Abstract

The ‘pathwise’ theory of filtering is concerned with casting the filtering equations in a form in which the filtered estimates can be computed separately for each sample path of the observation process. This was originally achieved in cases where the observation noise is independent of the signal process. This paper presents a pathwise theory for the case where the signal is a diffusion on a finite-dimensional manifold and there is correlation with the observation noise. A geometric setting is natural for this problem, which also brings in Kunita’s decomposition theorem for solutions of SDEs and a family of observation-dependent multiplicative functionals of the signal process.

1 Introduction

Aside from minor revision and slight improvements, this chapter is a reprint of the author’s paper Davis (1981), which appeared in a now unavailable conference proceedings volume (Hazewinkel and Willems, 1981). The editors of the present volume have been kind enough to include it here because the ‘pathwise filtering’ approach continues to be of both theoretical and practical importance. In 1978, J.M.C. Clark made the key observation that without some continuity of the filter estimate with respect to the observation sample path, the nonlinear filtering equations are of no practical use; the argument is summarized in §2 below. He showed how the required continuity could be obtained by a simple integration by parts, re-expressing the filtering equations in a form in which no stochastic integrals with respect to the observation process appear. The filtering equations then make sense for any observation sample path that is merely a continuous function, and continuity can be established. The technical details of this argument were completed in Clark and Crisan (2005). They cover the case of multidimensional observations when the observation noise is independent of the signal process, due to a fortunate commutativity property of certain operators appearing in the problem. The purpose of this paper is to analyse the case where there is correlation between the signal and observation noise. Then commutativity is lost, and a pathwise theory is only possible for scalar observations. In deriving it we are inevitably led to a geometric approach which is perhaps of some independent interest—apart from the useful by-product of showing that the filtering equations apply equally well to signals evolving on finite-dimensional manifolds.

The paper is laid out as follows. The next section gives some background on the filtering equations and the need for a pathwise approach. §3 describes the coordinate-free representation of the signal SDE and gives some Lie-algebraic calculations that will be needed.
later. §4 covers the case of independent signal and observation noise and establishes the essential connection between the Kallianpur-Striebel (KS) formula and observation-dependent multiplicative functional transformations of the signal process semigroup. This is the basis for the pathwise approach. The final section, §5, addresses the correlated noise case using an SDE decomposition technique due to Kunita (1981). The main results of the paper are the formulas (5.11), giving the correct formulation of the KS formula in this context, and (5.14),(5.15) expressing the KS formula as an observation-dependent multiplicative functional transformation.

2 Background and summary

The setting is the conventional nonlinear filtering problem of calculating recursively estimates

$$
\mathbb{E}[f(X_t)|Y_s, 0 \leq s \leq t]
$$

where $X_t$ is a Markov process and $Y_t$ is a real-valued ‘observation process’ given by

$$
dY_t = h(X_t)dt + dW_t^0
$$

Here $h$ is a bounded function (additional smoothness assumptions will be imposed later) and $W_t^0$ is a standard Brownian motion. The introductory article in this volume can be consulted for general background and most of the standard results in filtering theory used below. Other references are Bain and Crisan (2008) and the brief introduction Davis and Marcus (1981).

Let us denote $Y_t = \sigma\{Y_s, s \leq t\}$ and

$$
\pi_t(f) = \mathbb{E}[f(X_t)|Y_t]
$$

The process $\pi_t$ should be thought of as the conditional distribution of $X_t$ given $Y_t$, so that

$$
\pi_t(f) = \int_{\mathcal{S}} f(x)\pi_t(dx).
$$

Here $\mathcal{S}$ is the state space for $X_t$. It is convenient to calculate an unnormalized form $\rho_t$ of this, $\pi_t$ then being given by

$$
\pi_t(f) = \frac{\rho_t(f)}{\rho_t(1)}.
$$

If $X_t$ and $W_t^0$ are independent, an appropriate unnormalized distribution can be obtained in two alternative forms:

(i) the Kallianpur-Striebel formula, giving $\rho_t$ non-recursively as a function-space integral:

$$
\rho_t(f) = \int_{\Xi} f(X_t)\exp\left(\int_0^t h(X_s)dY_s - \frac{1}{2}\int_0^t h^2(X_s)ds\right)\nu(dx) \quad (2.3)
$$

(Here $(\Xi,\nu)$ is the sample path probability space.)

