The Annals of Probability
1988, Vol. 16, No. 1, 80–107

UNIQUE CHARACTERIZATION OF CONDITIONAL DISTRIBUTIONS IN NONLINEAR FILTERING

BY T. G. KURTZ[1] AND D. L. OCONE[2]

University of Wisconsin-Madison and Rutgers University

Let (X, Y) solve the martingale problem for a given generator A. This paper studies the problem of uniquely characterizing the conditional distribution of $X(t)$ given observations $\{Y(s)|0 \leq s \leq t\}$. We define a *filtered martingale problem for* A and we show, given appropriate hypotheses on A, that the conditional distribution is the unique solution to the filtered martingale problem for A. Using these results, we then prove that the solutions to the Kushner–Stratonovich and Zakai equations for filtering Markov processes in additive white noise are unique under fairly general circumstances.

1. Introduction. A filtering model, in its most general form, is a pair of stochastic processes $(X, Y) = \{(X(t), Y(t))|t \geq 0\}$, in which X represents a signal which cannot be observed directly and Y an observation process related to X. Thus, at time t, the σ-algebra $\mathscr{F}_t^Y := \sigma\{Y(s)|0 \leq s \leq t\}$ provides all the available information about $X(t)$, and the conditional distribution of $X(t)$ given $\mathscr{F}_t^Y$ is the most detailed description of our knowledge of $X(t)$. A central goal of filtering theory is to characterize this conditional distribution effectively.

In this paper, we study how to characterize the conditional distribution uniquely for filtering models in which (X, Y) solves a martingale problem in the sense of Stroock and Varadhan (1979). This includes the commonly studied cases of partially observed Markov processes and, more particularly, of Markov processes observed in additive white noise. We approach the issue by defining a "filtered martingale problem," a martingale-type problem involving a probability–measure-valued process and the generator for the martingale problem for (X, Y). This definition is motivated by the simple fact that (π_t, Y), where π_t is the conditional distribution of $X(t)$ given $\mathscr{F}_t^Y$, solves this filtered martingale problem. This observation is by itself not essentially new; rather, it is a useful formalization of a standard technique, namely martingale representation, used to obtain filtering equations. Our main contribution—summarized in Theorems 3.2 and 3.3—is to give fairly mild conditions on the original martingale problem solved by (X, Y) that insure that the conditional distribution is, in fact, the *unique* solution to the filtered martingale problem. On the strength of these results, we are then able to obtain new uniqueness theorems for the Kushner–Stratonovich and Zakai equations for the conditional distribution of a

Received September 1985; revised December 1986.
[1]Partially supported by NSF grant DMS-8401360.
[2]Partially supported by NSF grant MCS-8301880.
AMS 1980 *subject classifications*. Primary 60G35, 62M20, 93E11; secondary 60H15, 60G44, 60G57.
Key words and phrases. Martingale problem, nonlinear filtering, Kushner–Stratonovich equation, Zakai equation, conditional distributions.

Markov process X, observed via $Y(t) = \int_0^t h(X(s))\,ds + W(t)$, where W is an independent Brownian motion. Previously, Szpirglas (1978) established uniqueness of solutions to the Kushner–Stratonovich and Zakai equations for bounded h. Also, Pardoux (1982) and Baras, Blankenship and Hopkins (1983) treated Zakai's equation for special cases of unbounded h when X is a diffusion process. By using techniques from p.d.e. theory, they established existence and uniqueness of solutions in various specific function classes after imposing technical conditions relating the growth of h to that of the diffusion coefficients. Baras, Blankenship and Hopkins (1983) did not prove their solutions were the actual unnormalized conditional densities, but this fact is a corollary of the results in this paper. Finally, Kallianpur and Karandikar (1984) obtained more general results for unbounded h by combining their finitely additive approach to nonlinear filtering with p.d.e. theory. They showed that if X is a nondegenerate diffusion and h is locally Hölder continuous, Zakai's equation has a unique solution in an exponential growth class and this solution is the unnormalized conditional density.

In this paper, we study the additive white noise model for signals X solving a martingale problem for a generator A_0. The hypotheses that we need to prove uniqueness are mild and do not necessarily require that h be bounded or continuous or that X be a diffusion. Basically, in the independent signal-observation noise case, the martingale problem for A_0 must be well posed, $E\int_0^T |h(x(s))|^2\,ds < \infty$, and if h is continuous, $D(A_0)$, the domain of A_0, must be sufficiently "dense" (see Theorems 4.1 and 4.2) or if h is not continuous, $R(\lambda - A_0)$, the range of $\lambda - A_0$, must be sufficiently "dense" for $\lambda > 0$ (see Theorems 4.6 and 4.7). With these assumptions, we obtain uniqueness for both the Kushner–Stratonovich and Zakai equations in the class of measure-valued solutions. Finally, we also prove uniqueness for diffusions observed in correlated noise (see Theorem 4.5). Some of the results of this paper were previously announced in Kurtz and Ocone (1985).

The paper is organized as follows. Section 2 presents the martingale problem theory on which the work is based. Section 3 defines the filtered martingale problem and establishes the main abstract results concerning uniqueness of the filtered martingale problem. Section 4 contains the applications to the Kushner–Stratonovich and Zakai equations for Markov processes observed in additive white noise.

In the remainder of this section, we define the notation to be used throughout the paper and state a lemma on the existence and regularity of the conditional distribution process. To begin, $(\Omega, \mathscr{F}, P)$ will denote a complete probability space on which all random variables will be defined. E, E_1 and E_2 will denote complete separable metric spaces, $B(E)$ the bounded Borel functions on E, $\mathscr{P}(E)$ the set of probability measures on the Borel sets of E and $D_E[0,\infty)$ the space of cadlag (right continuous with left limit) E-valued paths. As usual, $\mathscr{P}(E)$ carries the topology of weak convergence and $D_E[0,\infty)$ the Skorohod topology. [Note that $\mathscr{P}(E)$ is metrizable.] Given $\mu \in \mathscr{P}(E)$ and a Borel measurable μ-integrable f on E, we write $\mu f := \int_E f(x)\mu(dx)$.

Let (X, Y) be an $E_1 \times E_2$-valued measurable process defined on $(\Omega, \mathscr{F}, P)$ and set

$$\mathscr{F}_t^X = \sigma\{X(s)|0 \le s \le t\} \vee \mathscr{N},$$
$$\mathscr{F}_t^Y = \sigma\{Y(s)|0 \le s \le t\} \vee \mathscr{N},$$

where $\mathscr{N}$ is the collection of P-null sets of $(\Omega, \mathscr{F})$. Given a process Z, $E\{Z(t)|\mathscr{F}_{t+}^Y\}$ will always mean the $\{\mathscr{F}_{t+}^Y\}$-optional projection of Z.

Lemma 1.1. (a) *There is a $\mathscr{P}(E_1)$-valued $\{\mathscr{F}_{t+}^Y\}$-optional process $\pi_t(\omega, dx)$ such that*

$$\pi_t f = E\left[f(X(t))|\mathscr{F}_{t+}^Y\right] \tag{1.1}$$

for all $f \in B(E_1)$ and equality in (1.1) *means that the two processes are indistinguishable.*

(b) *If the sample paths of X are in $D_{E_1}[0, \infty)$, then π_t has a version with paths in $D_{\mathscr{P}(E_1)}[0, \infty)$ and for all but countably many t,*

$$\pi_t(f) = E\left[f(X(t))|\mathscr{F}_t^Y\right] \quad a.s. \tag{1.2}$$

for all $f \in B(E_1)$.

Remarks. (i) For technical reasons, namely, right continuity of $\{\mathscr{F}_{t+}^Y\}$, we choose to work with $E[f(X(t))|\mathscr{F}_{t+}^Y]$ rather than $E[f(X(t))|\mathscr{F}_t^Y]$.

(ii) For the processes that we shall work with later, one may always assume that X is cadlag and, hence, that π_t is also. We always work with this cadlag version.

(iii) In the applications of Section 4 to Markov processes observed in white noise, $\mathscr{F}_{t+}^Y = \mathscr{F}_t^Y$, so $\pi_t(f) = E[f(X(t))|\mathscr{F}_t^Y]$ for all t.

Proof of Lemma 1.1. Part (a) and the conclusion that π_t has a cadlag version if X is cadlag are proved by Yor (1977). To prove (1.2), observe that if Z is a process with sample paths in $D_S[0, \infty)$, where S is some metric space, then

$$I_0 = \{t > 0|P(Z_t \ne Z_{t-}) > 0\}$$

is countable. Now $\mathscr{P}(E_1)$ with the weak topology is metrizable. Thus, if $\pi(\omega, \cdot) \in D_{\mathscr{P}(E_1)}([0, \infty))$ for all ω,

$$I = \{t > 0|P(\pi_t \ne \pi_{t-}) > 0\}$$

is countable. Choose some $t \notin I$. Then

$$\pi_t = \pi_{t-} = \lim_{s \uparrow t} \pi_s$$

a.s. and, hence, since $\mathscr{F}_t^Y$ includes the P-null sets by definition, π_t is $\mathscr{F}_t^Y$-measurable. This concludes the proof. □

We shall also have need for the following notational convention. Let $f(x, t, \omega)$ be a bounded function on $E_1 \times [0, \infty) \times \Omega$, measurable with respect to the product σ-algebra. Then

$$\pi_{t,\omega} f(\cdot, t, \omega) := \int_{E_1} f(x, t, \omega)\pi_t(\omega, dx).$$

It is easily seen that $\pi_{t,\omega} f(\cdot, t, \omega)$ is a measurable process; certainly this is true for indicators $1_A(x)1_B(t)1_C(\omega)$ and their linear combinations, and it follows for bounded measurable f by a Dynkin class-type argument. Further, by a similar argument, if $f(x, t, \omega)$ is $E_1 \times \mathscr{F}_t^Y$-measurable for every t,

$$\pi_{t,\omega} f(\cdot, t, \omega) = E\left[f(X(t), t, \omega) | \mathscr{F}_{t+}^Y \right](\omega).$$

For $f(x, Y(t, \omega))$, where f is in $B(E_1 \times E_2)$—the case most commonly encountered—we write

$$\pi_{t,\omega} f(\cdot, Y(t, \omega)) = \pi_t f(\cdot, Y(t))$$

and we have

$$\pi_t f(\cdot, Y(t)) = E\left[f(X(t), Y(t)) | \mathscr{F}_{t+}^Y \right].$$

2. Martingale problems. Let $A \subset B(E) \times B(E)$. We think of A as a (perhaps multivalued) operator. The domain of A is defined by $\mathscr{D}(A) = \{f: (f, g) \in A\}$ and the range of A is defined by $\mathscr{R}(A) = \{g: (f, g) \in A\}$. Multivalued operators will arise only in certain technical parts of the proofs and, since in most situations A will be single-valued, we will write Af instead of g for $(f, g) \in A$.

A measurable stochastic process Z is a solution of the martingale problem for A if there exists a filtration $\{\mathscr{F}_t\}$ such that

$$f(Z(t)) - \int_0^t Af(Z(s))\, ds \tag{2.1}$$

is an $\{\mathscr{F}_t\}$-martingale for each $f \in \mathscr{D}(A)$. If A is multivalued, this requirement should be interpreted to mean that

$$f(Z(t)) - \int_0^t g(Z(s))\, ds \tag{2.2}$$

is a martingale for each g such that $(f, g) \in A$. For $\nu \in \mathscr{P}(E)$, Z is a solution of the martingale problem for (A, ν) if Z is a solution of the martingale problem for A and $Z(0)$ has distribution ν. We say that uniqueness holds for the martingale problem for (A, ν) if any two solutions have the same finite-dimensional distributions. The martingale problem for A is *well posed* if for each $\nu \in \mathscr{P}(E)$, a solution of the martingale problem for (A, ν) exists and uniqueness holds. In almost all situations of interest, a solution of the martingale problem for A has a modification with sample paths in $D_E[0, \infty)$. [See Ethier and Kurtz (1986) for a detailed discussion.] Consequently, we will assume that all solutions considered here have that property, that is, are solutions of the $D_E[0, \infty)$ *martingale problem.* We remark that uniqueness may hold for the $D_E[0, \infty)$ martingale problem while failing to hold for the more general martingale problem, but this is rare.

