The Annals of Probability
1986, Vol. 14, No. 2, 480–489

REVERSE TIME DIFFERENTIATION AND SMOOTHING FORMULAE FOR A FINITE STATE MARKOV PROCESS

BY ROBERT J. ELLIOTT

University of Hull

The paper investigates the reverse time differentiation of a stochastic exponential that occurs in smoothing, when the signal is a finite state Markov process and the observation process is a diffusion.

1. Introduction. In this paper we discuss the reverse time differentiation of a stochastic exponential expression that arises in the theories of filtering and smoothing when the signal process is a finite state Markov chain and the observation process is a diffusion. Section 2 contains a description of the filtering and smoothing problems in this situation and indicates why the reverse differential is of interest. The methods of this paper are inspired by those of Kunita [2] and [3]. In [2] Kunita discusses backward and forward stochastic differential equations, and in [3] uses these results to construct solutions of stochastic partial differential equations by the Feynman–Kac method. Anderson and Rhodes in [1] discuss smoothing formulae for various kinds of signal and observation processes and derive similar stochastic partial differential equations for the smoothed estimate. The methods of [1] require the reverse time differentiation of certain stochastic expressions, that is the differentiation with respect to the initial time and state; however, [1] does not include any discussion of backward stochastic integrals.

Pardoux obtains similar results for filtering, smoothing, and prediction formulae in [4]–[7] using existence results from partial differential equations, when the signal and observation process are diffusions. However, as Pardoux observes, Kunita's theory can be used to write down such solutions. By deriving the reverse time differential formula this paper enables similar solutions to be written down when the signal is a finite state Markov process and the observation is a diffusion.

2. Filtering and smoothing. We first define the Markov process, describe related filtering and smoothing equations, and indicate why reverse time differentiation of a related exponential is of interest. The author is greatly indebted to the referee for help in clarifying the definition of the Markov process.

Write $e_i = (0, 0, \ldots, 1, \ldots 0)'$ for the ith unit column vector in R^N. We shall define a Markov process on a probability space $(\Omega, \mathscr{F}, P)$, with a state space $S = \{e_1, \ldots, e_N\}$ and infinitesimal generator given by the matrix $A_t = (a_{ij}(t))$, $1 \le i,\ j \le N$. Write U for the totality of maps u: $S \to S$ and let $p = p(t)$ be a Poisson point process on U defined on $(\Omega, \mathscr{F}, P)$ such that the intensity measure

Received June 1984; revised April 1985.
AMS 1980 *subject classifications*. 93E14, 60H05, 93C10, 60H15, 60C35.
Key words and phrases. Filtering, smoothing, backward Itô and Stratonovich integral, stochastic partial differential equation.

$n_p(dt, du)$ of its counting measure $N_p(dt, du)$ is given by

$$n_p\big(dt, \{u: u(e_i) = e_j\}\big) = a_{ij}(t)\,dt, \qquad i \neq j.$$

Let D_p be the domain of the point process $p = p(t)$. Because the intensity measure is finite it is almost surely a discrete subset of $(0, \infty)$. For each $s > 0$ label the points of $D_p \cap (s, \infty)$ as $\{\sigma_n(s), n = 1, 2, \dots\}$ such that $s < \sigma_1(s) < \sigma_2(s) < \cdots$ and define a U valued process $X_{s,t}$ in the following way:

$$\begin{aligned} X_{s,t} &= \text{the identity map} \quad \text{if } s \leq t < \sigma_1(s) \\ &= p(\sigma_1(s)) \quad \text{if } \sigma_1(s) \leq t < \sigma_2(s) \\ &= p(\sigma_k(s)) \circ p(\sigma_{k-1}(s)) \dots \circ p(\sigma_1(s)) \quad \text{if } \sigma_k(s) \leq t < \sigma_{k+1}(s). \end{aligned}$$

