality on moments of $V$, we can complete the proof proceeding as in the above reference. A close examination of the various constants appearing shows that $C_2, C_3$ depend only on $C_T, C_1$. □

We will now derive the Feynman–Kac formula which gives a useful representation of a solution to a PDE. This formula is used in the literature under various sets of conditions [see, e.g., Friedman (1976)]. It was proved in Kallianpur–Karandikar (1984a) under the first set of conditions given in the next result. An outline of the proof under the second set of conditions was given in Kallianpur–Karandikar (1983b).

THEOREM A.3. *Suppose $\{X_t\}$ is a diffusion process with drift coefficient $b$ and diffusion coefficient $a$ and suppose that (A.24) holds. Let $c: [0,\infty) \times \mathbf{R}^d \to \mathbf{R}$ be a continuous function. Suppose that $v \in C^{1,2}([0,T] \times \mathbf{R}^d)$ is a (classical) solution to the PDE*

$$\text{(A.26)} \qquad \frac{\partial v}{\partial t} + \frac{1}{2}\sum_{i,j=1}^d a_{ij}\frac{\partial^2 v}{\partial x_i \partial x_j} + \sum_{i=1}^d b_i \frac{\partial v}{\partial x_i} + cv = 0.$$

*Suppose that either (A.27) holds or that the conditions (A.22), (A.23), (A.28), (A.29) hold where*

$$\text{(A.27)} \qquad c \text{ is bounded above}, \ v \text{ is bounded},$$

$$\text{(A.28)} \qquad |v(t,x)| \le \exp(K_1(1+|x|)), \qquad (t,x) \in [0,T] \times \mathbf{R}^d$$

*for a suitable constant $K_1 < \infty$,*

$$\text{(A.29)} \qquad c(t,x) \le K_2(1+|x|), \qquad (t,x) \in [0,T] \times \mathbf{R}^d$$

*for some constant $K_2 < \infty$.*

*Then, for all $0 \le s \le T$,*

$$\text{(A.30)} \qquad v(s, X_s) = E_\Pi\left(v(T, X_T)\exp\left(\int_s^T c(t, X_t)\, dt\right) \Big| \sigma(X_s)\right).$$

PROOF. Fix $0 \le s < T$. Using "integration by parts formula" for local martingales and the fact that $v$ satisfies (A.26), it can be shown that

$$N(t) = v(t, X_t)\exp\left(\int_s^t c(u, X_u)\, du\right), \qquad s \le t \le T$$

is a local martingale. For the details of this step, see Kallianpur–Karandikar (1984a). To complete the proof, we will show that $N(t)$ is a martingale which clearly implies (A.30).

If (A.27) holds, then $N$ is a bounded local martingale and hence is a martingale. If (A.22), (A.23), (A.28), (A.29) hold then

$$E_{\Pi}\Big(\sup_{s\le t\le T}|N(t)|\Big) \le E_{\Pi}\exp\Big((K_1+K_2)\Big(\sup_{s\le t\le T}|X(t)+1|\Big)\Big) < \infty$$

by Theorem A.2. Hence $N(t)$ is a martingale. This completes the proof. □

## REFERENCES


ALLINGER, F. and MITTER, S. K. (1981). New results on the innovations problem for nonlinear filtering. *Stochastics* 4 339–348.

BALAKRISHNAN, A. V. (1976). Applied functional analysis. *Applications of Mathematics*, Vol. 3. Springer-Verlag, New York.

BALAKRISHNAN, A. V. (1977). Radon–Nikodym derivatives of a class of weak distributions on Hilbert spaces. *Appl. Math. Optim.* 3 209–225.

BALAKRISHNAN, A. V. (1982). A likelihood ratio formula for random fields with application to physical geodesy. *Appl. Math. Optim.* 8 97–102.

BALAKRISHNAN, A. V. (1980). Non-linear white noise theory. *Multivariate Analysis*, 5. (P. R. Krishnaiah, ed.) North-Holland, Amsterdam.

BARAS, J. S., BLANKENSHIP, G. L. and MITTER, S. K. (1981). Nonlinear filtering of diffusion processes. *Proc. IFAC Congr.*, Kyoto, Japan.