Let ν_t denote the distribution of $Z(t)$. Then the fact that (2.1) is a martingale implies

$$\nu_t f = \nu_0 f + \int_0^t \nu_s Af\,ds, \qquad f \in \mathscr{D}(A). \tag{2.3}$$

We will need conditions under which, given ν_0, (2.3) determines ν_t for $t \geq 0$. We give conditions of two types. The first condition is essentially that the closure of A generates a semigroup. We say that $M \subset B(E)$ is separating [for $\mathscr{P}(E)$] if $\nu, \mu \in \mathscr{P}(E)$ and $\nu f = \mu f$ for all $f \in M$ implies $\nu = \mu$.

PROPOSITION 2.1. *Suppose $\mathscr{R}(\lambda - A)$ is separating for each $\lambda > 0$. If $\{\nu_t\}$ and $\{\mu_t\}$ satisfy* (2.3), *are weakly right continuous and $\nu_0 = \mu_0$, then $\nu_t = \mu_t$ for all $t \geq 0$.*

PROOF. By (2.3),

$$\begin{aligned}\lambda\int_0^\infty e^{-\lambda t}\nu_t f\,dt &= \nu_0 f + \lambda\int_0^\infty e^{-\lambda t}\int_0^t \nu_s Af\,ds\,dt\\ &= \nu_0 f + \lambda\int_0^\infty\int_s^\infty e^{-\lambda t}\,dt\,\nu_s Af\,ds\\ &= \nu_0 f + \int_0^\infty e^{-\lambda s}\nu_s Af\,ds.\end{aligned} \tag{2.4}$$

Consequently,

$$\int_0^\infty e^{-\lambda t}\nu_t(\lambda f - Af)\,dt = \nu_0 f, \qquad f \in \mathscr{D}(A). \tag{2.5}$$

Since $\mathscr{R}(\lambda - A)$ is separating, (2.5) implies that ν_0 uniquely determines the measure $\int_0^\infty e^{-\lambda t}\nu_t\,dt$. Since this holds for each $\lambda > 0$, if $\{\nu_t\}$ is weakly right continuous, the uniqueness of the Laplace transform implies ν_0 determines ν_t, $t \geq 0$. □

The second set of conditions is taken from Echeverria (1982). He treated the particular case $\nu_0 Af = 0$, for which uniqueness gives $\nu_t = \nu_0$, $t \geq 0$, i.e., ν_0 is a stationary distribution. [In the following statement and throughout the paper if E is locally compact, $\hat{C}(E)$ denotes the continuous functions which vanish at infinity.]

PROPOSITION 2.2. *Let E be locally compact and separable. Suppose that $A \subset \hat{C}(E) \times \hat{C}(E)$, that $\mathscr{D}(A)$ is an algebra and dense in $\hat{C}(E)$ and that the $D_E[0,\infty)$ martingale problem for A is well posed. If $\{\nu_t\}$ satisfies* (2.3) *and Z is a solution of the martingale problem for (A, ν_0), then for all $t \geq 0$, ν_t is the distribution of $Z(t)$.*

PROOF. See Ethier and Kurtz [(1986), Chapter 4, Proposition 9.19].

3. The filtered martingale problem. Let $E = E_1 \times E_2$ and $A \subset B(E) \times B(E)$. If $Z = (X, Y)$ is a solution of the $D_E[0,\infty)$ martingale problem for A and

π_t is the conditional distribution of $X(t)$ given $\mathscr{F}_{t+}^Y$, then

$$\pi_t f(\cdot, Y(t)) - \int_0^t \pi_s A f(\cdot, Y(s))\, ds \tag{3.1}$$

is an $\{\mathscr{F}_{t+}^Y\}$-martingale for each $f \in \mathscr{D}(A)$. We think of (π, Y) as a process with sample paths in $D_{\mathscr{P}(E_1)\times E_2}[0,\infty)$ and we want to use this martingale property to characterize the distribution of (π, Y) and, hence, to characterize π as a functional of Y. Simply requiring (3.1) to be a martingale with respect to an arbitrary filtration is clearly not sufficient to characterize π. For example, (δ_X, Y) would satisfy this requirement. Consequently, we introduce the notion of a filtered martingale problem. A process (μ, U) with sample paths in $D_{\mathscr{P}(E_1)\times E_2}[0,\infty)$ is a solution of the filtered martingale problem for A if μ is $\{\mathscr{F}_{t+}^U\}$-adapted and

$$\mu_t f(\cdot, U(t)) - \int_0^t \mu_s A f(\cdot, U(s))\, ds \tag{3.2}$$

is an $\{\mathscr{F}_{t+}^U\}$-martingale for each $f \in \mathscr{D}(A)$.

Note that the assumption that (μ, U) is $\{\mathscr{F}_{t+}^U\}$-adapted implies that for each t there exists a Borel measurable function H_t: $D_{E_2}[0,\infty) \to \mathscr{P}(E_1)$ such that

$$\mu_t = H_t(U) = H_t(U(\cdot \wedge s)) \quad \text{a.s.} \tag{3.3}$$

for every $s > t$. Since μ has sample paths in $D_{\mathscr{P}(E_1)}[0,\infty)$, for all but countably many t, $\mu_t = \mu_{t-}$ a.s. and, hence, μ_t is $\mathscr{F}_t^U$-measurable. For such t, it follows that there exists a Borel measurable G_t: $D_{E_2}[0,\infty) \to \mathscr{P}(E_1)$ such that

$$G_t(U) = G_t(U(\cdot \wedge t)) = H_t(U) \quad \text{a.s.} \tag{3.4}$$

To connect (μ, U) with the original process, we require

$$E[\mu_0 f(\cdot, U(0))] = E[f(X(0), Y(0))] \tag{3.5}$$

for all $f \in B(E)$. Then $U(0)$ and $Y(0)$ have the same distribution. Furthermore, if uniqueness holds for solutions of the filtered martingale problem satisfying (3.5), then (π, Y) must have the same distribution as (μ, U) and, hence, $\pi_t = H_t(Y)$ a.s. Consequently, the filtered martingale problem gives a method of characterizing nonlinear filters.

Let (X, Y) be a solution of the $D_E[0,\infty)$ martingale problem for A and let (μ, U) be a solution of the filtered martingale problem for A satisfying (3.5). Let $\nu_t \in \mathscr{P}(E)$ satisfy $\nu_t f = E[\mu_t f(\cdot, U(t))]$. Then $\{\nu_t\}$ satisfies (2.3) and under the conditions of either Proposition 2.1 or 2.2, we have

$$E[\mu_t f(\cdot, U(t))] = E[f(X(t), Y(t))], \qquad t \ge 0,\ f \in \mathscr{B}(E). \tag{3.6}$$

It follows from (3.6) that for each $t > 0$, $U(t)$ and $Y(t)$ have the same distribution. This, of course, is not enough for our purposes, but it does indicate how we will proceed. Let $V_0(t) = t$ and for $i = 1, \dots, m$, let $g_i \in \hat{C}(E \times [0,\infty))$ be nonnegative and define

$$V_i(t) = \int_0^t g_i(U(s), s)\, ds = \int_0^t g_i(U(s), V_0(s))\, ds.$$

Define W_i in the same way as V_i, but with U replaced by Y. Observe that

$$\mathscr{F}_t^Y = \sigma\left\{Y(t), \int_0^t g(Y(s), s)\, ds : g \in \hat{C}(E \times [0,\infty)),\ g \ge 0\right\} \vee \mathscr{N}. \tag{3.7}$$

If for each m and each choice of $g_0, \ldots, g_m$, we can show that

$$E\left[\mu_t f(\cdot, U(t), V_0(t), \ldots, V_m(t))\right] = E\left[f(X(t), Y(t), W_0(t), \ldots, W_m(t))\right] \tag{3.8}$$

for each $t \geq 0$ and each f in a collection of functions that is separating in $B(E_1 \times E_2 \times [0, \infty)^{m+1})$, then $(U, V_0, \ldots, V_m)$ and $(Y, W_0, \ldots, W_m)$ have the same one-dimensional distributions and, hence, using (3.4), for all but countably many t,

$$\begin{aligned} &E\left[f(X(t), Y(t), W_0(t), \ldots, W_m(t))\right] \\ &\quad = E\left[G_t(Y) f(\cdot, Y(t), W_0(t), \ldots, W_m(t))\right] \\ &\quad = E\left[G_t(Y(\cdot \wedge t)) f(\cdot, Y(t), W_0(t), \ldots, W_m(t))\right]. \end{aligned} \tag{3.9}$$

Since $G_t(Y(\cdot \wedge t))$ is $\mathscr{F}_t^Y$-measurable, it follows that $G_t(Y) = \pi_t$ a.s. for t for which $\pi_t = \pi_{t-}$ a.s., that is, for all but countably many t. Consequently, by (3.4), for all but countably many t (i.e., for all t such that both $\pi_t = \pi_{t-}$ and $\mu_t = \mu_{t-}$),

$$\pi_t = H_t(Y) \quad \text{a.s.} \tag{3.10}$$

The right continuity of π and μ implies (3.10) holds for all t. The significance of the preceding analysis is that if, given the observed process Y, an algorithm (e.g., solving the Kushner–Stratonovich equation) produces a functional H_t such that $(H_t(Y), Y)$ is a solution of the filtered martingale problem for A with $E[H_0(Y) f(\cdot, Y(0))] = E[f(X(0), Y(0))]$, then uniqueness of the filtered martingale problem implies $\pi_t = H_t(Y)$. Turning now to the problem of proving uniqueness, note that (3.8) is of the same form as (3.6) with U replaced by $(U, V_0, \ldots, V_m)$. Consequently we can use Propositions 2.1 and 2.2 to give conditions for (3.8) to hold and, hence, conditions for uniqueness.

Let $\hat{C}^1([0, \infty))$ denote the collection of functions such that $f, f' \in \hat{C}([0, \infty))$. Let $g_0 = 1$ and let $g_1, \ldots, g_m$ be as in (3.8). For $f \in \mathscr{D}(A)$ and $f_i \in \hat{C}^1([0, \infty))$, $i = 0, \ldots, m$ (recall, we are assuming $g_i \geq 0$), define

$$f(x, y, w_0, \ldots, w_m) = f(x, y) \prod_{i=0}^{m} f_i(w_i) \tag{3.11}$$

and

$$\begin{aligned} &Bf(x, y, w_0, \ldots, w_m) \\ &\quad = Af(x, y) \prod_{i=1}^{m} f_i(w_i) + f(x, y) \sum_{i=0}^{m} g_i(y, w_0) f_i'(w_i) \prod_{j \neq i} f_j(w_j). \end{aligned} \tag{3.12}$$

Let $E_0 = E \times [0, \infty)^{m+1}$. Extend B linearly and note that if E is locally compact and separable and $\mathscr{D}(A)$ is an algebra that is dense in $C(E)$, then $\mathscr{D}(B)$ is an algebra that is dense in $C(E_0)$. Furthermore, we claim that if the $D_E[0, \infty)$ martingale problem for A is well posed, then the $D_{E_0}[0, \infty)$ martingale problem for B is well posed and, consequently, if A satisfies the conditions of Proposition 2.2, then B does also. Existence of solutions for B follows from existence for A by the preceding definition of Z_i. To see that uniqueness for A implies uniqueness for B, note that if $(X, Y, W_0, \dots, W_m)$ is a solution of the martingale problem for B, then (X, Y) is a solution for A. Uniqueness for B is then a consequence of Lemma 3.1. The proof is left to the reader.

LEMMA 3.1. *Let $\{\mathscr{F}_t\}$ be a filtration and let U and V be right-continuous processes adapted to $\{\mathscr{F}_t\}$. If*

$$f(U(t)) - \int_0^t V(s) f'(U(s))\, ds \tag{3.13}$$

is an $\{\mathscr{F}_t\}$-martingale for each $f \in \hat{C}'(\mathbb{R})$, then

$$U(t) = U(0) + \int_0^\infty V(s)\, ds. \tag{3.14}$$

We now give the main uniqueness theorems.