Here $\circ$ denotes the composition of the mappings. Suppose $\Phi(t, s)$ is the transition matrix associated with A, so

$$\frac{d\Phi}{dt}(t, s) = A_t \Phi(t, s), \qquad \Phi(s, s) = I,$$

and

$$\frac{d\Phi}{ds}(t, s) = -\Phi(t, s) A_s, \qquad \Phi(t, t) = I,$$

where I is the $N \times N$ identity matrix. For each $x \in S$, $X_{s,t}(x)$ is an S-valued Markov process, and it follows from the above definition that $E[X_{s,t}(x)] = \Phi(t, s)x$, so in particular $P(X_{s,t}(e_i) = e_j) = e_j'\Phi(t, s)e_i$. Furthermore, for any $s < t < u$, $X_{t,u}(X_{s,t}(x)) = X_{s,u}(x)$ a.s., and for any times $t_1 < t_2 < \cdots < t_n$ the $X_{t_i, t_{i+1}}$ are independent; these two properties correspond to similar properties possessed by the stochastic flows generated by stochastic differential equations.

Write

$$\mathscr{G}_t^s = \sigma\{X_{s,u}(x): s \leq u \leq t\},$$

$$G_t = G_t^0, \qquad G^s = \bigvee_{t \geq s} G_t^s \quad \text{and} \quad G = G^0.$$

We suppose the finite state Markov process X cannot be observed directly. Instead there is a noisy observation process $\{y_t\}$, $t \geq 0$, where

$$dy_t = h(X_{s,t}(x), t)\,dt + \alpha(t)\,dw_t, \qquad y_0 = 0. \tag{2.1}$$

Here, for simplicity, we suppose y is one dimensional h is a bounded function with a bounded derivative in t, and w is a Brownian motion independent of X on the probability space $(\Omega, \mathscr{F}, P)$. We suppose α is a measurable function of t with a bounded inverse. For $0 \leq s \leq t$ write:

$$\mathscr{F}_t^s = \sigma\{y_v - y_u: s \leq u \leq v \leq t\}, \qquad \mathscr{F}_t = \mathscr{F}_t^0,$$

$$\mathscr{F}_t^s = \bigvee_{t \geq s} \mathscr{F}_t^s \quad \text{and} \quad \mathscr{F} = \mathscr{F}^0.$$

Consider the exponential

$$\Lambda_{s,t}(x) = \exp\left\{\int_s^t h(X_{s,u}(x), u)\alpha^{-1}(u)\, dy_u - \tfrac{1}{2}\int_s^t h(X_{s,u}(x), u)^2 \alpha^{-2}(u)\, du\right\}.$$

Write $\Lambda_t(x) = \Lambda_{0,t}(x)$, and define a new probability measure $\tilde{P}_{s,x}$ on $(\Omega, \mathscr{F})$ by putting

$$\left.\frac{d\tilde{P}_{s,x}}{dP}\right|_{\mathscr{G}^s \vee \mathscr{F}_t^s} = \Lambda_{s,t}^{-1}(x).$$

Write $\tilde{w}_t = \int_0^t \alpha^{-1}(u)\, dy_u$. Then under $\tilde{P}_{s,x}$, $X_{s,t}(x)$ remains a finite state Markov process with the same law and $\tilde{w}_t - \tilde{w}_s$ is a Brownian motion independent of X. E will denote expectation with respect to P and $\tilde{E}_{s,x}$ expectation with respect to $\tilde{P}_{s,x}$. Suppose F: $\{e_i\} \to R$ is any bounded function. The filtering problem discusses, [writing $X_{0,t}(x) = X_t$],

$$E[F(X_t)|\mathscr{F}_t].$$

Using Bayes' rule this can be written:

$$E[F(X_t)|\mathscr{F}_t] = \frac{\tilde{E}_{0,x}[\Lambda_t(x)F(X_t)|\mathscr{F}_t]}{\tilde{E}_{0,x}[\Lambda_t(x)|\mathscr{F}_t]},$$

so the investigation of $\tilde{E}_{0,x}[\Lambda_t(x)F(X_t)|\mathscr{F}_t]$ plays a central role in filtering.