(ii) the Zakai equation, giving $\rho_t$ in recursive form as the solution of a measure-valued stochastic differential equation:

$$
\begin{align*}
\frac{d\rho_t(f)}{dt} &= \rho_t(Af)dt + \rho_t(hf)dY_t \\
\sigma_0(f) &= \pi(f)
\end{align*} \quad (2.4)
$$

($A$ is the differential generator of the $X_t$ process, $hf(x) = h(x)f(x)$ and $\pi$ is the distribution of $X_0$).

\footnote{It is of course not obvious \textit{a priori} that a regular conditional distribution exists.}
Clark’s (1978) argument about the need for a pathwise approach and a continuity result is as follows. It has to do with questions of stochastic modelling. First, recall some facts about the conditional expectation (2.2). For this discussion, fix a time \( t > 0 \). Since \( Y_t \) generated by \( Y = \{ Y_s, 0 \leq s \leq t \} \), \( \pi_t(f) \) is a functional of the continuous process \( Y \), i.e. there is a measurable function \( \phi : C[0, t] \to \mathbb{R} \) such that

\[
\mathbb{E}[f(X_t)|Y_t] = \phi(Y) \quad \text{a.s.} \tag{2.5}
\]

\( \phi \) is not uniquely defined, in that any other function \( \phi' \) such that \( \phi'(Y) = \phi(Y) \) a.s. would be an equally good version of the conditional expectation. Here ‘a.s.’ refers to the distribution of \( Y \) on \( C[0, t] \), and this distribution has the same null sets as Wiener measure. In particular the set of functions with bounded variation is a null set. Now in the observation equation (2.1), \( Y \) is a mathematical model for \( \tilde{Y}_t = \int_0^t Z_s ds \) where \( Z_t \) is a physical observation

\[
Z_t = h(X_t) + N_t
\]

and \( N_t \) is (physical) ‘wide band’ noise. As an estimate for \( f(X_t) \) we then plan to take \( \phi(\tilde{Y}) \). But since \( \tilde{Y} \) has bounded variation, \( \phi \) is undefined for \( \tilde{Y} \), and indeed on the whole set of physical sample paths. Thus nonlinear filtering theory cannot be applied in practice unless we are able to choose a particular version of the conditional expectation which has ‘nice’ properties. Specifically, what is required is a function \( \phi : C[0, t] \to \mathbb{R} \) such that

\[\begin{align*}
(\text{i}) & \quad (2.5) \text{ holds.} \\
(\text{ii}) & \quad \phi \text{ is continuous with respect to the supremum norm on } C[0, t].
\end{align*}\]

Then \( \phi(\tilde{Y}) \) is a ‘sensible’ estimator, in that the mean square error \( \mathbb{E}[(f(X_t) - \phi(\tilde{Y}))^2] \) is close to the predicted value \( \mathbb{E}[(f(X_t) - \phi(Y))^2] \) as long as the distributions of \( \tilde{Y} \) are close to those of \( Y \) in the sense of weak convergence, and this certainly includes all the usual bounded-variation approximations to Bownian motion.

In Davis (1980, 1982) it was shown that such ‘robust’ filtering algorithms could be produced for a very wide class of Markov signal processes \( X_t \), when the signal and observation noise \( W_t^Y \) are independent. The main purpose of this paper is to extend the results to certain cases where there is correlation between signal and observation noise. This cannot be done at the same level of generality as in Davis (1980), and the signals we consider are diffusions on finite-dimensional manifolds. Such signals appear in important areas of application of nonlinear filtering theory, for example in alignment problems in inertial navigation where the signal is an orientation, represented by a quaternion vector. Also, the coordinate-free signal description (introduced in §3 below) adds insight even for \( \mathbb{R}^d \)-valued diffusions.