THEOREM 3.2. *Let E_1 and E_2 be complete separable metric spaces and let $E = E_1 \times E_2$. Let $A \subset B(E) \times B(E)$. Suppose the $D_E[0, \infty)$ martingale problem for A is well posed and $\mathscr{R}(\lambda - A)$ is bounded-pointwise dense in $B(E)$ for each $\lambda > 0$ (i.e., $B(E)$ is the smallest collection of bounded Borel measurable functions containing $\mathscr{R}(\lambda - A)$ that is closed under bounded-pointwise convergence of sequences). Let B be defined as in* (3.11) *and* (3.12). *Then* (a) *B satisfies the conditions of Proposition* 2.1; (b) *if (X, Y) is a solution of the $D_E[0, \infty)$ martingale problem for A and (μ, U) is a solution of the filtered martingale problem for A with sample paths in $D_{\mathscr{P}(E_1) \times E_2}[0, \infty)$ and satisfying* (3.5), *then there exist functions H_t satisfying* (3.3) *and* (3.10) *and (μ, U) has the same distribution as (π, Y) (i.e., uniqueness holds for the filtered martingale problem for A).*

REMARK. The range condition implies uniqueness holds for the $D_E[0, \infty)$ martingale problem. [See Ethier and Kurtz (1986), Chapter 4, Section 4.] Note also that if (X, Y) is a solution of the martingale problem for A, then it is a solution for the linear span of A, so without loss of generality we may assume A is linear. For linear A, existence of solutions of the martingale problem for (A, ν) for all $\nu \in \mathscr{P}(E)$ implies A is dissipative, that is, $\|\lambda f - Af\| \geq \lambda \|f\|$ for all $\lambda > 0$, and $(\lambda - A)^{-1}$ exists and satisfies $\|(\lambda - A)^{-1}\| \leq \lambda^{-1}$. ($\|\cdot\|$ denotes the sup norm.)

PROOF OF THEOREM 3.2. Let $\hat{B}$ denote the bounded-pointwise closure of the linear span of B. We must show that $\mathscr{R}(\lambda - B)$ is separating. This will be true if and only if $\mathscr{R}(\lambda - \hat{B})$ is separating.

As noted earlier, without loss of generality, we can assume A is linear. Let Z_z denote a solution of the martingale problem for (A, δ_z). Then for $h \in \mathscr{R}(\lambda - A)$, there is an $(f, g) \in A$ such that

$$f(z) = (\lambda - A)^{-1} h(z) = E\left[\int_0^\infty e^{-\lambda s} h(Z_z(s))\, ds\right]. \tag{3.15}$$

Let $\hat{A}$ denote the bounded-pointwise closure of A in $B(E) \times B(E)$. Then any solution of the martingale problem for A is a solution for $\hat{A}$ and, similarly, for the filtered martingale problem. Since $\mathscr{R}(\lambda - A)$ is bounded-pointwise dense in $B(E)$, it follows from (3.15) that $\mathscr{R}(\lambda - \hat{A}) = B(E)$ [bounded-pointwise convergence of a sequence $\{h_n\} \subset \mathscr{R}(\lambda - A)$ to a function h implies convergence of the corresponding sequence $\{f_n\}$ to a function f and $(f, \lambda f - h) \in \hat{A}$]. We may, therefore, assume that $A = \hat{A}$ and, hence, that $\mathscr{R}(\lambda - A) = B(E)$.

Let A_0 be the bounded-pointwise closure of the linear span of

$$\{(f(x,y)f_0(w_0), f_0(w_0)Af(x,y) + f(x,y)f_0'(w_0)) \colon f \in \mathscr{D}(A), \quad f_0 \in \hat{C}^1([0,\infty))\}. \tag{3.16}$$

We claim that $\mathscr{R}(\lambda - A_0) = B(E \times [0,\infty))$. To see this, let $a \geq 0$ and $f_0(w_0) = e^{aw_0}$. Then for $h \in B(E)$, there is an $f \in \mathscr{D}(A)$ such that

$$\begin{aligned} \lambda f(x,y)f_0(w_0) - f_0(w_0)Af(s,y) - f(x,y)f_0'(w_0) \\ = e^{-aw_0}((\lambda + a)f(x,y) - Af(x,y)) \\ = e^{-aw_0}h(x,y). \end{aligned} \tag{3.17}$$

Linear combinations of functions of the form $e^{-aw_0}h(x,y)$ form an algebra and $\beta(E \times [0,\infty))$ [the Borel subsets of $E \times [0,\infty)$] is the smallest σ-algebra with respect to which all such functions are measurable. It follows that this collection of functions is bounded-pointwise dense in $B(E \times [0,\infty))$ [see Ethier and Kurtz (1986), Appendix, Corollary 4.4].

If $f_0 \in \mathscr{D}(A_0)$ and $a_1, \dots, a_m \geq 0$, then

$$f(x,y,w_0,\dots,w_m) = f_0(x,y,w_0)\prod_{i=1}^m e^{-a_i w_i} \in \mathscr{D}(\hat{B}) \tag{3.18}$$

and

$$\begin{aligned} \lambda f(x,y,w_0,\dots,w_m) - \hat{B}f(x,y,w_0,\dots,w_m) \\ = \prod_{i=1}^m e^{-a_i w_i}\Bigg(\lambda f_0(x,y,w_0) - Af(x,y,w_0) \\ + f(x,y,w_0)\sum_{i=1}^m a_i g_i(y,w_0)\Bigg). \end{aligned} \tag{3.19}$$

Let $V(y,w_0) = \Sigma a_i g_i(y,w_0)$. We claim that for $h \in B(E \times [0,\infty))$, there exists $f_0 \in \mathscr{D}(A_0)$ such that

$$\lambda f_0 - A_0 f_0 + V f_0 = h. \tag{3.20}$$

To see this, let $c = \|V\|$ and note that since $V \geq 0$, $c \geq \|V - c\|$. Define $Kf_0 = ((\lambda + c) - A_0)^{-1}(V - c)f_0$. Then K is a bounded linear operator on $B(E \times [0,\infty))$ with $\|K\| \leq c/(\lambda + c) < 1$ and f_0 satisfies (3.20) if and only if

$$f_0 + Kf_0 = ((\lambda + c) - A_0)^{-1}h. \tag{3.21}$$

Existence of a solution of (3.21) follows from the fact that $\|K\| < 1$. It now follows that $\mathscr{R}(\lambda - \hat{B}) = B(E \times [0,\infty)^{m+1})$.

If (X, Y) is a solution of the $D_E[0, \infty)$ martingale problem for A, (μ, U) is a solution of the filtered martingale problem for A satisfying (3.5) and $W_0, \dots, W_m$ and $V_0, \dots, V_m$ are as previously defined, then $(\lambda, U, V_0, \dots, V_m)$ is a solution of the filtered martingale problem for $\hat{B}$ satisfying

$$E\left[\mu_0 f(\cdot, U(0), V_0(0), \dots, V_m(0))\right] = E\left[f(X(0), Y(0), W_0(0), \dots, W_m(0))\right].$$

Proposition 2.1 implies (3.8) by the same argument that gave (3.6) and part (b) follows. □

THEOREM 3.3. *Let E_1 and E_2 be locally compact separable metric spaces and let $E = E_1 \times E_2$. Let $A \subset \hat{C}(E) \times \hat{C}(E)$ and suppose A satisfies the conditions of Proposition* 2.2. *Let B be defined as in* (3.11) *and* (3.12). *Then* (a) *B satisfies the conditions of Proposition* 2.2; (b) *if (X, Y) is a solution of the $D_E[0, \infty)$ martingale problem for A and (μ, U) is a solution of the filtered martingale problem for A with sample paths in $D_{\mathscr{P}(E_1)\times E_2}[0, \infty)$ and satisfying* (3.5), *then there exist functions H_t satisfying* (3.3) *and* (3.10) *and (μ, U) has the same distribution as (π, Y) (i.e., uniqueness holds for the filtered martingale problem for A).*

PROOF. We have already noted that if A satisfies the conditions of Proposition 2.2, then B does also. Part (b) follows as in the proof of Theorem 3.2. □

COROLLARY 3.4. *Under the conditions of either Theorem* 3.2 *or* 3.3, *suppose $\mathscr{D}(A) \subset \overline{C}(E)$ and let (X, Y) be a solution of the $D_E[0, \infty)$ martingale problem for A and (μ, U) be a process with sample paths in $D_{\mathscr{P}(E_1)\times E_2}[0, \infty)$ satisfying* (3.5). *Suppose that μ is $\{\mathscr{F}_{t+}^U\}$-adapted, that τ is an $\{\mathscr{F}_{t+}^U\}$-stopping time and that for each $f \in \mathscr{D}(A)$*

$$\mu_{t\wedge\tau} f(\cdot, U(t \wedge \tau)) - \int_0^{t\wedge\tau} \mu_s Af(\cdot, U(s))\, ds \tag{3.22}$$

is an $\{\mathscr{F}_{t+}^U\}$-martingale. Then for each $t \geq 0$,

$$\mu_t \chi_{\{t<\tau\}} = H_t(U)\chi_{\{t<\tau\}} \quad a.s., \tag{3.23}$$

where H_t is as in (3.10).

PROOF. A process $(\tilde{\mu}, \tilde{U})$ can be constructed on an enlarged sample space such that $(\tilde{\mu}_{t\wedge\tau}, \tilde{U}(t \wedge \tau)) = (\mu_{t\wedge\tau}, U(t \wedge \tau))$, $t \geq 0$, and $(\tilde{\mu}, \tilde{U})$ is a solution of the filtered martingale problem for A [see Ethier and Kurtz (1986), Chapter 4, Lemma 5.16]. Then (3.23) follows by Theorem 3.2 or 3.3. □

4. Uniqueness for the equations of nonlinear filtering. In this section, we apply the filtered martingale problem characterization of π_t to derive uniqueness theorems for the Kushner–Stratonovich and Zakai equations for the following filtering problem. The signal X is taken to be an E-valued cadlag solution of the martingale problem for (A_0, π_0); here, E is a locally compact complete separable metric space, $A_0 \subset B(E) \times B(E)$ and $\pi_0 \in \mathscr{P}(E)$. Let h: $E \to \mathbb{R}^P$ be a Borel function. The additive white noise observation model is then

$$Y(t) = \int_0^t h(X(s))\,ds + W(t), \tag{4.1}$$

where

(4.2) W is a Brownian motion independent of X.

We will be interested in studying the filtering problem on the time interval $[0, T]$ for some given $T > 0$ and we assume, in addition, that

$$E\int_0^T |h(X(s))|^2\,ds < \infty. \tag{4.3}$$

[All random processes are defined on $(\Omega, \mathscr{F}, P)$ as before.] Later, we also study an example in which X and W are correlated.

This filtering model is stated in a time independent form—neither h nor A_0 depends explicitly on t. However, there is no loss of generality because the time dependent case can always be recast into the autonomous form by the standard trick of including time as a component of X. Hence, all results that follow may be interpreted for the time dependent case.

The assumptions (4.2) and (4.3) imply that the formula

$$\frac{dP_0}{dP} = \exp\left[-\int_0^T h(X(s))\,dW(s) - \frac{1}{2}\int_0^T |h(X(s))|^2\,ds\right]$$

defines a new probability measure P_0 which is mutually absolutely continuous with respect to P and with respect to which Y is a Brownian motion independent of X. From this observation, it follows easily that $\mathscr{F}_t^Y = \mathscr{F}_{t+}^Y$ for all $t \le T$ and, hence, that $\pi_t f = E[f(X(t))|\mathscr{F}_t^Y]$ for all $t \le T$. We shall use this right continuity of $\{\mathscr{F}_t^Y\}$ without further comment.