The smoothing problem discusses

$$E[F(X_s)|\mathscr{F}_t] \quad \text{for } s \le t,$$

that is, we have observations through to some later time t than the time s at which we wish to estimate the state X. Again transforming to a measure $\tilde{P}_{0,x}$ we have

$$E[F(X_s)|\mathscr{F}_t] = \frac{\tilde{E}_{0,x}[F(X_s)\Lambda_t(x)|\mathscr{F}_t]}{\tilde{E}_{0,x}[\Lambda_t(x)|\mathscr{F}_t]}.$$

Similarly to Lemma 3.9 of [5] one can show:

$$\begin{aligned}\tilde{E}_{0,x}[F(X_s)\Lambda_t(x)|\mathscr{F}_t] &= \tilde{E}_{0,x}[F(X_s)\Lambda_s(x)\Lambda_t^s(X_s)|\mathscr{F}_t] \\ &= \tilde{E}_{0,x}[F(X_s)\Lambda_s(x)\Lambda_t^s(X_s)|\mathscr{F}_s \vee \mathscr{F}_t^s \vee \{X_s\}|\mathscr{F}_t] \\ &= \tilde{E}_{0,x}\left[F(X_s)\Lambda_s(x)\tilde{E}_{s,X_s}[\Lambda_t^s(X_s)|\mathscr{F}_t^s]|\mathscr{F}_t\right].\end{aligned}$$

Consequently in smoothing the investigation of $\tilde{E}_{s,X_s}[\Lambda_t^s(X_s)|\mathscr{F}_t^s]$ plays a central role. That is, for a fixed t and for $X_s = x \in \{e_i\}$ we must investigate

$$\tilde{E}_{s,x}[\Lambda_t^s(x)|\mathscr{F}_t^s] = \tilde{E}_{s,x}\left[\exp\left\{\int_s^t h(X_{s,u}(x), u)\, d\tilde{w}_u - \tfrac{1}{2}\int_s^t h(X_{s,u}(x), u)^2\, du\right\}\right],$$

where $\tilde{w}$ is a $\tilde{P}_{s,x}$ Brownian motion and $X_{s,u}(x)$ is a finite state Markov process independent of $\tilde{w}$. In the following section we determine how such an expression varies with s. In particular we discuss the differentiation in s of an exponential expression of this form and show it satisfies a backward stochastic differential equation.

3. Reverse time differentiation of an exponential. Consider, as above, a finite state Markov process defined on a probability space $(\Omega, \mathscr{F}, P)$ and with state space $e_1, \ldots, e_N$. Suppose the process is in state x at time $s \geq 0$, and again write $X_{s,t}(x)$ for its state at time $t \geq s$. Recall $\mathscr{G}_t^s(x) = \sigma\{X_{s,u}(x): s \leq u \leq t\}$.

Suppose $w_t = (w_t^1, \ldots, w_t^m)$, $t \geq 0$, is an m-dimensional Brownian motion, defined on $(\Omega, \mathscr{F}, P)$, which is independent of the Markov process X. Write, for $s \leq t$,

$$\mathscr{F}_t^s = \sigma\{w_v - w_u: s \leq u \leq v \leq t\}.$$

For $0 \leq j \leq m$ (as in the observation process of Section 2) suppose that h^j is a bounded real valued function on $\{e_i\} \times [\theta, \infty)$. Then h^j is represented as a row vector of functions $h^j(u) = (h^j(e_1, u), h^j(e_2, u), \ldots, h^j(e_N, u))$, and we further suppose that the derivative $dh^j(t)/dt$ exists and is a bounded function of the same form. Then h^j evaluated a state x and time $u \in [0, \infty]$ can be represented as $\langle h^j(u), x\rangle = h^j(x, u)$. In filtering and smoothing problems $h^0(x, u)$ is often of the form $-\frac{1}{2}\sum_{j=1}^m h^j(x, u)^2$.

Write $\omega_t^0 = t$. Then the theory of filtering and smoothing, when the signal process is the above Markov process and the observation process is a diffusion, leads one to consider an exponential expression of the form

$$\Lambda_{s,t}x = \exp\left\{\sum_{j=0}^m \int_s^t h^j(X_{s,u}(x), u)\, dw_u^j\right\}.$$

Because x and w are independent it is immaterial whether the integrals are interpreted as Itô or Stratonovich integrals. Write

$$\hat{\Lambda}_{s,t}(x) = E_P[\Lambda_{s,t}(x) | \mathscr{F}_t^s].$$

We have seen in Section 2 that the theory of smoothing requires one to investigate the derivative in s of $\hat{\Lambda}_{s,t}(x)$. Following the method of Kunita [2], this is the problem studied in the present section. We first state as lemmas two simple inequalities.