BARAS, J. S., BLANKENSHIP, G. L. and HOPKINS, W. E. (1983). Existence, uniqueness and asymptotic behavior of solutions to a class of Zakai equations with unbounded coefficients. *IEEE Trans. Automat. Control* 28 203–214.

BESALA, P. (1979). Fundamental solution and Cauchy problem for a parabolic system with unbounded coefficients. *J. Differential Equations* 33 26–38.

BODANKO, W. (1966). Sur le probleme de Cauchy et les problèmes de Fourier pour les equations paraboliques dans un domaine non borné. *Ann. Polon. Math.* 18 79–94.

CLARK, J. M. C. (1969). Conditions for one to one correspondence between an observation process and its innovation. Technical Report, Center for Computing and Automation, Imperial College, London.

CLARK, J. M. C. (1978). The design of robust approximations to the stochastic differential equations of nonlinear filtering. In *Communications Systems and Random Process Theory*. (J. K. Skwirzynski, ed.) NATO Advanced Study Institute Series. Sijthoff and Noordhoff, Alphen aan den Rijn.

DAVIS, M. H. A. (1979). Pathwise solutions and multiplicative functionals in nonlinear filtering. *18th IEEE Conference on Decision and Control*, Fort Lauderdale, Florida.

DAVIS, M. H. A. (1980). On a multiplicative functional transformation arising in nonlinear filtering theory. *Z. Wahrsch. verw. Gebiete* 54 125–139.

DUNFORD, and SCHWARTZ (1958). *Linear Operators*. Vol. 1. Interscience, New York.

DYNKIN, D. B. (1964). *Markov Processes*. Vol. 1. Springer-Verlag, New York.

ELLIOTT, R. J. (1982). Stochastic calculus and applications. *Applications of Mathematics*. Vol. 18. Springer-Verlag, Berlin.

FRIEDMAN, A. (1964). *Partial Differential Equations of Parabolic Type*. Prentice-Hall, New York.

FRIEDMAN, A. (1975). *Stochastic Differential Equations and Applications*. Vol. 1. Academic, New York.

FUJISAKI, M., KALLIANPUR, G. and KUNITA, H. (1972). Stochastic differential equations for the nonlinear filtering problem. *Osaka J. Math.* 1 19–40.

GOHBERG, I. C., and KREIN, M. G. (1970). Theory and applications of Volterra operators in Hilbert space. *American Mathematical Society, Translations of Mathematical Monographs*. Vol. 24.

GROSS, L. (1960). Integration and nonlinear transformations in Hilbert space. *Trans. Amer. Math. Soc.* 94 404–440.

GROSS, L. (1962). Measurable functions on Hilbert space. *Trans. Amer. Math. Soc.* **105** 372–390.

IKEDA, N. and WATANABE, S. (1981). *Stochastic Differential Equations and Diffusion Processes.* North-Holland, Amsterdam.

KAILATH, T. (1968). An innovations approach to least squares estimation. *IEEE Trans. Automat. Control* **AC-13** 646–655.

KALLIANPUR, G. (1980). *Stochastic Filtering Theory.* Springer-Verlag, New York.

KALLIANPUR, G. and KARANDIKAR, R. L. (1983a). A finitely additive white noise approach to nonlinear filtering. *Appl. Math. Optim.* **10** 159–185.

KALLIANPUR, G. and KARANDIKAR, R. L. (1983b). Some recent developments in nonlinear filtering theory. *Acta Appl. Math.* **1** 399–434.

KALLIANPUR, G. and KARANDIKAR, R. L. (1984a). The nonlinear filtering problem for the unbounded case. *Stochastic Process Appl.* **18**, 57–66.

KALLIANPUR, G. and KARANDIKAR, R. L. (1984b). Measure valued equations for the optimum filter in finitely additive nonlinear filtering theory. *Z. Wahrsch. verw. Gebiete* **66** 1–17.

KALLIANPUR, G. and KARANDIKAR, R. L. (1984c). Markov property of the filter in the finitely additive white noise approach to nonlinear filtering. *Stochastics* **13** 177–198.

