Stochastics, 1984, Vol. 13, pp. 1–23
0090-9491/84/1302-0001 \$18.50/0

Filtering of Diffusions Controlled Through their Conditional Measures†

V. E. BENEŠ
Bell Laboratories, Murray Hill, New Jersey 07974

and

I. KARATZAS‡
Columbia University, New York, New York 10027

(*Accepted for publication November 18, 1983*)

For the closed-loop nonlinear filtering problem with control in separated form (a functional of the conditional distribution measure), the Kallianpur–Striebel formula yields a stochastic equation for the unnormalized conditional distribution, given the past of the observations. Existence, uniqueness and measurability of solutions to this equation are discussed. The results give a partially positive answer to the question of admissibility of separated control laws. However, "pathological" nonanticipative but noncausal solutions appear, after the manner of Tsirelson's example.

1. INTRODUCTION

In this paper we are concerned with the "closed-loop" nonlinear filtering and control problem of studying the conditional distribution $v_t(A)=P_u\{x_t \in A \,|\, y_s; s \leqq t\}$, where $\{x_t\}$ is an R^r-diffusion process

†Presented at the 20th IEEE Conference on Decision and Control, San Diego, California, December 1981 (invited paper).

‡This research was supported in part by the National Science Foundation under MCS-81-03435.

satisfying the stochastic differential equation

$$dx_t = f(t, x_t, u_t)\,dt + dw_t \tag{1.1}$$

and $\{y_t\}$ is the observation process in R^1

$$y_t = \int_0^t h(s, x_s)\,ds + b_t; 0 \leqq t \leqq 1. \tag{1.2}$$

The process $\{(w_t, b_t); 0 \leqq t \leqq 1\}$ is Wiener in R^{r+1} and independent of the random variable x_0, which is assumed to have a known distribution μ_0. Here $\{u_t\}$ is a process adapted to the family $\mathscr{F}_t^y = \sigma(y_s; 0 \leqq s \leqq t)$ of σ-fields and takes values in a "control set" $\Gamma \subseteq R^m$; it provides feedback control, which closes the loop in Eqs. (1.1), (1.2).

The Kallianpur–Striebel formula (Eq. (2.4)) computes the conditional distribution $v_t(\cdot)$, by considering an appropriate unnormalized version $\mu_t(\cdot)$ with $v_t(A) = \mu_t(A)/\mu_t(R^r)$, $A \in \mathrm{Borel}_r$, in the form

$$\mu_t(A) = \int_{R^r} E\Bigg[1_{\{z+w_t \in A\}} \exp\Bigg\{ \int_0^t f(s, z+w_s, u_s) \cdot dw_s$$
$$-\tfrac{1}{2}\int_0^t |f(s, z+w_s, u_s)|^2\,ds + \int_0^t h(s, z+w_s)\,dy_s$$
$$-\tfrac{1}{2}\int_0^t h^2(s, z+w_s)\,ds \Bigg\} \Bigg] k_0(dz),$$

for any $\mathscr{F}_t^y$-adapted process $\{u_t\}$.† On the other hand, whenever μ_0 has a density $p_0(x)$, then it is checked from the above formula that μ_t also has a density, namely: $\rho_t(x) = \int_{R^r} q_t(x; z) p_0(z)\,dz$, with $q_t(x; z)\,dx = E[1_{\{z+w_t \in dx\}} \exp\{\ldots\}]$ where the exponent is the same as on the right-hand side of the Kallianpur–Striebel formula. The function ρ is called an unnormalized conditional density and satisfies, under proper conditions, the Zakai stochastic partial differential equation [15]:

†Juxtaposition denotes inner product in R^r: $x \cdot y = \sum_{i=1}^r x_i y_i$, while ∇, Δ stand for gradient, Laplacian, respectively.

$$\begin{aligned} d_t\rho_t(x) = [\tfrac{1}{2}\Delta - f(t,x,u_t)\cdot\nabla \\ - tr\{\nabla f(t,x,u_t)\}]\rho_t(x)\,dt \\ + h(t,x)\rho_t(x)\,dy_t; t>0 \\ \rho_0(x) = p_0(x). \end{aligned} \tag{1.3}$$

It has been suggested ([9, 5, 4]) that, for purposes of optimization, the control u_t at time t need depend on the past $\{y_s; s \leqq t\}$ of the observation sample path only through the conditional density ρ_t; this is equivalent to saying that the latter should be thought of as a "sufficient statistic" containing all the pertinent information. In [4] the interested reader can find some first steps toward a rigorous theory along these lines, originally proposed in a heuristic fashion by Mortensen [9]. Such a program calls, ultimately, for the solution of Eq. (1.3) where u_t is replaced by $U(t, \rho_t)$ (a functional of the conditional density), in the strong sense that ρ_t is to be adapted to $\mathscr{F}_t^y$.

This program seems to be hard: the resulting equation is a stochastic nonlinear partial differential equation with nonlocal drift and potential terms. Our aim in this paper is to show that by switching attention from the Zakai equation to the Kallianpur–Striebel (K.S.) formula with u_t replaced by $v(t, \mu_t)$, the interesting questions can be formulated in their "natural" setting and attacked successfully. We call a control process $\{u_t\}$ of the form $u_t = v(t, \mu_t)$, $0 \leqq t \leqq 1$, for a suitable function v, a *separated* control process; for such $\{u_t\}$, (K.S.) becomes a *stochastic equation for the measure-valued process* $\mu = \{\mu_t; 0 \leqq t \leqq 1\}$. We shall be concerned with questions of existence, uniqueness, and $\mathscr{F}^y$-measurability of the solution to this equation.

2. FORMULATION AND SUMMARY

Let us start with a basic probability space $(\Omega, \mathscr{F}, \tilde{P}; \mathscr{F}_t)$ and an associated increasing family of σ-fields $\{\mathscr{F}_t; 0 \leqq t \leqq 1\}$, such that $\mathscr{F}_1 = \mathscr{F}$. On this space is given an $(r+1)$-dimensional Wiener process $\{(w_t, y_t); 0 \leqq t \leqq 1\}$ and an R^r-valued random variable x_0, independent of the Wiener process and with a specified Lebesgue–Stieltjes measure μ_0. The basic state process is given by

$$\{x_t\}=\{x_0+w_t; 0\leqq t\leqq 1\}.$$

We make the following assumptions on the drift functions f, h appearing in Eqs. (1.1), (1.2):

A.1: $f(t,x,u)$: $[0,1]\times R^r\times\Gamma\to R^r$ is a bounded, continuous function, with Γ a compact, convex subset of some Euclidean space R^m.