LEMMA 3.1. *For* $n = 1, 2, \ldots,$

$$E\left[\Lambda_{s,t}^{2n}(x)\right] \leq e^{C_n(t-s)}$$

for constants C_n.

PROOF. With $\circ$ denoting the Stratonovich integral

$$\begin{aligned}\Lambda_{s,t}(x) &= 1 + \sum_{j=0}^m \int_s^t \Lambda_{s,u}(x) h^j(X_{s,u}(x), u) \circ dw_u^j \\ &= 1 + \sum_{j=0}^m \int_s^t \Lambda_{s,u}(x) h^j(X_{s,u}(x), u)\, dw_u^j \\ &\quad + \frac{1}{2}\sum_{j=1}^m \int_s^t \Lambda_{s,u}(x) h^j(X_{s,u}(x), u)^2\, du.\end{aligned}$$

Also, writing $h^j(u)$ for $h^j(X_{s,u}(x), u)$,

$$\Lambda^2_L s, t(x) = 1 + 2\sum_{j=0}^{m} \int_s^t \Lambda^2_{s,u}(x) h^j(u)\, dw_u^j + \sum_{j=1}^{m} \int_s^t \Lambda^2_{s,u}(x) h^j(u)^2\, du.$$

Because the h^j are uniformly bounded

$$E\left[\Lambda^2_{s,t}(x)\right] \leq 1 + C_1 \int_s^t E\left[\Lambda^2_{s,u}(x)\right] du$$

and the result follows by Gronwall's inequality. Inequalities for higher powers are established the same way. □

LEMMA 3.2. *For any power M*

$$E\left[|X_{s,t}(x) - x|^M\right] \leq \text{Const.}|t - s|.$$

PROOF. Recall that the matrix A_t is assumed to be uniformly bounded and, if $\phi_{ij}(t, s)$ denotes the (i, j) entry of the transition matrix,

$$\left.\frac{\partial \phi_{ij}(t, s)}{\partial t}\right|_{t=s} = a_{ij}(t) \quad \text{if } i \neq j.$$

Therefore,

$$\begin{aligned} E\left[|X_{s,t}(x) - x|^M\right] &\leq 2^{M/2}\text{Prob}\{X_{s,t}(x) \neq x\} \\ &\leq \text{Const.}\left(\sum_{i \neq j,\, i,\, j=1}^{N} \phi_{ij}(t, s)\right) \\ &\leq \text{Const.}|t - s|. \quad \square \end{aligned}$$

REMARK 3.3. Recall that if $f(u)$, $0 \leq u \leq t$, is an $\mathscr{F}_t^{\,u}$ predictable process such that $\int_s^t E[f(u)^2]\, du < \infty$ then the backward Itô integral (see [2]), is defined, if $f(u)$ is continuous in probability, as

$$\int_s^t f(u)\, \hat{d}w_u^j := \lim_{|\Delta| \to 0} \sum_{k=0}^{n-1} f(t_{k+1})\left(w_{t_{k+1}}^j - w_{t_k}^j\right).$$

Here $\Delta = \{s = t_0 < t_1 < \cdots < t_n = t\}$ and $|\Delta| = \max_k |t_{k+1} - t_k|$.

It is a backward martingale, i.e., if $s' < s$

$$E\left[\int_{s'}^t f(u)\, \hat{d}w_u^j | \mathscr{F}_t^{\,s}\right] = \int_s^t f(u)\, \hat{d}w_u^j.$$

The backward Stratonovich integral is defined as

$$\int_s^t f(u) \circ \hat{d}w_u^j = \lim_{|\Delta| \to 0} \sum_{k=0}^{n-1} \tfrac{1}{2}\left(f(t_{k+1}) + f(t_k)\right)\left(w_{t_{k+1}}^j - w_{t_k}^j\right).$$