Finally, it will be convenient to define the *innovations process*

$$I(t) = Y(t) - \int_0^t \pi_s h\,ds.$$

It is a standard result from filtering theory [see, e.g., Liptser and Shiryayev (1977)] that I is an $\{\mathscr{F}_t^Y\}$-Wiener process on $(\Omega, \mathscr{F}, P)$. [Here and in what follows, $\pi_s h$ is the vector $(\pi_s h_1, \dots, \pi_s h_p)$.]

By explicitly representing the martingale

$$\pi_t f - \int_0^t \pi_s(A_0 f)\,ds$$

as a stochastic integral with respect to I, one obtains the Kushner–Stratonovich equation

$$\pi_t f = \pi_0 f + \int_0^t \pi_s(A_0 f)\,ds + \int_0^t [\pi_s(hf) - \pi_s(h)\pi_s(f)]\,dI(s) \quad \text{for all } f \in \mathscr{D}(A_0). \tag{4.4}$$

Note that the condition "for all $f \in \mathscr{D}(A_0)$" is included as part of (4.4). This is important: (4.4) is essentially a stochastic differential equation for π_t written in weak form, that is, for π_t applied to a domain of test functions f, and it is necessary to specify this domain precisely. The point of applying martingale problem techniques is to find a set of test functions small enough to handle easily, but large enough to insure uniqueness.

Let $\mathscr{M}^+(E)$ denote the set of positive bounded Borel measures on E. Define the $\mathscr{M}^+(E)$-valued process

$$\sigma_t := \exp\left[\int_0^t \pi_s h\, dY(s) - \tfrac{1}{2}\int_0^t |\pi_s h|^2\, ds\right]\pi_t.$$

The assumption (4.3) guarantees that

$$E\int_0^T |\pi_s h|^2\, ds < \infty$$

and, hence, that σ_t is a well defined almost surely nonzero process for $t \le T$. It can be shown that $\sigma_t f$ is the numerator in the Kallianpur–Striebel formula for $\pi_t f$, that is,

$$\sigma_t f = E_0\left[f(X(t)) \frac{dP}{dP_0} \middle| \mathscr{F}_t^Y \right]$$

and $\pi_t f = \sigma_t f / \sigma_t 1$, where E_0 denotes expectation with respect to P_0. By Itô's rule, σ_t satisfies the Duncan–Mortensen–Zakai equation

$$\sigma_t f = \pi_0 f + \int_0^t \sigma_s(A_0 f)\, ds + \int_0^t \sigma_s(hf)\, dY(s) \tag{4.5}$$

for every f in $\mathscr{D}(A_0)$.

Equation (4.5) has the advantage of being linear in σ and of being driven by Y rather than the innovations process. Equations (4.4) and (4.5) extend to more general situations than the filtering model defined here. Standard references giving derivations and extensions are Fujisaki, Kallianpur and Kunita (1972) and Liptser and Shiryayev (1977).

Application of the filtered martingale problem characterization of π_t in Theorem 3.3 leads to the following uniqueness theorems for (4.4) and (4.5).

THEOREM 4.1. *Assume in addition to* (4.2) *that*

(a) A_0 *satisfies the conditions of Proposition* 2.2, *i.e.*, $A_0 \subset \hat{C}(E) \times \hat{C}(E)$, $\mathscr{D}(A_0)$ *is a dense algebra of* $\hat{C}(E)$ *and the martingale problem for* A_0 *is well posed*;

(b) $f(x)h_i(x) \in \hat{C}(E)$ *for all* $f \in \mathscr{D}(A_0)$, $1 \le i \le p$.

Let $\{\mu_t\}$ *be an* $\{\mathscr{F}_t^Y\}$*-adapted cadlag* $\mathscr{P}(E)$*-valued process such that*

$$\mu_t f = \pi_0 f + \int_0^t \mu_s(A_0 f)\, ds + \int_0^t [\mu_s(hf) - \mu_s(h)\mu_s(f)]\, dI^\mu(s) \tag{4.6}$$

for every $f \in \mathscr{D}(A_0)$, $t \le T$,

where $I^\mu(t) = Y(t) - \int_0^t \mu_s h\, ds$. *Then* $\mu_t = \pi_t$ *for all* $t < T$ *a.s.*

THEOREM 4.2. *Make the same assumptions as in Theorem* 4.1. *Let* $\{\rho_t\}$ *be an* $\{\mathscr{F}_t^Y\}$*-adapted cadlag* $\mathscr{M}^+(E)$*-valued process satisfying*

$$\rho_t f = \pi_0 f + \int_0^t \rho_s(A_0 f)\, ds + \int_0^t \rho_s(hf)\, dY(s) \qquad (4.7)$$

$$\text{for all } f \in \mathscr{D}(A_0),\ t \le T,$$

and

$$\rho_t 1 = 1 + \int_0^t \rho_s(h)\, dY(s). \qquad (4.8)$$

Then $\rho_t = \sigma_t$ *for all* $t < T$ *a.s.*

REMARK 1. Since h may be unbounded, $\mu_s h$ does not automatically make sense. Therefore, we regard as implicit in the assumption that μ_t satisfy (4.6) the conditions that h be a.s. μ_t-integrable for all $t \le T$ and

$$\int_0 |\mu_t h|^2\, dt < \infty \quad \text{a.s.}$$

so that all stochastic integrals are well defined. Similar remarks hold for $\rho_t h$.

REMARK 2. (4.7) and (4.8) may be rephrased as saying that (4.7) holds for all $f \in \mathscr{D}(A_0 \cup \{(1,0)\})$.

REMARK 3. The assumption (4.3), that $E\int_0^T |h(X(s))|^2\, ds < \infty$ is standard in derivations of the Kushner–Stratonovich and Zakai equations. A careful examination of the proofs to follow show that we really only use

$$E\int_0^T |h(X(s))|\, ds < \infty,$$

which guarantees existence of $\pi_s(h)$ and is needed in Lemma 4.4, and

$$\int_0^T |\pi_s(h)|^2\, ds < \infty \quad \text{a.s.},$$

which is needed to make the stopping time arguments and to assure the existence and positivity of $\sigma_t 1$. Thus, our results prove that if solutions to the filtering equations exist under these slightly less restrictive assumptions, they must coincide with the actual conditional distributions.

REMARK 4. In Theorem 4.1, hypotheses (a) and (b) are implied by the alternative hypotheses:

(a′) $A_0 \subset C_c(E) \times \hat{C}(E)$, where $C_c(E)$ denotes the set of continuous functions with compact support, $\mathscr{D}(A_0)$ is a dense subalgebra of $\hat{C}(E)$ and the martingale problem for A_0 is well posed.

(b′) $h_i(x) \in C(E)$ for $1 \le i \le p$.

These conditions have the advantage that (b′) involves only $h_1, \dots, h_p$ and not $\mathscr{D}(A_0)$. We shall use (a′) and (b′) in the first example (which follows the proof of Theorem 4.2).

REMARK 5. Since $\{\sigma_t\}$ is the solution of a stochastic p.d.e., one can view $\{\sigma_t\}$ as a diffusion in $M^+(E)$ and can try to formulate a corresponding martingale problem. That is, one can seek an operator $\mathscr{A} \subset B(\mathscr{M}^+(E)) \times B(M^+(E))$ such that the martingale problem for $\mathscr{A}$ is well posed and $\{\sigma_t\}$ is a solution for this martingale problem. This has been done by Hijab (1986). As the domain of $\mathscr{A}$, he takes those functions on $\mathscr{M}^+(E)$ of the form

$$F(\mu) = \varphi(\mu f_1, \dots, \mu f_n),$$

where n is arbitrary, $\varphi \in C^\infty(\mathbb{R}^n)$ and $f_1, \dots, f_n \in \mathscr{D}(A_0)$. The appropriate definition of $\mathscr{A}$ on $\mathscr{D}(\mathscr{A})$ is found by using Itô's rule on $F(\sigma_t)$ for $\mathscr{F} \in \mathscr{D}(\mathscr{A})$. This leads to

$$\begin{aligned}\mathscr{A}F(\rho) = &\sum_{i=1}^{n} \partial_i \varphi(\rho f_1, \dots, \rho f_n) \rho A_0 f_i \\ &+ \sum_{i,j=1}^{n} \partial_{ij}^2 \varphi(\rho f_1, \dots, \rho f_n)\langle h, f_i\rangle\langle h, f_j\rangle.\end{aligned}$$

Hijab shows that the martingale problem for $\mathscr{A}$ is well posed if $h_i \in \mathscr{D}(A_0)$, $1 \le i \le p$.

The strategy of the proof of Theorem 4.1 is to use the stochastic integral representation of $\mu_t f - \int_0^t \mu_s(A_0 f)\,ds$, given by (4.6), to verify that μ_t provides a solution to the filtered martingale problem. However, for this we need an operator A which specifies a martingale problem for the joint process (X, Y), while the filtering equations (4.6) and (4.7) involve only A_0, the generator of X. This problem is easily amended by observing that if ψ is a function in $C_c^\infty(\mathbb{R}^p)$, the infinitely differentiable functions with compact support, and if $(f, g) \in A_0$, Itô's rule implies

$$\begin{aligned}f(X(t))\psi(Y(t)) - \int_0^t \big[&g(X(s))\psi(Y(s)) \\ &+ f(X(s))\big[\tfrac{1}{2}\Delta\psi(Y(s)) + h(X(s))\nabla\psi(Y(s))\big]\big]\,ds\end{aligned}$$

is an $\{\mathscr{F}_t^X \vee \mathscr{F}_t^Y\}$-martingale. It follows that (X, Y) solves the martingale problem for $(A, \pi_0 \times \delta_0)$, where

$$\begin{aligned}A = \operatorname{span}\big\{\big(f(x)\psi(y), g(x)\psi(y) + f(x)\big[\tfrac{1}{2}\Delta\psi(y) + h(x)\nabla\psi(y)\big]\big) \big| \\ (f, g) \in A_0, \psi \in C_c^\infty(\mathbb{R}^p)\big\}\end{aligned}$$

(span here means linear span). We need to show that A satisfies the hypotheses of Theorem 3.3 if A_0 satisfies the assumptions of Theorem 4.1. The following "obvious," but not so easy to prove, fact will be needed [for a proof see Ethier and Kurtz (1986), Chapter 4, Theorem 10.1].

LEMMA 4.3. *Let $A_i \subset B(E_i) \times B(E_i)$, $i = 1, 2$, where E_1 and E_2 are complete separable metric spaces and where $(1, 0) \in A_i$ for $i = 1$ and $i = 2$. Define*

$$B \subset B(E_1 \times E_2) \times B(E_1 \times E_2),$$

$$B = \{(f_1 f_2, g_1 f_2 + g_2 f_1) | (f_1, g_1) \in A_1, (f_2, g_2) \in A_2\}.$$

Then if uniqueness holds for the martingale problems for A_1 and A_2, it holds for the martingale problem for B, also.

The operator B corresponds to the process (X_1, X_2), where X_i solves the martingale problem for A_i, $i = 1,2$, and X_1 and X_2 are independent, given independent initial measures.