A.2: $h(t,x)$: $[0,1]\times R^r\to R^1$ and its partial derivatives

$$\frac{\partial}{\partial t}h,\ \frac{\partial}{\partial x_i}h,\ \frac{\partial^2}{\partial x_i\,\partial x_j}h$$

are bounded, continuous functions.

The symbol K is used henceforth as a generic upper bound for any of these quantities.

DEFINITION 2.1 A process $\{u_t\}$ is called an *admissible control process* if it takes values in the space Γ and, for each $0\leqq t\leqq 1$, u_t is an $\mathscr{F}_t^y$-measurable random variable. The class of all such processes is denoted by $\mathscr{A}$.

Similarly, $\{u_t\}$ is called a *wide-sense control process* if it is Γ-valued and adapted to $\mathscr{F}_t=\sigma(x_0, w_s, y_s; s\leqq t)$. The resulting class is denoted by $\mathscr{U}$. □

Finally, $P(\mathring{P})$ denotes the restriction of the probability measure $\tilde{P}$ to the σ-field $\mathscr{F}_1^w=\sigma(w_t; t\leqq 1)$ $(\mathscr{F}_1^x=\sigma(x_0+w_t; t\leqq 1)$, respectively), and $E(\mathring{E})$ the corresponding expectation.

For any $u\in\mathscr{A}$, a new measure can be defined by $P_u(E)=\int_E \exp\zeta_0^1(u;x_0)\,d\tilde{P}$, where

$$\zeta_0^t(u;z)=\int_0^t f(s,z+w_s,u_s)\cdot dw_s-\tfrac{1}{2}\int_0^t |f(s,z+w_s,u_s)|^2\,ds$$

$$+\int_0^t h(s,z+w_s)\,dy_s-\tfrac{1}{2}\int_0^t h^2(s,z+w_s)\,ds; z\in R^r. \tag{2.1}$$

According to Girsanov's theorem (e.g. [7], Ch. 6), P_u is a probability measure under which the process

$$\begin{bmatrix} w_t^u \\ b_t \end{bmatrix} = \begin{bmatrix} w_t \\ y_t \end{bmatrix} - \begin{bmatrix} \int_0^t f(s, x_0 + w_s, u_s)\, ds \\ \int_0^t h(s, x_0 + w_s)\, ds \end{bmatrix}; \quad 0 \leqq t \leqq 1$$

is Wiener and independent of x_0, with $P_u x_0^{-1} = \mu_0$. For any $g: R^r \to R^1$ bounded measurable, Bayes' rule says that

$$E_u[g(x_t) \mid \mathscr{F}_t^y] = \frac{\tilde{E}[g(x_t) \exp \zeta_0^t(u; x_0) \mid \mathscr{F}_t^y]}{\tilde{E}[\exp \zeta_0^t(u; x_0) \mid \mathscr{F}_t^y]} \triangleq \frac{\sigma_t(g)}{\sigma_t(1)}. \tag{2.2}$$

But under $\tilde{P}$, the processes $\{x_t\}$ and $\{y_t\}$ are independent, so that the x path can be integrated out to give

$$\begin{aligned} \sigma_t(g) &= \mathring{E}[g(x_0 + w_t) \exp \zeta_0^t(u; x_0)] \\ &= \int_{R^r} E[g(z + w_t) \exp \zeta_0^t(u; z)] \mu_0(dz). \end{aligned} \tag{2.3}$$

Therefore, the conditional distribution measure is given by

$$\nu_t(A) = \mu_t(A)/\mu_t(R^r); A \in \mathrm{Borel}_r, \text{ where:}$$

$$\begin{aligned} \mu_t(A) = \int_{R^r} E\Bigg[1_{\{z + w_t \in A\}} \exp \Bigg\{ & \int_0^t f(s, z + w_s) \cdot dw_s \\ & - \tfrac{1}{2} \int_0^t |f(s, z + w_s, u_s)|^2\, ds \\ & + \int_0^t h(s, z + w_s)\, dy_s \\ & - \tfrac{1}{2} \int_0^t h^2(s, z + w_s)\, ds \Bigg\} \Bigg] \mu_0(dz). \end{aligned} \tag{2.4}$$

We have already referred to formula (2.4) as the Kallianpur–Striebel formula, and to $\mu_t(\cdot)$ as the unnormalized conditional distribution measure at time t, given the observation record $\{y_s; 0 \leqq s \leqq t\}$. For a fixed t, μ_t is an $\mathscr{F}_t^y$-measurable, $\mathscr{M}^r$-valued random variable, where $\mathscr{M}^r$, the space of positive, finite measures on Borel$_r$, is a Polish space in the weak topology that can be metrized by the Prokhorov metric:

$$\sum_{i=1}^{\infty} 2^{-i} \frac{|(m, g_i) - (n, g_i)|}{\|g_i\|}; \ (m, g) = \int_{R^r} g(x) m(dx).$$

Here, $\{g_i\}_{i=1}^{\infty}$ is a dense subset of the space of real-valued functions on R^r which are bounded and uniformly continuous (c.f. [10], Chapter II). It can be shown (e.g. [3]) that the g_i's may be chosen to be infinitely continuously differentiable, and that instead of 2^{-i} one can use a different sequence $\{\varphi_i\}_{i=1}^{\infty}$ of weights, in an equivalent metric

$$\rho(m, n) = \sum_{i=1}^{\infty} \varphi_i \frac{|(m, g_i) - (n, g_i)|}{\|g_i\|}$$

for the weak topology. We shall require that the φ_i's have the property

$$\sum_{i=1}^{\infty} \varphi_i \left(1 + \frac{\|\nabla g_i\|}{\|g_i\|} \right) < \infty,$$

where $\|g\| = \sup_{x \in R^r} |g(x)|$, $\|\nabla g\| = \sup_{x \in R^r} |\nabla g(x)|$.

The process $\mu = \{\mu_t; 0 \leqq t \leqq 1\}$ takes values in the space $C = C([0,1]; \mathscr{M}^r)$ of continuous functions from $[0,1]$ into $\mathscr{M}^r$; under the metric

$$d(\mu, \tilde{\mu}) = d_1(\mu, \tilde{\mu}); d_t(\mu, \tilde{\mu}) \triangleq \sup_{0 \leqq s \leqq t} \rho(\mu_s, \tilde{\mu}_s),$$

the space (C, d) is complete metric.