The two integrals are related by the formula

$$\int_s^t f(u) \circ \hat{d}w_u^j = \int_s^t f(u)\, \hat{d}w_u^j + \tfrac{1}{2}\left(\langle f, w^j \rangle_t - \langle f, w^j \rangle_s\right).$$

NOTATION 3.4. For $1 \le i \le N$ we shall write

$$H^i(s,t) = \exp\left\{ \sum_{j=0}^{m} \int_s^t h^j(X_{s,u}(e_i), u)\, dw_u^j \right\} = \Lambda_{s,t}(e_i),$$

and H for the row vector $(H^1, H^2, \dots, H^N)$. Then $\Lambda_{s,t}(x) = \langle H(s,t), x\rangle$ and, with

$$\hat{\Lambda}_{s,t}(x) = E_P[\Lambda_{s,t}(x) | \mathscr{F}_t^s]$$

and

$$\hat{H}_{s,t} = E_P[H(s,t) | \mathscr{F}_t^s]$$

we have

$$\hat{\Lambda}_{s,t}(x) = \langle \hat{H}(s,t), x\rangle.$$

THEOREM 3.5. *$\hat{\Lambda}_{s,t}(x)$ satisfies the following reverse time stochastic differential equation*:

$$\hat{\Lambda}_{s,t}(x) - 1 = \int_s^t \hat{\Lambda}_{u,t}(A_u x)\, du + \sum_{j=0}^{m} \int_s^t \hat{\Lambda}_{u,t}(x) h^j(x,u) \circ \hat{d}w_u^j.$$

PROOF. The method is an adaptation of that in Kunita's paper [2]. Consider a partition $\Delta = \{s = t_0 < t_1 < \cdots < t_n = t\}$ and write $|\Delta| = \max_k |t_{k+1} - t_k|$. Then

$$\Lambda_{s,t}(x) - 1 = \sum_{k=0}^{n-1} \left(\Lambda_{t_k,t}(x) - \Lambda_{t_{k+1},t}(x)\right).$$

Now

$$\Lambda_{t_k,t}(x) = \Lambda_{t_k,t_{k+1}}(x), \qquad \Lambda_{t_{k+1},t}\left(X_{t_k,t_{k+1}}(x)\right).$$

So the kth term in the above sum is

$$\Lambda_{t_k,t}(x) - \Lambda_{t_{k+1},t}(x) = \Lambda_{t_k,t_{k+1}}(x)\left(\Lambda_{t_{k+1},t}\left(X_{t_k,t_{k+1}}(x)\right) - \Lambda_{t_{k+1},t}(x)\right)$$
$$+ \Lambda_{t_{k+1},t}(x)\left(\Lambda_{t_k,t_{k+1}}(x) - 1\right) = J_k + K_k, \quad \text{say}.$$

Write $\hat{J}_k = E_P[J_k | \mathscr{F}_t^s]$, etc. Then we shall show that

$$\sum_{k=0}^{n-1} \hat{J}_k \to \int_s^t \hat{\Lambda}_{u,t}(A_u x)\, du \quad \text{as } |\Delta| \to 0,$$

and

$$\sum_{K=0}^{n-1} \hat{K}_k \to \sum_{j=0}^{m} \int_s^t \hat{\Lambda}_{u,t}(x) h^j(x,u) \circ d\omega_u^j \quad \text{as } |\Delta| \to 0.$$

Now

$$\mathcal{J}_k = \left(\Lambda_{t_k, t_{k+1}}(x) - 1\right)\left(\Lambda_t t_{k+1}, t\left(X_{t_k, t_{k+1}}(x)\right) - \Lambda_{t_{k+1}, t}(x)\right)$$
$$+ \left(\Lambda_{t_{k+1}, t}\left(X_{t_k, t_{k+1}}(x)\right) - \Lambda_{t_{k+1}, t}(x)\right) = \mathcal{J}_k^{(1)} + \mathcal{J}_k^{(2)}, \quad \text{say.}$$

We can write

$$\mathcal{J}_k^{(2)} = \left\langle H(t_{k+1}t), X_{t_k, t_{k+1}}(x) - x\right\rangle.$$