LEMMA 4.4. *Assume that A_0 and h satisfy the conditions of Theorem 4.1. Then A satisfies the conditions of Proposition 2.2. In particular, the martingale problem for A is well posed.*

PROOF. Assumption (b) of Theorem 4.1, $f(x)h_i(x) \in \hat{C}(E)$, $1 \le i \le p$, implies that $A \subset \hat{C}(E \times \mathbb{R}^p) \times \hat{C}(E \times \mathbb{R}^p)$. $\mathscr{D}(A)$ is clearly an algebra and, by applying the Stone–Weierstrass theorem to the one-point compactification of $E \times \mathbb{R}^p$, $\mathscr{D}(A)$ is dense in $\hat{C}(E \times \mathbb{R}^p)$. One may also easily check that A satisfies the positive maximum principle. It remains to show that the $D_{E\times\mathbb{R}^p}[0,\infty)$ martingale problem for A is well posed. Existence of solutions follows by explicit construction: If B is a Brownian motion independent of X and Y_0 and (X_0, Y_0) has the distribution ν, then

$$\left(X(t), Y(t) = Y_0 + \int_0^t h(X(s))\,ds + B(t)\right)_{t\ge 0}$$

solves the martingale problem for (A, ν). To show uniqueness, let

$$A_1 = \left\{\left(\psi, \tfrac{1}{2}\Delta\psi\right) | \psi \in C_c^\infty(\mathbb{R}^p) + \text{constants}\right\}.$$

It is well known [Stroock and Varadhan (1979)] that the martingale problem for A_1 is well posed and its solution is merely $\mathbb{R}^p$-valued Brownian motion. Since the martingale problem for A_0 is well posed, that for $A_0 = A_0 \cup \{(1,0)\}$ is also and, therefore, by Lemma 4.3, uniqueness holds for the martingale problem for

$$B = \left\{\left(f(x)\psi(y), g(x)\psi(y) + f(x)\tfrac{1}{2}\Delta\psi(g)\right) | (f,g) \in A_0,\ \psi \in C_c^\infty(\mathbb{R}^p)\right\} \cup \{(1,0)\}.$$

We shall show that if (X, Y) solves the martingale problem for A, then (X, Y), where

$$Z(t) = Y(t) - \int_0^t h(X(s))\,ds,$$

solves the martingale problem for $(B, \pi_0 \times \delta_0)$. This will complete the proof since then Z will be a Brownian motion independent of X, which is a uniquely determined solution for the martingale problem for A_0.

Set $\mathscr{G}_t = \mathscr{F}_t^X \vee \mathscr{F}_t^Y$. Our task is to show that if $t_1 < t_2$,

$$E\left[f(X(t_2))\psi(Z(t_2)) | \mathscr{G}_{t_1}\right] = f(X(t_1))\psi(Z(t_1)) + E\left[\int_{t_1}^{t_2} B[f(X(s))\psi(Z(s))]\,ds \,|\, \mathscr{G}_{t_1}\right].$$

The following identities will be crucial. First, for any $\eta \in \mathbb{R}^p$ and $s < t$,

$$\begin{aligned} &E[f(X(t))\psi(Y(t)+\eta)|\mathscr{G}_s] \\ (4.9)\qquad &= f(X(s))\psi(Y(s)+\eta) + E\left[\int_s^t Af(X(u))\psi(Y(u)+\eta)\,du|\mathscr{G}_s\right], \end{aligned}$$

since (X, Y) solves the martingale problem for A. The terms $f(x)\psi(y+\eta)$ and $Af(x)\psi(y+\eta)$ are continuous in η and bounded and, hence, one can conclude that for any $\mathscr{G}_s$-measurable random variable (with values in $\mathbb{R}^p$),

$$\begin{aligned} &E[f(X(t))\psi(Y(t)+H)|\mathscr{G}_s] \\ (4.10)\qquad &= f(X(s))\psi(Y(s)+H) + E\left[\int_s^t Af(X(u))\psi(Y(u)+H)\,du|\mathscr{G}_s\right]. \end{aligned}$$

Now observe that

$$\begin{aligned} &E[f(X(t_2))\psi(Z(t_2))|\mathscr{G}_{t_1}] \\ &= E\left[f(X(t_2))\psi\left(Y(t_2) - \int_0^{t_2} h(X(u))\,du\right)\Big|\mathscr{G}_{t_1}\right] \\ (4.11)\qquad &= E\left[f(X(t_2))\psi(Y(t_2)) - \int_0^{t_2} f(X(t_2))\right. \\ &\qquad \left.\times \nabla\psi\left(Y(t_2) - \int_0^s h(X(u))\,du\right)h(X(s))\,ds|\mathscr{G}_{t_1}\right]. \end{aligned}$$

The next step is to apply (4.8) with H replaced by $\int_0^s h(X(u))\,du$ to $f(X(t_2))\nabla\psi(Y(t_2) - \int_0^s h(X(u))\,du)$ in the second term of the right-hand side of (4.11). Here we shall write

$$Z(t, s) = Y(t) - \int_0^s h(X(u))\,ds;$$

note that $Z(s, s) = Z(s)$. The result is

$$\begin{aligned} &E\left[\int_0^{t_2} f(X(t_2))\nabla\psi(Z(t_2, s))h(X(s))\,ds|\mathscr{G}_{t_1}\right] \\ &= \int_0^{t_1} f(X(t_1))\nabla\psi(Z(t_1, s))h(X(s))\,ds \\ (4.12)\qquad &\quad + E\left[\int_{t_1}^{t_2} f(X(s))\nabla\psi(Z(s))h(X(s))\,ds|\mathscr{G}_{t_1}\right] \\ &\quad + E\left[\int_0^{t_2}\int_{t_1\vee s}^{t_2} Af(X(u))\nabla\psi(Z(u, s))h(X(s))\,du\,ds|\mathscr{G}_{t_1}\right]. \end{aligned}$$

Now

$$\begin{aligned} &\int_0^t f(X(t_1))\nabla\psi(Z(t_1, s))h(X(s))\,ds \\ (4.13)\qquad &= f(X(t_1))[\psi(Y(t_1)) - \psi(Z(t_1))] \end{aligned}$$

and

$$\begin{aligned}(4.14)\qquad &\int_0^{t_2}\int_{t_1\vee s}^{t_2} Af(X(u))\nabla\psi(Z(u,s))h(X(s))\,du\,ds\\ &\quad=\int_{t_1}^{t_2}[Af(X(u))\psi(Y(u)) - Af(X(u))\psi(Z(u))]\,du.\end{aligned}$$

Substitution of (4.13) and (4.14) into (4.12) and then (4.11) yields

$$\begin{aligned}&E[f(X(t_2))\psi(Z(t_2))|\mathscr{G}_{t_1}]\\ &\quad= f(X(t_1))\psi(Z(t_1)) + E\biggl[\int_{t_1}^{t_2} - f(X(s))\nabla\psi(Z(s))h(X(s))\\ &\qquad\qquad + Af(X(s))\psi(Z(s))\,ds|\mathscr{G}_{t_1}\biggr]\\ &\qquad + E[f(X(t_2))\psi(Y(t_2))|\mathscr{G}_{t_1}] - f(X(t_1))\psi(Y(t_1))\\ &\qquad - E\biggl[\int_{t_1}^{t_2}Af(X(s))\psi(Y(s))\,ds|\mathscr{G}_{t_1}\biggr]\\ &\quad= f(X(t_1))\psi(Z(t_1)) + E\biggl[\int_{t_1}^{t_2}Bf(X(s))\psi(Z(s))\,ds|\mathscr{G}_{t_1}\biggr].\end{aligned}$$

□

PROOF OF THEOREM 4.1. If $\psi \in C_c^\infty(\mathbb{R}^p)$ and $f \in \mathscr{D}(A_0)$, an application of Itô's rule shows that

$$\begin{aligned}(4.15)\qquad \mu_t[f(\cdot)\psi(Y(t))] &= (\pi_0 f)\psi(0) + \int_0^t \mu_s A[f(\cdot)\psi(Y(s))]\,ds\\ &\quad + \int_0^t[(\mu_s f)\nabla\psi(Y(s)) + \psi(Y(s))\\ &\qquad\times[\mu_s(hf) - \mu_s h\mu_s f]]\,dI^\mu(s).\end{aligned}$$

Let $\xi(s) = \mu_s(h) - \pi_s(h)$ and set

$$\tau_N = T\wedge \inf\left\{t\,\middle|\,\int_0^t|\xi(s)|^2\,ds \ge N\right\} \wedge \inf\left\{t\,\middle|\,\int_0^t|\mu_s h|^2\,ds \ge N\right\}.$$

Define a new measure Q_N on $(\Omega, \mathscr{F})$ by

$$\frac{dQ_N}{dP} = \exp\left[\int_0^{\tau_N}\xi(s)\,dI(s) - \frac{1}{2}\int_0^{\tau_N}|\xi(s)|^2\,ds\right],$$

where I is the innovations process. Girsanov's theorem implies that Q_N is a probability measure and that

$$I(t) - \int_0^{t\wedge\tau_N}\xi(s)\,ds$$

is an $\{\mathscr{F}_t^Y\}$-Brownian motion on $(\Omega, \mathscr{F}, Q_N)$. However, $I^\mu(t\wedge\tau_N) = I(t\wedge\tau_N) - \int_0^{t\wedge\tau_N}\xi(s)\,ds$ and, thus,

$$\int_0^{t\wedge\tau_N}[\mu_s f\,\nabla\psi(Y(s)) + \psi(Y(s))[\mu_s hf - \mu_s h\mu_s f]]\,dI^\mu(s)$$

is an $\{\mathscr{F}_t^Y\}$-martingale on $(\Omega, \mathscr{F}, Q_N)$, since for $f \in \mathscr{D}(A_0)$ the integrand is bounded on $[0, \tau_N]$. It follows from (4.15) that

$$\mu_{t\wedge\tau_N}\big(\tilde{f}(\cdot, Y(t\wedge\tau_N))\big) - \int_0^{t\wedge\tau_N}\mu_s A\tilde{f}(\cdot, Y(s\wedge\tau_N))\,ds \tag{4.16}$$

is an $\{\mathscr{F}_t^Y\}$-martingale for all $\tilde{f}\in\mathscr{D}(A)$ on $(\Omega, \mathscr{F}, Q_N)$. Since A satisfies the conditions of Proposition 2.2 (Lemma 4.3), Theorem 3.3 and Corollary 3.4 imply that

$$\begin{aligned}\mu_t\chi_{\{t<\tau_N\}} &= H_t(Y)\chi_{\{t<\tau_N\}}\\ &= \pi_t\chi_{\{t<\tau_N\}} \quad \text{a.s. } Q_N,\end{aligned} \tag{4.17}$$

where H_t is the $\mathscr{P}(E_1)$-valued functional defined in Section 3. Since $Q_N \ll P$, (4.17) holds a.s. with respect to P also. Finally, a.s. P, there exists $N(\omega)$ such that $n \geq N(\omega)$ implies $\tau_n(\omega) = T$. Taking $N\to\infty$ in (4.17), therefore, implies $\mu_t = \pi_t$ for $t < T$ a.s. P. □

Proof of Theorem 4.2. We shall work on the probability space $(\Omega, \mathscr{F}, P_0)$ on which $Y(\cdot)$ is a Brownian motion; since $P_0 \sim P$, all almost sure statements are valid with respect to either P_0 or P. First, observe that, because $d\sigma_t 1 = \sigma_t h\,dY$, $\sigma_t 1$ is a nonnegative continuous local $\{\mathscr{F}_t^Y\}$-martingale on $(\Omega, \mathscr{F}, P_0)$. Furthermore, $\sigma_T 1 > 0$ a.s. and it follows that $\min_{[0,T]}\sigma_t 1 > 0$ a.s., because a nonnegative continuous local martingale will stay at 0 after it hits 0.