An integration by parts with respect to $\{y_t\}$ in (2.4) has the effect of defining μ_t for all $y \in C_{[0,1]}$, not just on a subset of full Wiener measure. In fact, writing

$$y_t h(t, z + w_t) = \int_0^t y_s \left\{ \frac{\partial}{\partial t} h(s, z + w_s) \right.$$

$$+\tfrac{1}{2}\Delta h(s,z+w_s)\Big\}\,ds+\int_0^t h(s,z+w_s)\,dy_s$$

$$+\int_0^t y_s\nabla h(s,z+w_s)\cdot dw_s$$

by virtue of the Itô formula, we have

$$\mu_t(A)=\int_{R^r} E\Bigg[1_{\{z+w_t\in A\}}\exp\Bigg\{\int_0^t f(s,z+w_s,u_s)\cdot dw_s$$

$$-\tfrac{1}{2}\int_0^t |f(s,z+w_s,u_s)|^2\,ds+\xi_0^t(-y\nabla h)$$

$$+y_t h(t,z+w_t)+\int_0^t l(s,z+w_s)\,ds\Bigg\}\Bigg]\mu_0(dz), \tag{2.4$'$}$$

with the notation $\xi_0^t(\phi)=\int_0^t\phi_s\cdot dw_s-\frac{1}{2}\int_0^t|\phi_s|^2\,ds$ and

$$l(t,z)=\tfrac{1}{2}y_t^2|\nabla h(t,z)|^2-y_t\Big\{\frac{\partial}{\partial t}h(t,z)+\tfrac{1}{2}\Delta h(t,z)\Big\}-\tfrac{1}{2}h^2(t,z).$$

The expression in (2.4)$'$ is parametrized in a simple way by the observation process sample path $\{y_t\}$.

We now consider a Borel measurable function $v(t,m)\colon[0,1]\times\mathscr{M}^r\to\Gamma$, and address the question of solving the resulting "separated" version of (2.4)$'$, i.e. with u_t replaced by $v(t,\mu_t)$:

$$\mu_t(A)=\int_{R^r} E\Bigg[1_{\{z+w_t\in A\}}\exp\Bigg\{\int_0^t f(s,z+w_s,v(s,\mu_s))\cdot dw_s$$

$$-\tfrac{1}{2}\int_0^t |f(s,z+w_s,v(s,\mu_s))|^2\,ds$$

$$+\xi_0^t(-y\nabla h)+y_t h(t,z+w_t)+\int_0^t l(s,z+w_s)\,ds\Bigg\}\Bigg]\mu_0(dz) \tag{2.5}$$

for some element μ of C. Equivalently, one can fix $\mu_0 \in \mathscr{M}^r$ and $y \in C_{[0,1]}$, define the operator $T: C \to C$ by $T_t\mu(A) =$ right-hand side of (2.5), and search for fixed points of T. We shall write $T(y, \mu)$ when dependence on the path y is of import.

In Section 3 we discuss the case of $v(t, \cdot)$ Lipschitz continuous on $\mathscr{M}^r$ and obtain existence, uniqueness and continuous dependence (on μ_0 and y) results for the solution of $\mu = T\mu$. In Section 4 the case of continuous $v(\cdot,\cdot)$ is addressed, for which existence and measurable selection results are established. Some counterexamples appear in Section 5.

3. THE LIPSCHITZ CASE

Throughout this section, and in addition to the assumptions of Section 2, it is assumed that the functions f, v satisfy a Lipschitz condition:

$$|f(t,z,u_1) - f(t,z,u_2)| \leqq K|u_1 - u_2|; \forall (t,z) \in [0,1] \times R^r; u_1, u_2 \in \Gamma \tag{3.1}$$

$$|v(t,m) - v(t,n)| \leqq K\rho(m,n); \forall t \in [0,1]; m, n \in \mathscr{M}^r. \tag{3.2}$$

It is proved in this section that, under the above assumptions, Eq. (2.5) has a unique solution in C such that, for each t, μ_t is a measure-valued, $\mathscr{F}_t^y$-measurable random variable. By composition of maps, $\{u_t = v(t, \mu_t)\}$ is then in $\mathscr{A}$.

PROPOSITION 3.1 *Under the assumptions of this section and for any* $\mu_0 \in \mathscr{M}^r, y \in C_{[0,1]}$, *there exists a constant* q *depending on* K *and* $a = \sup_{0 \leqq t \leqq 1} |y_t|$ *such that, for any two elements* μ, $\tilde{\mu}$ *in* C:

$$\rho^4(T_t\tilde{\mu}, T_t\mu) \leqq qt \int_0^t \rho^4(\tilde{\mu}_s, \mu_s)\, ds; 0 \leqq t \leqq 1. \tag{3.3}$$

Proof With the notation

$$u_t = v(t, \mu_t), \tilde{u}_t = v(t, \tilde{\mu}_t)$$

$$f_t = f(t, z + w_t, u_t), \tilde{f}_t = f(t, z + w_t, \tilde{u}_t)$$

$$I_t = \xi_0^t(f), \tilde{I}_t = \xi_0^t(\tilde{f})$$

$$J_t = \tilde{I}_t - I_t \tag{3.4}$$

we have for any bounded, continuous function $g: R^r \to R^1$:

$$|(T_t\tilde{\mu}, g) - (T_t\mu, g)| \leqq \|g\| \int_{R^r} (E(\exp\{2\zeta_0^t(u; z)\}))^{1/2}.$$

$$(E|\exp J_t - 1|^2)^{1/2} \mu_0(dz).$$

It is easily checked from (2.1), (2.4)′ that

$$\zeta_0^t(u; z) = \xi_0^t(f - y\nabla h) + y_t h(t, z + w_t) + \int_0^t e(s, z + w_s; u_s)\, ds \tag{3.5}$$

with $e(t, z; u) = l(t, z) + y_t f(t, z, u) \cdot \nabla h(t, z)$.