KALLIANPUR, G., KANNAN, D. and KARANDIKAR, R. L. (1984). Analytic and sequential Feynman integrals on Abstract Wiener and Hilbert spaces and a Cameron-Martin formula. Technical Report 53, Center for Stochastic Processes, Department of Statistics, University of North Carolina at Chapel Hill.

KALLIANPUR, G. and STRIEBEL, C. (1968). Estimation of stochastic processes: arbitrary system process with additive white noise observation errors. *Ann. Math. Statist.* **39** 785–801.

KALLIANPUR, G. and STRIEBEL, C. (1969). Stochastic differential equations occurring in the estimation of continuous parameter stochastic processes. *Theory of Probability and its Applications*, 4, Vol. 14, 567–594.

KOREZLIOGLU, H. and MARTIAS, C. (1984). Martingale representation and nonlinear filtering equation for distribution-valued processes. *Filtering and Control of Random Processes, Lecture-Notes in Control and Information Sciences.* Vol. 61. Springer-Verlag, Berlin.

KRYLOV, N. V. and ROZOVSKII, B. L. (1978). On the conditional distribution of diffusion processes. *Math. USSR-Izv.* **12** 336–356.

KRYLOV, N. V. and ROZOVSKII, B. L. (1981). Stochastic evolution equations. *J. Soviet Math.* **16** 1233–1276.

KUNITA, H. (1971). Asymptotic behavior of the nonlinear filtering errors of Markov processes. *J. Multivariate Analysis* **1** 365–393.

KUNITA, H. (1983). Stochastic partial differential equations connected with nonlinear filtering. *Nonlinear Filtering and Stochastic Control, Lecture Notes in Math.* 972. (S. K. Mitter and A. Moro, ed.). Springer-Verlag, New York.

KUSHNER, H. (1967). Dynamical equations for optimal nonlinear filtering. *J. Differential Equations* **3** 179–190.

LIPTSER, R. S. and SHIRYAEV, A. N. (1977). *Statistics of Random Processes.* Vol. 1. Springer-Verlag, New York.

MEYER, P. A. (1976). Un cours sur les integrales stochastiques. *Seminaire de Probabilities X, Lecture Notes in Math.* 511. Springer-Verlag, New York.

MORTENSON, R. E. (1966). Optimal control of continuous-time stochastic systems. Report ERL-66-1, Electronics Research Laboratory, College of Engineering, University of California, Berkeley.

OCONE, D. (1984). Remarks on the finite energy condition in additive white noise filtering. *Systems and Control Letters* **5** 197–203.

PARDOUX, E. (1979). Stochastic partial differential equations and filtering of diffusion processes. *Stochastics* **3**, 127–168.

PARDOUX, E. (1982). Equations du filtrage non lineaire de la prediction et du lissage. *Stochastics* **6** 193–231.

PARDOUX, E. and TALAY, D. (1984). Discretization and simulation of stochastic differential equations, Publication de Mathematiques appliquees, Marseille-Toulon (*Acta Appl. Math.*).

ROZOVSKII, B. L. (1975). On stochastic partial differential equations. *Math. USSR Sbornik* **25** 295–322 (English translation).

SEGAL, I. E. (1956). Tensor algebras over Hilbert spaces. *Trans. Amer. Math. Soc.* **81** 106–134.

SEO, J. H. and MAZUMDAR, R. R. (1984). On the innovations problem of a finitely additive white noise approach to nonlinear filtering. Department of System Science, University of California, Los Angeles.

STROOCK, D. W. and VARADHAN, S. R. S. (1979). *Multidimensional Diffusion Processes*. Springer-Verlag, New York.

SZPIROLAS, J. (1978). Sur l'équivalence d'équations differentielles stochastiques a valeurs mesures intervenant dans le filtrage markovien non linéaire. *Ann. Inst. Henri Poincaré, Sec. B* **XIV** 33–59.

ZAKAI, M. (1969). On the optimal filtering of diffusion processes. *Z. Wahrsch. verw. Gebiete* **11** 230–243.

CENTER FOR STOCHASTIC PROCESSES
DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27514