Let ρ_t be a solution of the Zakai equation (4.7). Our strategy will be to show that $\mu_t = \rho_t/\rho_t 1$ satisfies the Kushner–Stratonovich equation, so that $\rho_t/\rho_t 1 = \pi_t$ and then to argue that $\rho_t 1 = \sigma_t 1$. However, in forming $\mu_t = \rho_t/\rho_t 1$, one must take care of the possibility that $\rho_t 1 = 0$. To this end, define for $\varepsilon > 0$,

$$\tau(\varepsilon) = \inf\{t | \rho_{t-}1 < \varepsilon\} \wedge T.$$

It suffices to prove that $\chi_{\{t<\tau_\varepsilon\}}\rho_t = \chi_{\{t<\tau_\varepsilon\}}\sigma_t$, $t < T$ a.s. Indeed, if this is the case,

$$\tau(\varepsilon) \geq \inf\{t|\sigma_t 1 < \varepsilon\} \wedge T$$

and since $\inf\{t|\sigma_t 1 < \varepsilon\} \wedge T = T$ for ε less than some $\varepsilon_0(\omega)$ almost surely, $\rho_t = \sigma_t$ for $t < T$ almost surely. Thus, to complete the proof, let μ_t be a cadlag $\{\mathscr{F}_t^Y\}$-adapted process such that $\mu_{t\wedge\tau(\varepsilon)} = \rho_{t\wedge(\varepsilon)}/\rho_{t\wedge\tau(\varepsilon)}1$. By Itô's rule,

$$\mu_{t\wedge\tau(\varepsilon)}f = \pi_0 f + \int_0^{t\wedge\tau(\varepsilon)}\mu_s(A_0 f)\,ds + \int_0^{t\wedge\tau(\varepsilon)}[\mu_s(hf) - \mu_s(h)\mu_s(f)]\,dI^\mu(s)$$

for all $f\in\mathscr{D}(A_0)$. By the proof of Theorem 4.1,

$$\chi_{\{t<\tau(\varepsilon)\}}\mu_t = \chi_{\{t<\tau(\varepsilon)\}}\pi_t \quad \text{a.s.}$$

Next consider the process $r_t = \rho_t 1/\sigma_t 1$—recall that $\inf_{[0,T]}\sigma_t 1 > 0$ a.s. Itô's rule implies

$$r_t = 1 + \int_0^t[\rho_s h/\sigma_s 1 - r_s\pi_s h]\,dI(s).$$

But for $s < \tau(\varepsilon)$, $\rho_s h/\sigma_s 1 = r_s(\rho_s h/\rho_s 1) = r_s \pi_s h$. Thus, $\chi_{\{t < \tau(\varepsilon)\}} r_t \equiv 1$ and

$$\begin{aligned}\chi_{\{t<\tau(\varepsilon)\}}\rho_t &= \left(\chi_{\{t<\tau(\varepsilon)\}}\rho_t 1\right)\pi_t \\ &= \left(\chi_{\{t<\tau(\varepsilon)\}}\sigma_t 1\right)\pi_t \\ &= \chi_{\{t<\tau(\varepsilon)\}}\sigma_t. \end{aligned}$$ □

EXAMPLE. Let $m(x)$: $\mathbb{R}^n \to \mathbb{R}^n$ and $g(x)$: $\mathbb{R}^n \to \mathbb{R}^n \otimes \mathbb{R}^d$ be globally Lipschitz and take

$$A_0 f = \tfrac{1}{2} tr\left[gg^T(x)\, D^2 f(x)\right] + m(x)\nabla f(x)$$

with $\mathscr{D}(A_0) = C_c^\infty(\mathbb{R}^n)$, where $C_c^\infty(\mathbb{R}^n)$ is the class of infinitely differentiable functions with compact support. According to Stroock and Varadhan (1979), the martingale problem for A_0 is well posed: Solutions are given by solutions of the stochastic differential equation

$$dX = m(X)\, dt + g(X)\, dB,$$

where B is a d-dimensional Brownian motion independent of the initial condition $x(0)$. If $h(x)$ is continuous and satisfies $E\int_0^T |h(x(s))|^2\, ds < \infty$, then the Kushner–Stratonovich and Zakai equations have unique solutions given by the normalized and unnormalized conditional distributions of the filtering problem associated to $Y(t) = \int_0^t h(X(s))\, ds + W(t)$, where W is independent of X.

The strategy used to obtain uniqueness in Theorems 4.1 and 4.2 extends also to cases involving correlation between signal and observation noise. Here, one must be more specific about the signal process than in Theorems 4.1 and 4.2 in order to handle the correlation effectively. We shall analyze the following example:

$$\begin{aligned} dX(t) &= m(X)\, dt + g(X)\, dB_1, \\ dY &= h(X)\, dt + G_1\, dB_1 + G_2\, dB_2, \\ X(0) &= x_0, \qquad Y(0) = 0. \end{aligned} \tag{4.18}$$

In (4.18), B_1 is a d-dimensional Brownian motion and B_2 is a p-dimensional Brownian motion independent of B_1. Both B_1 and B_2 are independent of the initial condition x_0, which has distribution π_0. We assume:

(i) $m(x)$: $\mathbb{R}^n \to \mathbb{R}^n$ and $g(x)$: $\mathbb{R}^n \to \mathbb{R}^{n\times d}$ are globally Lipschitz.

(ii) $h(x)$: $\mathbb{R}^n \to \mathbb{R}^d$ is continuous and

$$E\int_0^T |h(X(s))|^2\, ds < \infty. \tag{4.19}$$

(iii) G_1 and G_2 are constant $p \times d$ and $p \times p$ matrices, respectively, and $G_1 G_1^T + G_2 G_2^T = I$.

(iv) $G_2 G_2^T > 0$.

Assumption (iii) is made only to give the filtering equations a nice form. If G_1

and G_2 did not satisfy this condition, by (iv) one could choose a symmetric matrix $p \times p$ such that $R \cdot R = G_1 G^T + G_2 G_2^T$ and replace Y by

$$\tilde{Y}(t) = R^{-1}Y(t) = \int_0^t R^{-1}h(X(s))\,ds + R^{-1}G_1 B_1(t) + R^{-1}G_2 B_2(t).$$

$\tilde{G}_1 = R^{-1}G_1$ and $\tilde{G}_2 = R^{-1}G_2$ now satisfy assumption (iii). Hence, (iii) represents no loss of generality.

Given the assumptions in (4.19), X is a solution of the well posed martingale problem associated to the operator A_0 with domain $\mathscr{D}(A_0) = C_c^\infty(\mathbb{R}^n)$ given by

$$A_0 f = \tfrac{1}{2} tr\left[gg^T(x)D^2 f(x)\right] + \nabla f(x)m(x)$$

[see Stroock and Varadhan (1979)].

Equation (4.20) in the following theorem is the Kushner–Stratonovich equation for this filtering problem [Fujisaki, Kallianpur and Kunita (1972)].

THEOREM 4.5. *Let μ_t be a right-continuous $P(\mathbb{R}^n)$-valued $\mathscr{F}_t^Y$-adapted process such that*

$$\begin{aligned} \mu_t(f) &= \pi_0(f) + \int_0^t \mu_s(A_0 f)\,ds \\ &\quad + \int_0^t \left[\mu_s(hf) - \mu_s(h)\mu_s(f) + \mu_s(\nabla f g G_1^T)\right] dI^\mu(s) \end{aligned} \tag{4.20}$$

for all $f \in \mathscr{D}(A_0)$. Then, given assumptions (4.19) (i)–(iv), $\mu_t = \pi_t$ *for all $t < T$ a.s.*

PROOF. As in the proof of Theorem 4.1, we must first construct a generator A for (X, Y). For the domain of A, we take the linear span of the set

$$\{f(x)\psi(y) | f \in C_c^\infty(\mathbb{R}^n),\ \psi \in C_c^\infty(\mathbb{R}^p)\}.$$

Clearly, A must have the form

$$\begin{aligned} Af(x)\psi(y) &= \psi(y)A_0 f(x) + f(x)\left[\tfrac{1}{2}\nabla\psi(y) + h^T(x)\nabla\psi(y)\right] \\ &\quad + \nabla f(x)g(x)G_1^T\nabla^T\psi(y). \end{aligned}$$

Suppose that the martingale problem for A is well posed; it certainly satisfies the other "Echeverria conditions" of Theorem 3.3. By Itô's rule,

$$\begin{aligned} &\mu_t(f(\cdot)\psi(Y(t))) \\ &\quad = \pi_0(f)\psi(0) + \int_0^t \mu_s(Af\psi(\cdot, Y(s)))\,ds \\ &\qquad + \int_0^t \left[\mu_s(hf(\cdot) + \nabla f g G_1^T)\psi(Y(s)) + \mu_s(f)\nabla\psi(Y(s))\right. \\ &\qquad\qquad \left. - \psi(Y(s))\mu_s(h)\mu_s(f)\right] dI^\mu(s). \end{aligned}$$

Arguing exactly as in Theorem 4.1, we may then conclude that $\mu_t = \pi_t$ for $t < T$ a.s. Therefore, it remains to show that the martingale problem for A is well posed in order to complete the proof. To do this, let B be the operator

$$Bf(x)\psi(y) = \psi(y)A_0 f(x) + f(x)\left[\tfrac{1}{2}\nabla\psi(y)\right] + \nabla f(x)g(x)G_1^T\nabla^T\psi(y),$$

defined on

$$\mathscr{D}(B) = \operatorname{span}\{f(x)\psi(y) | f \in C_c^\infty(\mathbb{R}^n),\ \psi \in C_c^\infty(\mathbb{R}^p)\}.$$

By the results of Stroock and Varadhan (1979), the martingale problem for B is well posed and its solutions are given by solutions to the stochastic differential equations

$$dX = m(X)\,dt + g(X)\,dB_1,$$
$$dZ = G_1\,dB_1 + G_2\,dB_2.$$

By arguing exactly as in Lemma 4.4, one may then conclude that if (X, Y) solves the martingale problem for A, then (X, Y), where $Z(t) = Y(t) - \int_0^t h(X(s))\,ds$, solves the martingale problem for B. Since the martingale problems for A_0 and B are well posed, it follows that the martingale problem for A is well posed, also. □

The uniqueness theorems obtained so far effectively require the observation function $h(x)$ to be continuous because their proofs rely upon Theorem 3.3 which includes the condition $A \subset \hat{C}(E) \times \hat{C}(E)$. We can relax the continuity assumption if we replace the hypotheses on A_0 by an assumption on $\mathscr{R}(\lambda - A_0)$ and invoke Theorem 3.2. For the case of independent observation noise and signal, we obtain Theorem 4.6: As in Theorems 4.1 and 4.2, the signal X is a cadlag solution to the martingale problem for (A_0, π_0), $Y(t) = \int_0^t h(X(s))\,ds + W(t) \in \mathbb{R}^p$, where W is a Brownian motion independent of X and $E\int_0^T |h(X(s))|^2\,ds < \infty$. In addition, we shall use X_x to denote the solution to the martingale problem for (A_0, δ_x).

THEOREM 4.6. *Assume that*:

(i) *The $D_E([0, \infty))$ martingale problem for A_0 is well posed and $\mathscr{R}(\lambda - A_0)$ is bounded-pointwise dense in $B(E)$ for all $\lambda > 0$.*

(ii) $$E\int_0^T |h(X_x(s))|\,ds < \infty \text{ for each } x \text{ in } E.$$

Let $\{\mu_t\}$ be a $\{\mathscr{F}_t^Y\}$-adapted process with paths in $D_{\mathscr{P}(E)}[0, T)$ such that

$$\int_0^T \mu_s^2(|h|)\,ds < \infty \quad a.s. \tag{4.21}$$

and $\{\mu_t\}$ satisfies the Kushner–Stratonovich equation (4.6) *for all f in $\mathscr{D}(A_0)$. Then $\mu_t = \pi_t$ for all $t < T$ a.s.*

THEOREM 4.7. *Let the assumptions of Theorem* 4.6 *hold. Let $\{\rho_t\}$ be a $\{\mathscr{F}_t^Y\}$-adapted process with sample paths in $D_{\mathscr{M}^+(E)}[0, T)$, such that $\int_0^T \rho_s^2(|h|)\,ds < \infty$ and $\{\rho_t\}$ satisfies* (4.7) *and* (4.8).

Then $\rho_t = \sigma_t$, $t < T$ a.s.

REMARK. Condition (4.21) insures that the term $\int_0^T \mu_s(hf)\, dI^\mu(s)$ in the Kushner–Stratonovich equation is well defined. It is also necessary for limit arguments in the proof.

PROOFS. Theorem 4.7 is proved from Theorem 4.6 exactly in the manner of proving Theorem 4.2 from Theorem 4.1.