The basis of our approach is that the unnormalized conditional distribution \( \rho_t \) can be represented in the form

\[\rho_t(f) = < T_{t,0}^Y U_t f, \pi > \]

where \( T_{t,0}^Y \) is the \( Y \)-dependent semigroup associated with a certain multiplicative functional transformation, \( U_t \) is a group of operators and \( \pi \) is the distribution of the initial state \( X_0 \). A recursive form of estimator can then be obtained by considering the forward equation corresponding to \( T_{t,s}^Y \) (see §4.3 below). Our main concern is therefore to calculate the generator of \( T_{t,s}^Y \) and this is most readily done by factoring the relevant multiplicative functional (§4.2).

The relation with ‘pathwise solutions’ of the Zakai equation is explored in §4.4.

In the case of independent signal and noise, \( U_t \) is the operator of ‘multiplication by \( \exp(t h(x)) \)’. Our main result is that, for the type of noise correlation considered in §5, the
situation is formally analogous to the independent case, but with $U_t$ now being the flow corresponding to a certain differential operator (see (4.12)). Showing this involves decomposing the signal equation in the way described by Kunita (1981) in order to elucidate precisely the dependence of $X_t$ on the observations $y$.

It must, regretfully, be pointed out that the results for correlated noise cannot, unlike those for the independence case, be extended to vector observations. This is because the corresponding operators $U_t^i$ do not in general commute whereas with no noise correlation they are multiplication operators which automatically commute.

### 3 The signal process

The formulation here follows that of Kunita (1981). The signal process $X_t$ evolves on a $\sigma$-compact, connected $C^\infty$ manifold $S$ of dimension $d$. Suppose $V_0, \ldots, V_r$ are $C^\infty$ vector fields on $S$ and $W^1, \ldots, W^r$ are independent scalar Brownian motions which are independent of $W^0_t$ of equation (2.1). Then $X_t$ is the solution of the stochastic differential equation

$$dX_t = V_0(X_t)dt + V_j(X_t) \circ dW^j_t. \quad (3.1)$$

This means that $X_t$ is the unique $S$-valued process satisfying

$$f(X_t) = f(X_0) + \int_0^t V_0 f(X_s)ds + \int_0^t V_j f(X_s) \circ dW^j_s, \quad 0 \leq t \leq \tau$$

for any real valued $C^\infty$ function $f$. In these equations the $\circ$ denotes the Stratonovich stochastic integral, and the convention of implied summation from $j = 1$ to $r$ is used\(^2\). The initial point $X_0$ is supposed to be a random variable with a given distribution $\pi$, independent of all other r.v.’s. In (3.1), $\tau$ is the lifetime and we assume conditions are such that $\tau = \infty$ a.s. (automatically true if $S$ is compact). The relation of (3.1) with the corresponding Ito equation is the following: for any $k$, $V_k f \in C^\infty(S)$ and hence from (3.1)

$$dV_k f(X_t) = V_0 V_k f(X_t)dt + V_j V_k f(X_t) \circ dW^j_t.$$  

Thus the joint quadratic variation of the semimartingales $V_k f(X_t)$ and $W^k_t$ is

$$d < V_k f, W^k >_t = V_k^2 f(X_t)dt$$

and the Ito version of (3.1) is therefore

$$df(X_t) = (V_0 + \frac{1}{2} \sum_j V_j^2)f(X_t)dt + V_j f(X_t)dW^j_t. \quad (3.2)$$

The process $X_t$ is a Markov diffusion process; its generator $A$ is an operator acting on $C^\infty$ functions such that the Dynkin formula\(^3\)

$$E_x f(X_t) - f(x) = E_x \int_0^t Af(X_s)ds$$

is satisfied for $f \in C^\infty_0(S)$ (the $C^\infty$ functions of compact support). In view of the time-homogeneity this is equivalent to saying that the process

$$C^f_t := f(X_t) - f(x) - \int_0^t Af(X_s)ds$$

\[^2\] and similarly sums over other repeated indices $k$ etc. All sums are over 1 to $r$ unless otherwise specified.

\[^3\] $E_x$ is the expectation starting at $X_0 = x$. 