We denote henceforth by q any constant depending on K and a, not necessarily the same throughout this paper. Evidently from (3.5), $2\zeta_0^t(u, z) \leqq \xi_0^t\{2(f - y\nabla h)\} + q$ and since, by the Girsanov theorem [7; chapter 6], $E \exp \xi_0^t(\varphi) = 1$ for bounded φ, we conclude that

$$(E \exp\{2\zeta_0^t(u; z)\})^{1/2} \leqq q, \quad \text{for all} \quad z \in R^r.$$

On the other hand, by the Cauchy inequality:

$$E|\exp J_t - 1|^2 = E[|\exp J_t - 1|^2 \cdot 1_{\{|J_t| \geqq 1\}} + |\exp J_t - 1|^2 \cdot 1_{\{|J_t| < 1\}}]$$

$$\leqq (E|\exp J_t - 1|^4 \cdot P(|J_t| \geqq 1))^{1/2} + 4E(J_t^2) \tag{3.6}$$

where we have used the estimate: $|e^x - 1| \leqq 2|x|$, for $|x| \leqq 1$. Estimation of the various terms on the right-hand side of (3.6) is rather straightforward; for instance, a.s. P;

$$J_t^2 \leqq 2\left[\int_0^t (\tilde{f}_s - f_s) \cdot dw_s\right]^2 + 2K^2 t \int_0^t |\tilde{f}_s - f_s|^2\, ds, \text{ and so}$$

$$EJ_t^2 \leqq 2(1 + K^2 t) E \int_0^t |\tilde{f}_s - f_s|^2\, ds \leqq q t^{1/2} \left(\int_0^t \rho^4(\tilde{\mu}_s, \mu_s)\, ds\right)^{1/2}. \tag{3.7}$$

Similarly, a.s. P; $J_t^4 \leqq 4[\int_0^t (\tilde{f}_s - f_s)\cdot dw_s]^4 + 4K^4 t^3 \int_0^t |\tilde{f}_s - f_s|^4\, ds$, and by [12; Thm. 4, p. 23]:

$$P(|J_t| \geqq 1) \leqq E(J_t^4) \leqq 144 \cdot t \int_0^t E|\tilde{f}_s - f_s|^4\, ds + 4K^3 t^3 \int_0^t E|\tilde{f}_s - f_s|^4\, ds$$

$$\leqq qt \cdot \int_0^t \rho^4(\tilde{\mu}_s, \mu_s)\, ds.$$

Finally, an application of Itô's formula yields

$$\exp(J_t) - 1 = \int_0^t \exp(J_s)(\tilde{f}_s - f_s)\cdot dw_s + \int_0^t \exp(J_s)(\tilde{f}_s - f_s)\cdot f_s\, ds$$

whence, after some simple algebra and using the above-mentioned theorem: $E(\exp(J_t) - 1)^4 \leqq qt$. Substitution into (3.6) gives

$$E|\exp(J_t) - 1|^2 \leqq qt^{1/2} \cdot \{\int_0^t \rho^4(\tilde{\mu}_s, \mu_s)\, ds\}^{1/2},$$

and

$$|(T_t\tilde{\mu}, g_i) - (T_t\mu, g_i)| \leqq \|g_i\| qt^{1/2} \left\{\int_0^t \rho^4(\tilde{\mu}_s, \mu_s)\, ds\right\}^{1/4},$$

for every g_i; $i \geqq 1$ in the definition of the metric $\rho(\cdot,\cdot)$. Multiplying both sides of the above inequality by $\varphi_i \|g_i\|^{-1}$ and adding over $i \geqq 1$, we obtain (3.3).

THEOREM 3.1 *Under the assumptions of this section, Eq.* (2.5) *has a unique solution process* μ *in* C, *for each given* μ_0 *in* $\mathscr{M}^r$ *and* $y \in C_{[0,1]}$. *Furthermore,* μ_t *is* $\mathscr{F}_t^y$*-measurable for each* $t \in [0, 1]$, *and the process* $u = \{v(t, \mu_t); 0 \leqq t \leqq 1\}$ *is admissible:* $u \in \mathscr{A}$.

Proof Uniqueness follows directly from inequality (3.3) by a Gronwall-type argument. Existence is proved by the Picard iteration procedure: one starts with $\mu^{(0)}$ such that $\mu_t^{(0)} = \mu_0$, for all $0 \leqq t \leqq 1$ and defines the sequence $\{\mu^{(n)}\}_{n=0}^{\infty} \subseteq C$ recursively by $\mu^{(n+1)} = T\mu^{(n)}$; $n \geqq 0$. If $e_n(t) \triangleq d_t^4(\mu^{(n+1)}, \mu^{(n)})$; $n \geqq 0$, then it follows from relation (3.3) that:

$$e_n(t) \leqq \|e_0\| \frac{(qt)^n}{n!},\ 0 \leqq t \leqq 1; n \geqq 0.$$

Standard arguments now show that $\{\mu^{(n)}\}_{n=0}^{\infty}$ is a Cauchy sequence in C, and it converges to some element μ of C, by completeness of this space. Note that, for each $n \geqq 0$, $\mu_t^{(n)}$ is $\mathscr{F}_t^y$-measurable, and therefore so is μ. Admissibility of μ follows by composition of maps. It is easily checked that μ is a fixed point of T.

THEOREM 3.2 *Continuous dependence on* μ_0, y.

Suppose the assumptions of this section are satisfied. Then

i) For a fixed $y \in C_{[0,1]}$, let $\{\mu_0^{(n)}\}_{n=1}^{\infty}$ be a sequence of probability measures in $\mathscr{M}^r$ converging to the probability measure μ_0 in the weak topology of that space. If $\{\mu^{(n)}\}_{n=1}^{\infty}$, μ are the corresponding solutions of Eq. (2.5) according to Theorem 3.1, then: $d(\mu^{(n)}, \mu) \to 0$, as $n \to \infty$.

ii) For a fixed μ_0 in $\mathscr{M}^r$, let y, $\tilde{y}$ be two elements of $C_{[0,1]}$ such that: $\sup_{0 \leqq t \leqq 1} |y_t| \triangleq \|y\| \leqq a$, $\|\tilde{y}\| \leqq a$ and $\|\tilde{y} - y\| \leqq \varepsilon$. If μ, $\tilde{\mu}$ are the corresponding solutions of Eq. (2.5), then: $d(\tilde{\mu}, \mu) \leqq q\varepsilon$, for some constant q depending only on a, K.