We shall carry out the proof of Theorem 4.6 for the case $T = \infty$, that is,

$$\int_0^\infty \mu_s^2(|h|)\, ds < \infty \quad \text{a.s.} \tag{4.22}$$

and

$$E\int_0^\infty |h(X_x(s))|\, ds < \infty \quad \text{for each } x \text{ in } E. \tag{4.23}$$

This incurs no loss of generality. If $T < \infty$, set $\tilde{X} = \{(X(t), t)\}$ and $\tilde{Y}(t) = \int_0^t \tilde{h}(\tilde{X}(s))\, ds + W(t)$, where $\tilde{h}(x, s) = h(x)1_{\{s \leq T\}}$. For this filtering problem, $\tilde{X}$ solves the martingale problem for $\tilde{A}_0$, where $\tilde{A}_0$ is the linear span of

$$\{(f(x)f_0(t), f_0(t)A_0 f(x) + f(x)f_0'(t) | f \in \mathscr{D}(A_0),\ f_0 \in \hat{C}_c^1[0, \infty))\}$$

and if μ_t satisfies the original Kushner–Stratonovich equation (4.6) involving A_0, $\tilde{\mu}_t = \mu_t \times \delta_t$ solves it for $\tilde{A}_0$:

$$\tilde{\mu}_t(F) = \pi_0(F(\cdot, 0)) + \int_0^t \tilde{\mu}_s(\tilde{A}_0 F)\, ds$$
$$+ \int_0^t [\tilde{\mu}_s(\tilde{h}F) - \tilde{\mu}_s(\tilde{h})\mu_s(F)]\, dI^\mu(s)$$

for $\mathscr{F} \in \mathscr{D}(A_0)$ and $t \leq T$. Clearly, (4.22) and (4.23) are satisfied for $\tilde{h}$; furthermore, from the proof of Theorem 3.2, $\tilde{A}_0$ is well posed and $\mathscr{R}(\lambda - \tilde{A}_0)$ is bounded-pointwise dense in $B(E \times \mathbb{R}_+)$. This rephrases the $T < \infty$ case as a $T = \infty$ case.

As before, the main problem is to construct a generator A for (X, Y) such that for each $\phi(x, y) \in \mathscr{D}(A)$,

$$\mu_t(\phi(\cdot, Y(t))) = \pi_0(\phi(\cdot, 0)) + \int_0^t \mu_s(A\phi(\cdot, Y(s)))\, ds + \int_0^t \eta_s\, dI^\mu(s) \tag{4.24}$$

for some process η_s. This time, however, we want A to satisfy the assumptions of Theorem 3.2—the $D_{E\times\mathbb{R}^p}[0, \infty)$ martingale problem for A must be well posed and $R(\lambda - A)$ must be bounded-pointwise dense in $B(E \times \mathbb{R}^p)$ for all $\lambda > 0$. The rest of the proof will be devoted to constructing A and proving these properties. Once this is done, the reader will see from the form of η_s in (4.24) that the proof used to complete Theorem 4.1, using Theorem 3.2 in place of Theorem 3.3, also completes the proof of Theorem 4.6, except for one technical detail. Corollary 3.4 used in the proof of Theorem 4.1 no longer applies because $\mathscr{D}(A)$ is not contained in $\overline{C}(E)$. However, this condition can be dropped if the random time τ in Corollary 3.4 is discrete; the proof is exactly the same using Lemma 5.16 of Chapter 4 of Ethier and Kurtz (1986). We shall fix the proof of Theorem 4.6 by showing that the stopping times τ_N, $N > 0$, defined in the proof of Theorem 4.1, can be replaced by discrete stopping times. First observe that we can construct a probability space $(\Omega, \mathscr{G}, Q)$ on which a process (μ, U) and stopping times τ_N are defined so that $(\mu_{\cdot \wedge \tau_N}, U(\cdot \wedge \tau_N))$ has the same distribution as it does under Q_N. (Q_N is also defined in the proof of Theorem 4.1.) One way to do this is to apply Tulcea's theorem to the sequence of conditional distributions for $(\mu_{\cdot \wedge \tau_{N+1}}, U(\cdot \wedge \tau_{N+1}))$ given $(\mu_{\cdot \wedge \tau_N}, U(\cdot \wedge \tau_N))$. Next select ε_N so that $Q(\tau_N + \varepsilon_N \geq \tau_{N+1}) < 2^{-N}$. By the Borel–Cantelli lemma $\tau_N + \varepsilon_N \geq \tau_{N+1}$ only for finitely many values of N a.s. On the event $\{\tau_N < k\varepsilon_N \leq \tau_N + \varepsilon_N\}$, define $\gamma_N = k\varepsilon_N$. Now define $M = \min\{N > N_0 | \gamma_N < \tau_{N+1}\}$ for some N_0. Then $\gamma := \gamma_M$ is a discrete stopping time and

$$\mu_{t \wedge \gamma} f(\cdot, U(t \wedge \gamma)) - \int_0^{t \wedge \gamma} \mu_s A f(\cdot, U(s))\, ds$$

is an $\{\mathscr{F}_{t+}^U\}$ martingale on (Ω, Q) for each f in $\mathscr{D}(A)$. Because γ is discrete we can now apply Corollary 3.4 to Theorem 4.6 to prove that

$$\mu_t = H_t(Y) = \pi_t,$$

Q_N a.s. on $\{t < \gamma\}$, and therefore on $\{t < \tau_N\}$. As $N \to \infty$, $\tau_N \to T$, and so $\mu_t = \pi_t$ for $t < T$.

The construction of A given in Theorem 4.1 no longer works here because the term $f(x)h(x)\nabla\psi(y)$, appearing there in $Af(x)\psi(y)$, is no longer necessarily bounded. However, we do know what the resolvent of A should be. If X_x solves the martingale problem for (A, δ_x), let $Y_y(t) = y + \int_0^t h(X_x(s))\, ds + W_1(t)$, where W_1 is a Brownian motion independent of X_x. Then we should have

$$F(x, y) = (\lambda - A)^{-1} G(x, y) = E\int_0^\infty e^{-\lambda t} G(X_x(t), Y_y(t))\, dt$$

and, in this case, we would get $AF(x, y) = \lambda F(x, y) - G(x, y)$. We shall use this to define A by considering the special case $G(x, y) = a(x)\exp(i\langle\theta, y\rangle)$ in which $a(x)$ is in $B(E)$ and $\theta \in \mathbb{R}^p$. For this, note that using

$$E\left[\exp i\langle\theta, Y_y(t)\rangle | \mathscr{F}^{X_x}\right] = \exp(i\langle\theta, y\rangle)\exp\left(-\tfrac{1}{2}|\theta|^2 t\right)\exp\left[i\left\langle\theta, \int_0^t h(X(s))\, ds\right\rangle\right],$$

we obtain the formula

$$E\int_0^\infty e^{-\lambda t} a(X_x(t))\exp\left[i\langle\theta, Y_y(t)\rangle\right] dt = f(x)\exp(i\langle\theta, y\rangle), \tag{4.25}$$

where

$$f(x) = E\int_0^\infty \exp\left[-\left(\lambda + \tfrac{1}{2}|\theta|^2\right)t\right] a(X_x(t))\exp\left[i\left\langle\theta, \int_0^t h(X_x(s))\, ds\right\rangle\right] dt. \tag{4.26}$$

Therefore, define

$$\begin{aligned} A = \operatorname{span}\{(&f(x)\exp(i\langle\theta, y\rangle), [\lambda f(x) - a(x)]\exp(i\langle\theta, y\rangle))| \\ &\theta \in \mathbb{R}^p, \lambda > 0, a(x) \in B(E) \text{ and } f \text{ is given by (4.26)}\}. \end{aligned}$$

We next must show that A has all the right properties. Clearly, $\mathscr{R}(\lambda - A) \supset \operatorname{span}\{a(x)\exp(i\langle\theta, y\rangle) | a(x) \in B(E),\ \theta \in \mathbb{R}^p\}$, which is bounded-pointwise dense in $B(E \times \mathbb{R}^p)$. To show that (X, Y) solves the martingale problem for $(A, \pi_0 \times \delta_0)$ and that A is well posed and satisfies (4.24), we need the Lemma

4.8, which we shall prove later. In this lemma, $\hat{A}_0$ denotes the bounded-pointwise closure of A_0. Recall from Theorem 3.2 that $\mathscr{R}(\lambda - \hat{A}_0) = B(E)$ for $\lambda > 0$.

LEMMA 4.8. *Let $a(x) \in B(E)$ and define $f(x)$ by (4.26). Then there are a sequence $\{f_k\}_{k=1}^{\infty} \subset \mathscr{D}(\hat{A}_0)$ and constants M, K such that:*

(i) $f(x) = \text{b.p.-lim}_{k\to\infty} f_k(x)$.
(ii) $\hat{A}_0 f_k(x) \le M + K|h(x)|$ *for all* $x \in E$.
(iii) $\lim_{k\to\infty} \hat{A}_0 f_k(x) + f_k(x)[-\frac{1}{2}|\theta|^2 + i\langle h(x), \theta\rangle] = \lambda f(x) - a(x)$ *for each* $x \in E$.

A is well posed. It is enough to show that (X_x, Y_y) solves the martingale problem for $(A, \delta_x \times \delta_y)$ for each $(x, y) \in E \times \mathbb{R}^P$. For $\nu \in \mathscr{P}(E \times \mathbb{R}^P)$, a solution to the martingale problem for (A, ν) is then $\int P_{x,y}\, d\nu(x, y)$, where $P_{x,y}$ is the measure on $D_{e\times\mathbb{R}^P}[0, \infty)$ induced by (X_x, Y_y). [$P_{x,y}(B)$ is measurable for each $B \in \mathscr{B}(D_{E\times\mathbb{R}^P}[0,\infty))$ because $R(\lambda - A)$ is dense in $B(E \times \mathbb{R}^P)$.] Uniqueness also is a consequence of the bounded-pointwise denseness of $\mathscr{R}(\lambda - A)$ (see the remark after Theorem 3.2).

For the next argument, we need the following fact: If $c \in \mathscr{D}(\hat{A}_0)$, then

$$\begin{aligned} V(t) = {} & c(X_x(t))\exp\left[i\langle\theta, Y_y(t)\rangle\right] \\ & - \int_0^t \left[\hat{A}_0 c(X_x(s)) + c(X_x(s))\left[-\tfrac{1}{2}|\theta|^2 + i\langle h(X_x(s)), \theta\rangle\right]\right] \\ & \qquad \times \exp\left[i\langle\theta, Y_y(s)\rangle\right] ds \end{aligned}$$

is a $\mathscr{F}_t^X \vee \mathscr{F}_t^Y$-martingale. Indeed, if the martingale $M(t) = c(X_x(t)) - \int_0^t \hat{A}_0 c(X_x(s))\, ds$ is right continuous, Itô's rule gives

$$\begin{aligned} V(t) - V(0) = {} & \int_0^t \exp\left[i\langle\theta, Y_y(s)\rangle\right] dM(s) \\ & + \int_0^t ic(X_x(s))\exp\left[i\langle\theta, Y_y(s)\rangle\right]\theta\, dW(s). \end{aligned}$$

Using (4.23) and the independence of X and W, one can prove that $V(t)$ is a martingale even if $M(t)$ is just a measurable process [see Ethier and Kurtz (1986), Chapter 2, Problem 22].

To show that (X_x, Y_y) solves the martingale problem for $(A, \delta_x \times \delta_y)$, we must show that given any $t_1 > t_2$ and $U \in \sigma\{X_x(s), Y_y(s) \colon s \le t\}$, a in $B(E)$ and f related to a by (4.26),

$$\begin{aligned} 0 = E\Bigg[\chi_U\Bigg[& f(X_x(t_1))\exp\left[i\langle\theta, Y_y(t_1)\rangle\right] - f(X_x(t_2))\exp\left[i\langle\theta, Y_y(t_2)\rangle\right] \\ & - \int_{t_1}^{t_2}\left[\lambda f(X_x(s)) - a(X_x(s))\right]\exp\left[i\langle\theta, Y_y(s)\rangle\right] ds\Bigg]\Bigg]. \end{aligned} \tag{4.27}$$

Let f_k be as in Lemma 4.7. Then (4.27) certainly holds if f is replaced by f_k and $\lambda f(x) - a(x)$ by $\hat{A}_0 f_k(x) + f_k(x)[-\frac{1}{2}|\theta|^2 + i\langle h(x), \theta\rangle]$. Now take $k \to \infty$, using dominated convergence, based on (4.23) and Lemma 4.8(ii), to get (4.27).