---

**Pathwise nonlinear filtering**
Pathwise nonlinear filtering

is a martingale. Now in (3.2) the stochastic integral is a martingale for \( f \in C_0^\infty(S) \) since \( V_j f \) is then bounded, and it follows that

\[
A = V_0 + \frac{1}{2} \sum_j V_j^2.
\]

\( A \) is sometimes called the \textit{extended generator} of \( X_t \): it is an extension of the infinitesimal generator of the semigroup of operators on \( C(S) \)

\[
T_t f(x) := E_x[f(X_t)]
\]

(3.3)

associated with the process \( X_t \).

For the sequel, we shall need to compute some Lie brackets. We suppose that the observation function \( h \) of (2.1) is in \( C_2^b(S) \). \( h \) will also denote the zeroth-order operator of ‘multiplication by \( h \)’, i.e. for \( f \in C^\infty(S) \)

\[
h f(x) = h(x) f(x)
\]

and similarly for other functions below. If \( D \) is any differential operator, \( \text{ad}_h D \) denotes the Lie derivative

\[
(\text{ad}_h D)f(x) = [h, D]f(x) = h(x)Df(x) - D(h f)(x)
\]

and \( \text{ad}_h^2 D = \text{ad}_h (\text{ad}_h D) \), etc. If \( V \) is a vector field then we find using the Leibnitz rule that

\[
\begin{align*}
\text{ad}_h V & = -(Vh) \\
\text{ad}_h^2 V & = 0 \\
\text{ad}_h^3 V & = -(Vh) \quad \text{ad}_h^2 V = -(Vh)V - V^2 h \\
\text{ad}_h^2 V^2 & = 2(Vh)^2 \\
\text{ad}_h^3 V & = 0.
\end{align*}
\]

Thus in particular

\[
\begin{align*}
\text{ad}_h A & = -(V_j h)V_j - Ah \\
\text{ad}_h^2 A & = \sum_j (V_j h)^2 \\
\text{ad}_h^k A & = 0, \quad k > 2.
\end{align*}
\]

4 Independent signal and noise

We consider the filtering problem over a fixed finite interval \([0, T]\). Recall that the observation equation (2.1) is

\[
Y_t = \int_0^t h(X_s)ds + W_t^0, \quad 0 \leq t \leq T.
\]

We will assume henceforth that

\[
h \in C_0^\infty(\mathbb{R}^r).
\]

(4.1)

\textit{Notation:} \( Y \) will denote the \( C[0, T]\)-valued random variable \( \{Y_t, t \in [0, T]\} \), so that \( Y_t \) is the value of \( Y \) at \( t \). We will use the notation \( y \) for an arbitrary but fixed element of \( C[0, T] \), with \( y(t) \) denoting the value at time \( t \) (in accordance with the traditions of real analysis). \( y \) may or may not be a sample function \( Y(\omega) \).
4.1 The KS formula as a multiplicative functional transformation

Let us return to the Kallianpur-Striebel formula, (2.3). In view of (3.1), (4.1) the real-valued process $h(X_t)$ is certainly a semimartingale, and we can write

$$
\int_0^t h(X_s)dY_s = h(X_t)Y_t - \int_0^t Y_s dh(X_s).
$$

Note that the right hand side of this equality involves no stochastic integration with respect to $dY$ and makes sense if $Y$ is replaced by any function $y \in C([0,T])$. Thus (2.3) can be written in the form

$$
\rho_t(f) = \mathbb{E} \left[ f(X_t)e^{\alpha_t^f(h(X_1)}\right]_{y=Y}
$$

(4.2)

where $\alpha_t^f(y)$ is defined for $s \leq t$ by

$$
\alpha_t^f(y) = \exp \left( -\int_s^t y(u)dh(X_u) - \frac{1}{2}\int_s^t h^2(X_u)du \right).
$$

(4.3)

In (4.2) the expectation is taken over the distribution of $X$ (regarded as a random variable taking values in $C([0,T];\mathbb{R}^r)$). If we now define $\phi(y) = \rho_t(f)/\rho_t(1)$ where $\rho_t(f)$ is given by (4.2), then this is the desired version of the conditional expectation, in that the two conditions of (2.6) hold (Clark and Crisan, 2005). It remains to show how $\rho_t(f)$ can be computed recursively. Associated with the process $X_t$ is a semigroup $(T_t)_{t \geq 0}$ of operators on $C(\mathbb{S})$ defined by (3.3) above. Now the process $\alpha_t^f(y)$ of (4.3) is a multiplicative functional of $X_t$, i.e. an adapted process satisfying

$$
\alpha_t^r = \alpha_t^s \alpha_t^s
$$

for $r \leq s \leq t$.