Proof To establish (i) we notice that for any function g_i as in the definition of the metric $\rho(\cdot,\cdot)$, we have

$$|(\mu_t^{(n)}, g_i) - (\mu_t, g_i)| \leqq \|g_i\| \int_{R^r} E[\exp \zeta_0^t(u; z) |\exp J_t^{(n)} - 1|] .$$

$$\cdot \mu_0^{(n)}(dz) + |(\mu_0^{(n)}, m_t^{(i)}) - (\mu_0, m_t^{(i)})|, \tag{3.8}$$

where u_t, I_t are given as in (3.4), $u_t^{(n)} \triangleq v(t, \mu_t^{(n)})$,

$$f_t^{(n)} \triangleq f(t, z + w_t, u_t^{(n)}), \quad I_t^{(n)} \triangleq \xi_0^t(f^{(n)}), \quad J_t^{(n)} \triangleq I_t^{(n)} - I_t,$$

and

$$m_t^{(i)}(z) \triangleq E[g_i(z + w_t) \exp \zeta_0^t(u; z)]$$

$$= E\left[g_i(z + w_t) \exp \left\{ \xi_0^t(f - y\nabla h) + y_t h(t, z + w_t) \right.\right.$$

$$\left.\left. + \int_0^t e(s, z + w_s; u_s)\, ds \right\} \right]$$

$$= \check{E}\left[g_i(z + w_t) \exp\{y_t h(t, z + w_t)\} \exp\left\{ \int_0^t e(s, z + w_s; u_s)\, ds \right\} \right],$$

by virtue of (3.5). Above, $\check{E}$ denotes expectation with respect to the probability measure $\check{P}$ which is absolutely continuous with respect to P, with Radon–Nikodým derivative $\zeta_0^1(f-y\nabla h)$. Standard arguments yield now

$$\|\nabla m_t^{(i)}\| \leqq q(\|g_i\|+\|\nabla g_i\|).$$

Let us consider the family of functions $\{k_t^{(i)}(\cdot)\}_{\substack{i\geqq 1\\ 0\leqq t\leqq 1}}$ defined by

$$k_t^{(i)}(z) \triangleq \frac{m_t^{(i)}(z)}{\|g_i\|+\|\nabla g_i\|},$$

which is uniformly bounded and equicontinuous at every point $z\in R^r$, since

$$\|k_t^{(i)}\|\leqq 1,\ \|\nabla k_t^{(i)}\|\leqq q$$

hold for every $i\geqq 1$ and $0\leqq t\leqq 1$. Therefore, by invoking Theorem II. 6.8 in [10], we conclude that

$$\varepsilon(n)\triangleq \sup_{i\geqq 1}\sup_{0\leqq t\leqq 1}|(\mu_0^{(n)},k_t^{(i)})-(\mu_0,k_t^{(i)})|$$

converges to zero, as $n\to\infty$. On the other hand, we have from Proposition 3.1:

$$\int_{R^r} E[\exp\zeta_0^t(u;z)|\exp J_t^{(n)}-1|]\mu_0^{(n)}(dz)\leqq qt^{1/4}\left\{\int_0^t \rho^4(\mu_s^{(n)},\mu_s)\,ds\right\}^{1/4},$$

and substituting back in (3.8) we obtain:

$$\frac{|(\mu_t^{(n)},g_i)-(\mu_t,g_i)|}{\|g_i\|}\leqq q\left(1+\frac{\|\nabla g_i\|}{\|g_i\|}\right)\left[t^{1/4}\left\{\int_0^t \rho^4(\mu_s^{(n)},\mu_s)\,ds\right\}^{1/4}+\varepsilon(n)\right]$$

as well as

$$\rho^4(\mu_t^{(n)},\mu_t)\leqq q\left[t\int_0^t \rho^4(\mu_s^{(n)},\mu_s)\,ds+\varepsilon^4(n)\right].$$

By the Gronwall inequality, $d(\mu^{(n)},\mu)\leqq q\cdot\varepsilon(n)\to 0$, as $n\to\infty$.

In order to establish (ii), we write: $\zeta_0^t(u;z)=I_t+L_t(y)$, $\zeta_0^t(\tilde{u};z)=\tilde{I}_t+L_t(\tilde{y})$, with the same notation as in (3.4) and

$$L_t(y)\triangleq\xi_0^t(-y\nabla h)+y_t h(t,z+w_t)+\tfrac{1}{2}\int_0^t y_s^2|\nabla h(s,z+w_s)|^2\,ds$$

$$-\int_0^t\left\{y_s\frac{\partial}{\partial t}h(s,z+w_s)+\tfrac{1}{2}y_s\Delta h(s,z+w_s)+\tfrac{1}{2}h^2(s,z+w_s)\right\}ds,$$

$N_t=L_t(y)-L_t(\tilde{y})$. We have

$$|(\tilde{\mu}_t,g)-(\mu_t,g)|\leqq\|g\|\left[\int_{R^r}\sqrt{E\{\exp 2\zeta_0^t(u;z)\}}\sqrt{E|\exp J_t-1|^2}\cdot\mu_0(dz)\right.$$

$$\left.+\int_{R^r}\sqrt{E\{\exp 2\zeta_0^t(\tilde{u};z)}\sqrt{E|\exp N_t-1|^2}\cdot\mu_0(dz)\right].$$

It is easily seen that

$$E|\exp N_t-1|^2\leqq\sqrt{E|\exp N_t-1|^2E(N_t^4)}+4E(N_t^2)\leqq q\varepsilon^2$$

so that:

$$\rho(\tilde{\mu}_t,\mu_t)\leqq q\left[\left\{t\int_0^t\rho^4(\tilde{\mu}_s,\mu_s)\,ds\right\}^{1/4}+\varepsilon\right];0\leqq t\leqq 1.$$

Assertion (ii) follows again by a Gronwall-type argument.

4. THE CONTINUOUS CASE

In this section, we study Eq. (2.5) under appropriate continuity assumptions on the function $v(t,m)$. We obtain existence and measurable selection results.

THEOREM 4.1 *Suppose $v(t,m)$ is a continuous function on $[0,1]\times\mathscr{M}^r$ into Γ. Then, for any given $y\in C_{[0,1]}$ and $\mu_0\in\mathscr{M}^r$, there exists an element μ of C satisfying Eq. (2.5): $\mu=T\mu$.*

Proof For any continuous $u: [0, 1] \to \Gamma$, we define

$$(Su)(t) \triangleq v(t, \mu_t(u)) \tag{4.1}$$

where

$$\mu_t(u)(A) = \int_{R^r} E\left[1_{\{z+w_t \in A\}} \exp\left\{ \int_0^t f(s, z+w_s, u(s)) \cdot dw_s \right.\right.$$

$$- \tfrac{1}{2} \int_0^t |f(s, z+w_s, u(s))|^2 \, ds$$

$$\left.\left. + \xi_0^t(-y\nabla h) + y_t h(t, z+w_t) + \int_0^t l(s, z+w_s) \, ds \right\} \right] \mu_0(dz). \tag{4.2}$$