A satisfies (4.24). Let $f(x)$ be related to $a(x)$ as in (4.26) and let $f_k(x) \to f(x)$ as $k \to \infty$ as in Lemma 4.8. By Itô's rule

$$\mu_t(f_k(\cdot)\exp[i\langle\theta, Y(t)\rangle])$$
$$= \pi_0(f_k(\cdot)) + \int_0^t \mu_s((\hat{A}_0 f_k(\cdot)) + f_k(\cdot)[-\tfrac{1}{2}|\theta|^2 + i\langle h(\cdot),\theta\rangle])$$
$$\times \exp[i\langle\theta, Y(s)\rangle]\, ds + \int_0^t \eta^k(s)\, dI^\mu(s),$$

$$\eta^k(s) = [\mu_s(f_k(\cdot))i\theta + \mu_s(hf_k) - \mu_s(h)\mu_s(f_k)]\exp(i\langle\theta, Y(s)\rangle).$$

Using (4.22) and Lemma 4.8, it is again valid to pass to the limit. One thereby obtains (4.24) for $\phi(x, y) = f(x)\exp(i\langle\theta, y\rangle)$ and

$$\eta_s = [\mu_s(f)i\theta + \mu_s(hf) - \mu_s(h)\mu_s(f)]\exp(i\langle\theta, y\rangle).$$

PROOF OF LEMMA 4.8. Let

$$f(x) = E\int_0^\infty e^{-(\lambda+|\theta|^2/2)t} a(X_x(t))\exp\left[i\left\langle\theta, \int_0^t h(X_x(s))\, ds\right\rangle\right] dt.$$

Let $h_k(x) = h(x)\chi_{\{|h(x)|\le k\}}$. We claim that

$$f_x(x) = E\int_0^\infty e^{-(\lambda+|\theta|^2/2)t} a(X_x(t))\exp\left[i\left\langle\theta, \int_0^t h_k(X_x(s))\, ds\right\rangle\right] dt$$

satisfies (i)–(iii) of Lemma 4.8. Statement (i) is immediate by dominated convergence. To prove the other statements it suffices to show

$$(\lambda + |\theta|^2/2 - \hat{A}_0)^{-1}[a + f_k i\langle h_k, \theta\rangle](x) = f_k(x). \tag{4.28}$$

Indeed, given (4.28), $f_k \in \mathscr{D}(\hat{A}_0)$ and $\hat{A}_0 f_k = (\lambda + |\theta|^2/2)f_k - a - if_k\langle h_k, \theta\rangle$. Statements (ii) and (iii) then follow easily. To prove (4.28) observe that

$$E[f_k(X_x(t))\langle h_k(X_x(t)),\theta\rangle]$$
$$= E\Bigg[\langle h_k(X_x(t)),\theta\rangle E\bigg[\int_0^\infty e^{-(\lambda+|\theta|^2/2)s} a(X_y(s))$$
$$\times\exp\left[i\left\langle\theta, \int_0^t h_k(X_y(r))\, dr\right\rangle\right] ds \,\Big|\, y = X_x(t)\bigg]\Bigg]$$
$$= E\bigg[\langle h_k(X_x(t)),\theta\rangle e^{(\lambda+|\theta|^2/2)t}\int_t^\infty e^{-(\lambda+|\theta|^2/2)s} a(X_x(s))$$
$$\times \exp\left[i\left\langle\theta, \int_t^s h_k(X_x(r))\, dr\right\rangle\right] ds.$$

It follows upon interchanging orders of integration that

$$E\left[\int_0^\infty e^{-(\lambda+|\theta|^2/2)t} f_k(X_x(t))i\langle h_k(X_x(t)),\theta\rangle\, dt\right]$$
$$= f_k(x) - E\left[\int_0^\infty e^{-(\lambda+|\theta|^2/2)s} a(X_x(s))\, ds\right].$$

Thus

$$\begin{aligned}(\lambda + |\theta|^2/2 - \hat{A}_0)^{-1}[a + f_k i\langle h_k, \theta\rangle] \\ = E\left[\int_0^\infty e^{-(\lambda+|\theta|^2/2)t}[a(X_x(t)) + f_k(X_x(t))i\langle h_k(X_x(t)), \theta\rangle]\, ds\right] \\ = f_k(x).\end{aligned}$$

□

In applying Theorem 4.6 or 4.7, it is useful to note that one really needs only to check that $\mathscr{R}(\lambda - A_0)$ is bounded-pointwise dense in $B(E)$ only for all $\lambda > \lambda_0$ for some $\lambda_0 > 0$ rather than for all $\lambda > 0$. To see this, we show that the condition "$\lambda > 0$" in Proposition 2.1 can be replaced by "$\lambda \geq \lambda_0$ for some $\lambda_0 > 0$." Then, since Theorem 3.2 and, hence, Theorems 4.6 and 4.7, depend ultimately on Proposition 2.1, we can make the same replacement in these theorems also. But Proposition 2.1 relies on the identification of the Laplace transforms

$$\int_0^\infty e^{-\lambda t}\nu_t\, dt = \int_0^\infty e^{-\lambda t}\mu_t\, dt, \tag{4.29}$$

for all $\lambda > 0$, to conclude that $\mu_t = \nu_t$ for all t, and it is easy to check that $\mu_t = \nu_t$ for all t even if (4.29) is shown to hold only for all $\lambda \geq \lambda_0$, where $0 < \lambda_0 < \infty$. This remark is important in the next example, which gives a class of A_0 and h satisfying the hypotheses of Theorem 4.6.

EXAMPLE. As in the previous example, we consider the filtering model

$$\begin{aligned} dX &= m(X)\, dt + g(X)\, dB, \\ X(0) &\text{ has distribution } \pi_0, \\ dY &= h(X)\, dt + dW, \\ Y(0) &= 0. \end{aligned} \tag{4.30}$$

Here m: $\mathbb{R}^n \to \mathbb{R}^n$, g: $\mathbb{R}^n \to \mathbb{R}^{n\times d}$, h: $\mathbb{R}^n \to \mathbb{R}^p$ and B and W are independent Brownian motion processes independent of $X(0)$. Assume that:

(4.31)
(i) $m(x)$ and $g(x)$ are bounded twice continuously differentiable with bounded first and second derivatives.
(ii) $|h(x)| \leq K(1 + |x|^\beta)$ for some integer $\beta \geq 2$.
(iii) $E|X(0)|^{2\beta} < \infty$.

Let

$$A_0 f(x) = \tfrac{1}{2}tr[gg^T(x)D^2f(x)] + \langle\nabla f(x), m(x)\rangle$$

be defined on the domain

$$\mathscr{D}(A_0) = \{f \in C^2(\mathbb{R}^n) | f \text{ and its derivatives up to second order are bounded}\}.$$

We claim that A_0 and h satisfy the conditions of Theorem 4.6. Indeed, assumptions (ii) and (iii) of (4.31) easily imply

$$E\int_0^T |h(X(s))|^2\,ds < \infty \text{ and } E\int_0^T |h(X_x(s))|^2\,ds < \infty$$

for all x and any $T < \infty$. It remains to show that for some λ_0, $0 \le \lambda_0 < \infty$, $\mathscr{R}(\lambda - A_0)$ is bounded-pointwise dense in $\mathscr{B}(\mathbb{R}^n)$. For this, let U_λ be the set of $a(x) \in C(\mathbb{R}^n)$ such that if $f_a(x) = E\int_0^\infty e^{-\lambda t} a(X_x(t))\,dt$, $f_a \in \mathscr{D}(A_0)$. Then given $a \in U_\lambda$,

$$\frac{d}{dh} Ef_a(X_x(h))\bigg|_{h=0} = A_0 f_a(x).$$

At the same time,

$$Ef_a(X_x(h)) = e^{\lambda h} E\int_h^\infty e^{-\lambda t} a(X_x(t))\,dt$$

and, hence,

$$\frac{d}{dh} Ef_a(X_x(h)) = \lambda f_a(x) - a(x).$$

Thus, for $a \in U_\lambda$, $(\lambda - A_0) f_a(x) = a(x)$, so that $U_\lambda \subset \mathscr{R}(\lambda - A_0)$, and to show $\mathscr{R}(\lambda - A_0)$ is bounded-pointwise dense in $B(\mathbb{R}^n)$, it suffices to show U_λ is bounded-pointwise dense. We claim that, in fact, there is a $\lambda_0 > 0$, such that for all $\lambda \ge \lambda_0$, $C_b^2(\mathbb{R}^n) = \{a \in C^2(\mathbb{R}^n) | a$ and its first and second derivatives are bounded$\} \subset U_\lambda$. Since $C_b^2(\mathbb{R}^n)$ is bounded-pointwise dense in $B(\mathbb{R}^n)$, this will complete the proof.

Thus, to finish, observe that assumption (i) implies that $\partial X_x^k(t)/\partial x_i$ and $\partial^2 X_x^k(t)/\partial x_i\,\partial x_j$ exist, where X_x^k denotes the kth component of X and, that, if $a \in C_b^2(\mathbb{R}^n)$, $u(x, t) = Ea(X_x(t)) \in C^2$ for each t with

$$\frac{\partial u}{\partial x_i}(x, t) = E\left\{ \sum_k a_{x_k}(X_x(t)) \frac{\partial X_x^k}{\partial x_i}(t) \right\},$$

$$\frac{\partial^2 u}{\partial x_i\,\partial x_j}(x, y) = E\left\{ \sum_{k,l} a_{x_k x_l}(X_x(t)) \frac{\partial X_x^k}{\partial x_i} \frac{\partial X_x^l}{\partial x_j}(t) + \sum_k a_{x_k}(X_x(t)) \frac{\partial^2 X_x^k}{\partial x_i\,\partial x_j}(t) \right\}$$

[see, for example, Friedman (1975)]. Moreover, one can show that there is a $K < \infty$ such that if $\zeta(t)$ represents any first or second derivative of $X_x(t)$ or difference approximation thereof,

$$E|\zeta(t)|^\alpha \le Ke^{Kt}$$

for $\alpha = 1, 2$. For $\lambda > \lambda_0 = K$, it follows that

$$f(x) = E\int_0^\infty e^{-\lambda t} a(X_x(t))\,dt \in C_b^2(\mathbb{R}^n) = \mathscr{D}(A_0).$$

REFERENCES

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* **AC-28** 203–214.

ECHEVERRIA, P. E. (1982). A criterion for invariant measures of Markov processes. *Z. Wahrsch. verw. Gebiete* **61** 1–16.

ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence.* Wiley, New York.

FRIEDMAN, A. (1975). *Stochastic Differential Equations and Applications* **1**. Academic, New York.

FUJISAKI, M., KALLIANPUR, G. and KUNITA, H. (1972). Stochastic differential equations for the nonlinear filtering problem. *Osaka J. Math.* **9** 19–40.

HIJAB, O. (1986). On partially observed control of Markov processes. Unpublished.

KALLIANPUR, G. and KARANDIKAR, R. L. (1984). The nonlinear filtering problem for the unbounded case. *Stochastic Process. Appl.* **18** 57–66.

KURTZ, T. G. and OCONE, D. L. (1985). A martingale problem for conditional distributions and uniqueness for the nonlinear filtering problem. *Lecture Notes in Control and Information Sci.* **69** 224–234. Springer, New York.

LIPTSER, R. S. and SHIRYAYEV, A. N. (1977). *The Statistics of Random Processes I. General Theory.* Springer, Berlin.

PARDOUX, E. (1982). Equations du filtrage nonlineaire, de la prediction, et du lissage. *Stochastics* **6** 193–231.

STROOCK, D. W. and VARADHAN, S. R. S. (1979). *Multidimensional Diffusion Processes.* Springer, Berlin.

SZPIRGLAS, J. (1978). Sur l'equivalence d'équations différentielles stochastique à valeurs mesures intervenant dans le filtrage markovien non linéaire. *Ann. Inst. Poincaré Sect. B (N. S.)* **14** 33–59.

YOR, M. (1977). Sur les théories du filtrage et de la prédiction. *Séminaire de Probabilités XI. Lecture Notes in Math.* **581** 257–297. Springer, Berlin.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WISCONSIN-MADISON
MADISON, WISCONSIN 53706

DEPARTMENT OF MATHEMATICS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903