So

$$\hat{\mathcal{J}}_k^{(2)} = \left\langle \hat{H}(t_{k+1}, t), (\Phi(t_k, t_{k+1}) - I)x\right\rangle$$
$$= \left\langle H(t_{k+1}, t), \int_{t_k}^{t_{k+1}} A_u \Phi(u, t_k)\, du\, x\right\rangle.$$

Therefore,

$$\sum_{k=0}^{n-1} \hat{\mathcal{J}}_k^{(2)} \to_{|\Delta| \to 0} \int_s^t \langle \hat{H}(u, t), A_u x\rangle\, du.$$

Consider

$$\mathcal{J}_k^{(1)} = \sum_{j=0}^{m} \left(\int_{t_k}^{t_{k+1}} \Lambda_{t_k, u}(x) h^j\left(X_{t_k, u}(x), u\right) \circ dw_u^j\right)\left\langle H(t_{k+1}, t), X_{t_k, t_{k+1}}(x) - x\right\rangle.$$

Then, using Schwarz's inequality and Lemma 3.1, for an $\alpha > 0$

$$E|\mathcal{J}_k^{(1)}| \leq \text{Const.}|t_{k+1} - t_k|^{\alpha+1}.$$

Therefore,

$$E\left|\sum_{k=0}^{n-1} \hat{\mathcal{J}}_k^{(1)}\right| \leq E\left|\sum_{k=0}^{n-1} \mathcal{J}_k^{(1)}\right| \leq \sum_{k=0}^{n-1} E|\mathcal{J}_k^{(1)}| \leq \text{Const.}|\Delta|^\alpha (t - s)$$

and so $\sum_{k=0}^{n-1} \hat{\mathcal{J}}_k^{(1)} \to 0$ as $|\Delta| \to 0$.

$$K_k = \Lambda_{t_{k+1}, t}(x)\left(\exp\left(\sum_{j=0}^{m} \int_{t_k}^{t_{k+1}} h^j\left(X_{t_k, u}(x), u\right) dw_u^j\right) - 1\right)$$

and by the mean value theorem this can be written:

$$K_k = \Lambda_{t_{k+1}, t}(x)\left(\sum_{j=0}^{m} \int_{t_k}^{t_{k+1}} h^j(u)\, dw_u^j \right.$$
$$\left. + \tfrac{1}{2}\Lambda_{t_k, \sigma}(x) \sum_{i, j=0}^{m} \left(\int_{t_k}^{t_{k+1}} h^j(u)\, dw_u^j\right)\left(\int_{t_k}^{t_{k+1}} h^i(u)\, dw_u^i\right)\right).$$

Here $t_k \leq \sigma \leq t_{k+1}$ and, as above, $h^j(u) = h^j(X_{t_k, u}(x), u)$.

Then

$$K_k = \sum_{p=1}^{8} K_k^{(p)},$$

where

$$K_k^{(1)} = \Lambda_{t_{k+1},t}(x) \sum_{j=0}^{m} h^j(x, t_{k+1})\left(w_{t_{k+1}}^j - w_{t_k}^j\right),$$

$$K_k^{(2)} = \Lambda_{t_{k+1},t}(x) \sum_{j=0}^{m} \left(\int_{t_k}^{t_{k+1}} h^j\left(X_{t_k,u}(x), u\right) - h^j(x, u)\right) dw_u^j,$$

$$K_k^{(3)} = \Lambda_{t_{k+1},t}(x) \sum_{j=0}^{m} \left(\int_{t_k}^{t_{k+1}} (h^j(x, u) - h^j(x, t_{k+1}))\, dw_u^j\right),$$

$$K_k^{(4)} = \Lambda_{t_{k+1},t}(x)\Lambda_{t_k,\sigma}(x)\left(\tfrac{1}{2}\left(\int_{t_k}^{t_{k+1}} h^0(u)\, du\right)^2 + \left(\int_{t_k h^0}^{t_{k+1}}(u)\, du\right)\left(\sum_{j=1}^{m} \int_{t_k}^{t_{k+1}} h^j(u)\, dw_u^j\right)\right),$$