We first prove that the closure $\bar{K}$ of the set

$$K = \{\mu_t(u) | u: [0, 1] \to \Gamma \text{ Borel measurable; } 0 \leqq t \leqq 1\} \tag{4.3}$$

is compact in the weak topology of $\mathscr{M}^r$; for that it is sufficient to establish the uniform tightness of K (Prokhorov's Theorem; cf. [10], Thm. II. 6.7). In fact, by the Cauchy inequality,

$$\mu_t^2(u)(A) \leqq q \int_{R^r} P\{z + w_t \in A\} \mu_0(dz)$$

for any $A \in \text{Borel}_r$. Taking $A = \{x: |x| > N\}$, we have the estimate:

$$\mu_t^2(u)(R^r / B_N) \leqq q\left[\mu_0(R^r / B_{\frac{N}{r+1}}) + \int_{\{z; |z| \leqq N/(r+1)\}} P\{|z + w_t| > N\} \mu_0(dz) \right],$$

where we denote by B_α the closed sphere of radius $\alpha > 0$ in R^r. Moreover, on $\{z: |z| \leqq N/(r+1)\}$ we have

$$P\{|z + w_t| > N\} \leqq \sum_{i=1}^r P\left\{ |z_i + w_i(t)| > \frac{N}{r} \right\} \leqq 2 \sum_{i=1}^r P\left\{ b_t > \frac{N}{r} - |z_i| \right\}$$

$$\leqq 2r\left[1-\Phi\left(\frac{N}{r(r+1)t^{1/2}}\right)\right]$$

with $\{b_t\}$ a one-dimensional Brownian motion. Standard estimates (e.g. [8], p. 4) now yield

$$\mu_t^2(u)(R^r/B_N) \leqq q\left[\mu_0(R^r/B_{\frac{N}{r+1}}) + \frac{r^2(1+r)}{N}\right]$$

for any $u: [0,1] \to \Gamma$ Borel measurable and any $t \in [0,1]$.

Because any probability measure on R^r is tight [10; Thm. II. 3.2], $N = N(\varepsilon)$ can be chosen in such a way that

$$\mu_0(R^r/B_{\frac{N}{r+1}}) \leqq \frac{\varepsilon^2}{2}, \quad N \geqq \frac{2r^2(1+r)}{\varepsilon^2},$$

for any given $\varepsilon > 0$. Thus

$$\sup_{\substack{u:[0,1]\to\Gamma \\ \text{Borel}}} \sup_{0\leqq t\leqq 1} \mu_t(u)(R^r/B_N) \leqq q\varepsilon,$$

which proves uniform tightness of K in (4.3).

The restriction of the function $v(t,m)$ to $[0,1] \times \bar{K}$ is continuous on a compact set (Tychonoff), hence uniformly continuous on it. Therefore, there exists a continuous (modulus) function $\varepsilon(h)\downarrow 0$ (as $h \downarrow 0$) such that, for any $0 \leqq t < t+h \leqq 1$; $m, n \in \bar{K}$:

$$|v(t+h,m) - v(t,n)| \leqq \varepsilon(h + \rho(m,n)).$$

Now for any g_i as in the definition of the metric $\rho(\cdot,\cdot)$, we have

$$|(\mu_{t+h}(u), g_i) - (\mu_t(u), g_i)| \leqq \|\nabla g_i\| \cdot \int_{R^r} E[|w_{t+h} - w_t| \exp \zeta_0^{t+h}(u;z)]\mu_0(dz)$$

$$+ \|g_i\| \cdot \int_{R^r} E[\exp \zeta_0^t(u;z) \cdot |\exp \zeta_t^{t+h}(u;z) - 1|]\mu_0(dz)$$

$$\leqq q\left[\|\nabla g_i\| \cdot h^{1/2} + \|g_i\| \int_{R^r} \sqrt{E|\exp \zeta_t^{t+h}(u;z) - 1|^2}\,\mu_0(dz)\right].$$

As in the proof of Proposition 3.1, we obtain

$$E\left|\exp\zeta_t^{t+h}(u;z)-1\right|^2 \leqq \sqrt{E\left|\exp\zeta_t^{t+h}(u;z)-1\right|^4 \cdot E\left|\zeta_t^{t+h}(u;z)\right|^4}$$

$$+4E\left|\zeta_t^{t+h}(u;z)\right|^2.$$

It is not hard to see from (3.5) that: $E\left|\zeta_t^{t+h}(u;z)\right|^2 \leqq q\cdot j^2(h)$, where

$$j(h) \triangleq h^{1/2} + \sup_{0\leqq t<t+h\leqq 1} |y_{t+h}-y_t|$$

converges to zero as $h\downarrow 0$, and that $E\left|\zeta_t^{t+h}(u;z)\right|^4 \leqq q\cdot j^4(h)$. Therefore,

$$\frac{\left|(\mu_{t+h}(u),g_i)-(\mu_t(u),g_i)\right|}{\|g_i\|} \leqq q\cdot j(h)\left(1+\frac{\|\nabla g_i\|}{\|g_i\|}\right)$$

holds for every $i\geqq 1$, and

$$\rho(\mu_{t+h}(u),\mu_t(u)) \leqq q\cdot j(h)$$

follows upon multiplying by the weights φ_i and then summing up over $i\geqq 1$.

Consequently, $\left|Su(t+h)-Su(t)\right| \leqq \varepsilon(h+q\cdot j(h)) \triangleq \delta(h)$ holds for any $0\leqq t<t+h\leqq 1$. Now we consider the subset of $C([0,1];R^m)$ defined by

$$B=\left\{u\colon[0,1]\to\Gamma \text{ continuous; } \sup_{0\leqq t<t+h\leqq 1}\frac{\left|u(t+h)-u(t)\right|}{\delta(h)}\leqq 1\right\}$$

which is convex and compact in the sup-norm topology. The operator S defined by (4.1), (4.2) maps B into itself, and therefore has a fixed point $u^*\in B$ by the Schauder theorem. The measure-valued process $\mu^*=\{\mu_t^*\}$ given by

$$\mu_t^*(A)=\int_{R^r} E\left[1_{\{z+w_t\in A\}}\exp\left\{\int_0^t f(s,z+w_s,u^*(s))\cdot dw_s\right.\right.$$

$$\left.\left. -\frac{1}{2}\int_0^t \left|f(s,z+w_s,u^*(s))\right|^2 ds + \xi_0^t(-y\nabla h)+y_t h(t,z+w_t)\right.\right.$$

$$+\int_0^t l(s, z+w_s)\,ds\Bigg\}\Bigg]\cdot\mu_0(dz)$$

satisfies Eq. (2.5).

THEOREM 4.2 *Let us suppose that $v(t,m)$: $[0,1]\times\mathcal{M}^r\to\Gamma$ is Borel measurable, that $v(t,\cdot)$: $\mathcal{M}^r\to\Gamma$ is continuous for every $t\in[0,1]$, and that the function $f(t,z,u)$ satisfies Roxin's condition*

$$f(t,z,\Gamma) \text{ is a convex subset of } R^r, \text{ for any } (t,z)\in[0,1]\times R^r. \quad (4.5)$$

Then, for any $y\in C_{[0,1]}$ and $\mu_0\in\mathcal{M}^r$, there exists an element μ of C satisfying Eq. (2.5).

Proof With the class $\mathcal{U}$ as in Definition 2.1, consider the subset of C

$$\mathcal{S}=\Bigg\{\mu\in C \,\Big|\, \mu_t(A)=\int_{R^r} E\Bigg[1_{\{z+w_t\in A\}}\exp\Bigg\{\int_0^t f(s,z+w_s,u_s)\cdot dw_s$$

$$-\tfrac{1}{2}\int_0^t |f(s,z+w_s,u_s)|^2\,ds+\xi_0^t(-y\nabla h)+y_t h(t,z+w_t)$$

$$+\int_0^t l(s,z+w_s)\,ds\Bigg\}\Bigg]\mu_0(dz); \text{ for all } 0\leqq t\leqq 1, A\in\text{Borel}_r; \text{ some } u\in\mathcal{U}\Bigg\}.$$

We claim $\mathcal{S}$ is a compact, convex subset of C. Convexity of $\mathcal{S}$ amounts to convexity of the set of densities

$$\mathcal{D}=\Bigg\{\exp\Bigg[\int_0^t f(s,z+w_s,u_s)\cdot dw_s-\tfrac{1}{2}\int_0^t|f(s,z+w_s,u_s)|^2\,ds\Bigg]; u\in\mathcal{U}\Bigg\},$$

which is proved as in [1; Theorem 3] under condition (4.5). Notice that the latter does *not* guarantee convexity of the set of densities $\mathcal{D}'=\{\exp[\cdots]; u\in\mathcal{A}\}\subseteq\mathcal{D}$.

To prove compactness, we first notice that for each $0\leqq t\leqq 1$, the family of measures

$$K_t=\{\mu_t;\mu\in\mathcal{S}\}\subseteq\mathcal{M}^r$$

is uniformly tight, and so $\bar{K}_t$ is compact; this is proved as in Theorem 4.1. In the same fashion we check the equicontinuity of the family $\mathscr{S}$. By the Ascoli theorem [11; p. 155], $\mathscr{S}$ is then a compact subset to C.

Now the continuous operator $T: C \to C$ defined in Section 2 maps the set $\mathscr{S}$ into itself. By the Tychonoff fixed point theorem [14], T has a fixed point in $\mathscr{S}$. □

It is natural to inquire whether a (not necessarily unique) measure-valued process μ, a solution of Eq. (2.5) under either Theorem 4.1 or 4.2, can be "selected" in a fashion that is measurable with respect to the observation sample path y. Our next result addresses this question in the framework of Theorem 4.1.

THEOREM 4.3 *Measurable selection result.*

Under the assumptions of Theorem 4.1, fix $\mu_0 \in \mathscr{M}^r$ and a compact set $\mathscr{E} \subseteq C$. Then there exists a Borel-measurable function $\phi: \mathscr{E} \to C$, such that

$$\phi(y) = T(y, \phi(y)); \ \forall y \in \mathscr{E}.$$

Proof We notice first that the family of functions

$$\mathscr{K} = \{\mu(u) \in C \mid u: [0, 1] \to \Gamma \text{ Borel measurable}\}$$

is compact in C, since it is equicontinuous and for each $t \in [0, 1]$,

$$\mathscr{K}_t = \{\mu_t(u) \in \mathscr{M}^r \mid u: [0, 1] \to \Gamma \text{ Borel measurable}\}$$

is uniformly tight; see the proof of Theorem 4.1. The set

$$H = \{(y, \mu) \in \mathscr{E} \times \mathscr{K} \mid \mu = T(y, \mu)\}$$

is Borel in $\mathscr{E} \times \mathscr{K}$, and each section

$$H(y) = \{\mu \in \mathscr{H} \mid \mu = T(y, \mu)\}; \quad y \in \mathscr{E}$$

is compact. By Yankov's section Theorem ([6]; Corollary to Theorem 17) there exists a section of H by a Borel graph. □

Remark It should be apparent that in spite of the partial measurability results given above, there are still major gaps in our theoretical understanding of how solutions μ arise from the observations y. Further indications of the bad state of affairs are given in the final section, in which a pathological "lack of causality" appears.

5. COUNTEREXAMPLES

In conclusion, we present examples which show that pathwise uniqueness may fail if the control is allowed to depend in a suitable peculiar way on the past of the observations. The construction is based on an idea [13] of B. S. Tsirelson, used by him to disprove the existence of strong nonanticipative solutions to certain stochastic DEs, and elaborated [2] by one of the present authors.

The simplest example consists of the state and observation equations

$$dx_t = u_t\,dt + dw_t \quad \text{(state, with } x_0 = x \text{ given)}$$

$$dy_t = x_t\,dt + db_t \quad \text{(observation)}$$

with b, w independent Wiener processes, and u_t a certain functional of the past of the conditional distribution of x_t, given $\{y_s, 0 \leqq s \leqq t\}$. In this case the conditional distribution has the Gaussian density

$$\frac{1}{\sqrt{2\pi r}} \exp\left\{-\frac{(z-m)^2}{2r}\right\}$$

with m and r sufficient statistics satisfying

$$\dot{r} = 1 - r^2,\ r(0) = 0, \quad \text{so} \quad r(t) = \tanh t$$

$$dm = (u - mr)\,dt + r\,dy_t, m(0) = x.$$

These facts are either well-known or easily proved.

We can now obtain counterexamples by choosing u to depend in a perverse way on m itself; this is permitted by our ground rules, because the only conditions u must satisfy are that it depend only on past features of p and have its range in a specified set. The counterexamples to be given will show that solutions possessing certain reasonable properties need not always exist. Two such properties, often sought or expected for solutions of stochastic DEs, are

i) causality
ii) nonanticipation.