$$K_k^{(5)} = \tfrac{1}{2}\Lambda_{t_{k+1},t}(x)\left(\Lambda_{t_k,\sigma}(x) - 1\right) \sum_{i,j=1}^{m} \left(\int_{t_k}^{t_{k+1}} h^j(u)\, dw_u^j\right)\left(\int_{t_k}^{t_{k+1}} h^i(u)\, dw_u^i\right),$$

$$K_k^{(6)} = \tfrac{1}{2}\Lambda_{t_{k+1},t}(x)\left(\sum_{i,j=1}^{m} \int_{t_k}^{t_{k+1}}\left(h^j(u)\, dw_u^j\right)\left(\int_{t_k}^{t_{k+1}} h^i(u)\, dw_u^i\right) - \sum_{j=1}^{m} \int_{t_k}^{t_{k+1}} h^j(u)^2\, du\right),$$

$$K_k^{(7)} = \tfrac{1}{2}\Lambda_{t_{k+1},t}(x)\left(\sum_{j=1}^{m} \int_{t_k}^{t_{k+1}}\left(h^j(u)^2 - h^j(x, t_{k+1})^2\right) du,\right.$$

$$K_k^{(8)} = \tfrac{1}{2}\Lambda_{t_{k+1},t}(x)\left(\sum_{j=1}^{m} h^j(x, t_{k+1})^2\right)(t_{k+1} - t_k).$$

Now

$$\sum_{k=0}^{n-1} \hat{K}_k^{(1)} \to_{|\Delta| \to 0} \sum_{j=0}^{m} \int_s^t \hat{\Lambda}_{u,t}(x) h^j(x, u)\, \hat{d}w_u^j$$

and

$$\sum_{k=0}^{n-1} \hat{K}_k^{(8)} \to_{|\Delta| \to 0} \tfrac{1}{2} \sum_{j=1}^{m} \int_s^t \hat{\Lambda}_{u,t}(x) h^j(x, u)^2\, du.$$

In discussing $\Sigma K_k^{(p)}$ for $2 \leq p \leq 7$ two principles are used. Firstly that a sum of the form $\Sigma\text{Const.}|t_{k+1} - t_k|^{\alpha+1}$ is less than $\text{Const.}(t - s)|\Delta|^{\alpha}$, and so tends to 0 as $|\Delta| \to 0$. For example, for $p = 4,5,7$, $E[|K_k^{(p)}|] \leq \text{Const.}|t_{k+1} - t_k|^{\alpha+1}$. Therefore,

$$E\left|\sum_{k=0}^{n-1} \hat{K}_k^{(p)}\right| \leq E\left|\sum_{k=0}^{n-1} K_k^{(p)}\right|$$

$$\leq E\sum_{k=0}^{n-1} |K_k^{(p)}| \leq \text{Const.}(t-s)|\Delta|^{\alpha},$$

and so $\sum_{k=0}^{n-1}\hat{K}_k^{(p)} \to_{|\Delta|\to 0} 0$ for $p = 4,5,7$. Secondly, for $p = 2,3,6$, $M_m^{(p)} = \sum_{k=m}^{n-1} K_k^{(p)}$ is a reverse time $\mathscr{F}_t^{t_m} \vee \mathscr{G}$ martingale. Therefore, given $\mathscr{G}$, its increments are orthogonal and

$$E\left[\left(M_0^{(p)}\right)^2|\mathscr{G}\right] = \sum_{k=0}^{n-1} E\left[\left(K_k^{(p)2}|\mathscr{G}\right]\right..$$

So

$$E\left[\left(\sum_{k=0}^{n-1} \hat{K}_k^{(p)}\right)^2\right] \leq E\left[\left(M_0^{(p)}\right)^2\right] \leq \sum_{k=0}^{n-1} E\left[\left(K_k^{(p)}\right)^2\right].$$

Consider for example $p = 2$. Then

$$E\left(K_k^{(2)}\right)^2 \leq \text{Const.}\left(E\left(\sum_{j=0}^{m} \int_{t_k}^{t_{k+1}}\left(h^j(u) - h^j(x,u)\right) dw_u^j\right)^4\right)^{1/2}.$$