A *causal* solution m of the equation $dm = u(t,m)\,dt - rm\,dt + r\,dy_t$ is one expressible in terms of y by a causal functional, i.e.,

$$m_t = \phi(t, y) \text{ with}$$

$$y_1 = y_2 \quad \text{for} \quad s \leqq t \Rightarrow \phi(t, y_1) = \phi(t, y_2).$$

A *nonanticipative* solution m would be one whose past $\sigma\{m_s, s \leqq t\}$ is independent of the future increments of the forcing process y, i.e.,

$$\sigma\{m_s, s \leqq t\} \perp \sigma\{y_u - y_t, u > t\}.$$

We shall assume that u is restricted to lie in a bounded set. From this follows a Girsanov transformation of the probability space, which makes the observation process y_t a Wiener process, and we can assume without any further loss of generality that such a transformation has been carried out. The problem then is to see what kind of pathology can occur in solutions of

$$dm = u(t,m)\,dt - rm\,dt + r\,dy_t, \quad y_t \text{ Wiener.} \tag{5.1}$$

Note first that the "time-dependent low-pass filter" term rm can be essentially eliminated by defining, for $f \in C[0, \infty)$

$$(Rf)_t = f_t + \int_0^t r_s f_s\,ds.$$

The operator R is causally invertible; in terms of R, (5.1) can be written as

$$d(Rm)_t = u(t,m)\,dt + r\,dy_t.$$

We shall take $u(t,m)$ to have the form $\alpha(t,Rm)$. The next step is to perform a time change. It is well-known [7] that for

$$T(t)=\int_0^t r_s^2\,ds$$

the composition

$$\int_0^{T^{-1}(\tau)} r_s\,dy_s=W_\tau$$

is a Wiener process, whenever y is a Wiener process. Thus putting

$$\zeta_\tau=(Rm)_{T^{-1}(\tau)}$$

our equation takes on the form, using

$$\frac{dT^{-1}}{d\tau}=r^{-2}(T^{-1}(\tau)),$$

$$\zeta_\tau=(Rm)_{T^{-1}(\tau)}=\int_0^{T^{-1}(\tau)}\alpha(s,(Rm)_s)\,ds+x+W_\tau,\quad W\text{ Wiener}$$

$$=\int_0^\tau \alpha(T^{-1}(u),\zeta)r^{-2}(T^{-1}(u))\,du+x+W_\tau$$

or in differential form

$$d\zeta_\tau=\alpha(T^{-1}(\tau),\zeta)r^{-2}(T^{-1}(\tau))\,d\tau+dW_\tau.$$

Now we chose α so that the drift for the ζ-equation is Tsirelson's.

Let $\{t\}$ denote the fractional part of t, and let t_k, $k\leqq 0$ be distinct points in monotone sequence condensing at 0, with $t_0=1$. Tsirelson's drift β is defined for $f\in C[0,1]$ as

$$\beta(\tau,f)=\left\{\frac{f_{t_{k+1}}-f_{t_k}}{t_{k+1}-t_k}\right\}\quad\text{on}\quad(t_{k-1},t_k].$$

We choose

$$\alpha(t,f)=r_t^2\left\{\frac{f_{T(t_{k+1})}-f_{T(t_k)}}{t_{k+1}-t_k}\right\}\quad\text{on}\quad t_{k-1}\leqq T(t)<t_k$$

so that

$$a(T^{-1},\,f)r_{T^{-1}}^{-2}=\beta(\tau,\,f).$$

The equation $d\zeta_\tau=\beta(\tau,\zeta)\,d\tau+dW_\tau$ which results is exactly Tsirelson's, and has no causal solution, although by Girsanov's theorem it does have a unique nonanticipating solution measure. Since the time-change T and the filter R preserve both causality and the property of being nonanticipating, it follows that the conditional mean Eq. (5.1) for m has no causal solution for our choice of control law, while it does have a nonanticipating one. Thus a seemingly innocuous choice of control law can lead to loss of the kind of causal solution found in the first part of this paper. This fact, not really surprising in view of its arising for stochastic DEs, indicates the need for care in talking about "solutions" of the filtering problem when there is feedback from the conditional density.

Since $dy-m\,dt$ in principle defines the associated innovations process dv, one can similarly ask whether there are causal solutions for m in terms of v. Constructions like Tsirelson's again show that in general causal solutions do not exist.

Acknowledgement

The authors are grateful to B. L. Rozovsky and M. H. A. Davis for their comments and suggestions.

References

[1] V. E. Beneš, Existence of optimal stochastic control laws, *SIAM J. Control* **9** (1971), 446–472.

[2] V. E. Beneš, Nonexistence of strong nonanticipative solutions to stochastic DEs: implications for functional DEs, filtering, and control, *Stoch. Proc. Appl.* **5** (1977), 243–263.

[3] V. E. Beneš, Solving the Zakai equation by Itô's method, *Lecture Notes in Control and Information Sciences* **43** (1982), 156–161.

[4] V. E. Beneš and I. Karatzas, On the relation of Zakai's and Mortensen's equations, *SIAM J. Control and Optim.* **21** (1983), 472–489.

[5] M. H. A. Davis, Nonlinear semigroups in the control of partially observable stochastic systems, *Lecture Notes in Mathematics* **695** (1978), 37–49.
[6] C. Dellacherie, Ensembles analytiques: théorèmes de séparation et applications, *Lecture Notes in Mathematics* **465** (1975), 336–372.
[7] R. S. Liptser and A. N. Shiryaev, *Statistics of Random Processes, Vol I: General Theory*, Springer Verlag, Berlin, 1977.
[8] H. P. McKean, Jr., *Stochastic Integrals*, Academic Press, New York, 1969.
[9] R. E. Mortensen, Stochastic optimal control with noisy observations, *Int. J. Control* **4**, (1966), 455–464.
[10] K. R. Parthasarathy, *Probability Measures on Metric Spaces*, Academic Press, New York, 1967.
[11] H. L. Royden, *Real Analysis*, Macmillan, New York, 1963.
[12] A. V. Skorokhod, *Studies in the Theory of Random Processes*, Addison-Wesley, Reading, Mass. 1965.
[13] B. S. Tsirel'son, An example of a stochastic differential equation not possessing a strong solution, *Teoriya Veroiatnostei Primen.* **20** (1975), 427–430.
[14] A. Tychonoff, Ein Fixpunktsatz, *Math. Ann.* **111** (1935), 767–776.
[15] M. Zakai, On the optimal filtering of diffusion processes, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **11** (1969), 230–243.