However, using Lemma 3.2

$$E\left|h^j(u) - h^j(x,u)\right|^M = E\left(\left\langle h^j, X_{t_k,u}(x) - x\right\rangle\right)^M \leq \text{Const.}(u - t_k)$$

and we see

$$E\left(K_k^{(2)}\right)^2 \leq \text{Const.}\left(t_{k+1} - t_k\right)^{\alpha+1},$$

where the constant is independent of k. Therefore, by the first principle above,

$$E\left[\left(\sum_{k=0}^{n-1} K_k^{(p)}\right)^2\right] \to_{|\Delta|\to 0} 0.$$

$(\sum K_k^{(3)})^2$ (using the uniform bound for the u-derivatives of the h^j) and $(\sum \hat{K}_k^{(6)})^2$ are treated similarly. Consequently, $\sum_{k=0}^{n-1}\hat{K}_k^{(p)} \to_{|\Delta|\to 0} 0$ for $p = 2,3,6$ and we have shown that

$$\sum_{k=0}^{n-1} \hat{K}_k \to_{|\Delta|\to 0} \sum_{j=0}^{m} \int_s^t \hat{\Lambda}_{u,t}(x) h^j(x,u)\, \hat{d}\omega_u^j$$

$$+ \tfrac{1}{2}\sum_{j=1}^{m} \int_s^t \hat{\Lambda}_{u,t}(x) h^j(x,u)^2\, du.$$

The stochastic integral here is a backward Itô integral; using the backward Stratonovich integral the above expression is just

$$\sum_{j=0}^{m} \int_s^t \hat{\Lambda}_{u,t}(x) h^j(x,u) \circ \hat{d}\omega_u^j$$

and the theorem is proved. □

REMARKS 3.6. Recall that a function on the state space $\{e_i\}$ of the Markov process is just represented by a vector $f \in R^N$, so $f(x) = \langle f, x \rangle$. Write, for $1 \le i \le N$,

$$F_{s,t}^i = f(X_{s,t}(e_i))\Lambda_{s,t}(e_i), \qquad \hat{F}_{s,t}^i = E_P[F_{s,t}^i | \mathscr{F}_t^s]$$

and $\hat{F}$ for the vector $(\hat{F}^1, \dots, \hat{F}^N)$. With $\hat{F}(x) = \langle \hat{F}, x \rangle$ similar techniques establish the following result:

THEOREM 3.7. *$\hat{F}_{s,t}(x)$ satisfies the following reverse time stochastic differential equation*:

$$\hat{F}_{s,t}(x) = f(x) + \int_s^t \hat{F}_{u,t}(A_u x)\,du + \sum_{j=0}^{m} \int_s^t \hat{F}_{u,t}(x) h^j(x,u) \circ \hat{d}w_u^j.$$

Acknowledgment. The author wishes to thank the Institut des Hautes Etudes Scientifiques, Bures-sur-Yvette, for its kind hospitality in April, 1984, when he was able to work on this paper in particularly pleasant conditions.

REFERENCES

[1] ANDERSON, B. D. O. and RHODES, I. B. (1983). Smoothing algorithms for nonlinear finite dimensional systems. *Stochastics* **9** 139–165.

[2] KUNITA, H. (1982). On backward stochastic differential equations. *Stochastics* **6** 293–313.

[3] KUNITA, H. (1983). Stochastic partial differential equations connected with nonlinear filtering. *Lecture Notes in Math.* **972** 100–169.

[4] PARDOUX, E. (1979). Stochastic partial differential equations and filtering of diffusion processes. *Stochastics* **3** 127–167.

[5] PARDOUX, E. (1982). Equations du filtrage non linéaire, de la prediction et du lissage. *Stochastics* **6** 193–231.

[6] PARDOUX, E. (1983). Equations of non-linear filtering, and applications to stochastic control with partial observation. *Lecture Notes in Math.* **972** 208–248.

[7] PARDOUX, E. (1984). Equations du lissage non linéaire. *Lecture Notes in Control and Information Sciences* **61** 206–218.

DEPTARTMENT OF STATISTICS
UNIVERSITY OF ALBERTA
EDMONTON, ALBERTA,
CANADA T6G 2G1