It is easily checked that, if we define

$$
T_{s,t}^y f(x) = \mathbb{E}_{s,t}[f(X_t)\alpha_t^s f(X_t)]
$$

(4.4)

then $T_{s,t}^y$ is another (two-parameter) semigroup of operators on $C(\mathbb{S})$. Note, however, that it is not Markovian, i.e. does not satisfy $T_{s,t}^0 1 = 1$. For $f \in C(\mathbb{S})$ and $\mu$ a measure on $\mathbb{S}$, denote $< f, \mu > = \int_{\mathbb{S}} f(x)\mu(dx)$. Then from (3.1) - (3.3) we see that

$$
\rho_t(f) = < T_{0,t}^y e^{\alpha_t^f(h)}f, \pi >.
$$

This provides us, in principle, with a recursive way of computing $\rho_t$. Let $U_{t,s}^y$ be the adjoint semigroup to $T_{s,t}^y$, defined by

$$
<T_{s,t}^y f, \mu > = < f, U_{t,s}^y \mu >
$$

and define

$$
\pi_t^y = U_{t,0}^y \pi.
$$

Then, formally, $\pi_t^y$ is the solution of the forward (Fokker-Planck) equation

$$
\frac{d}{dt} \pi_t^y = (A_t^y)^* \pi_t^y, \quad \pi_0^y = \pi.
$$

(4.5)

Here $A_t^y$ is the differential generator of $T_{s,t}^y$. Now (4.5) is a recursive equation for $\pi_t^y$, and $\rho_t$ is given by

$$
\rho_t(f) = < e^{\alpha_t^f(h)}f, \pi_t^y >.
$$

(4.6)

Notice that (4.5), (4.6) constitute a recursive filter in a form in which no stochastic integration is involved. The forward equation (4.5) has been investigated in detail for the case $\mathbb{S} = \mathbb{R}^d$ in Pardoux (1979) and Pardoux (1982). The remainder of this section is devoted to explicit calculation of the generator $A_t^y$ of the semigroup $T_{s,t}^y$. 

Pathwise nonlinear filtering
4.2 Factorization of multiplicative functionals

This section follows up some ideas contained in a paper of Mitter (1979). We introduce three simple types of multiplicative functional (MF), all relative to the Markov process $X_t$, called the Gauge, Feynman-Kac and Girsanov types, and explore the relations between them. Further general information on MFs can be found in the books of Blumenthal and Getoor (1968) and Dynkin (1965). The MF $\alpha_t^y(y)$ of (4.3) is of course time-varying in that it depends on the sample path $y$, but here we shall discuss time-invariant MFs. A MF $\beta_t^s$ is time invariant if for any $r$, and $s \leq t$,

$$\beta_{t+s}^r = \beta_t^s \circ \theta_r$$

where $\theta_r$ is the shift operator $(\theta_r x)_s = X_{r+s}$. In particular this implies that $\beta_t^s = \beta_t^0 \circ \theta_s$, so that $\beta_t^s$ is really a one-parameter functional; indeed denoting $\beta_t = \beta_t^0$ we can write the multiplicative property as

$$\beta_{t+s}^t = \beta_t \beta_s \circ \theta_t.$$

Let $T_t^\beta$ be the semigroup corresponding to $\beta$, defined by

$$T_t^\beta f(x) = \mathbb{E}_x[f(X_t^\beta)].$$

We wish to consider the generator of $T_t^\beta$. If $\beta$ satisfies

$$\mathbb{E}_x[\beta_t^s] \leq 1$$

(or, equivalently, $T_t^\beta 1 \leq 1$) then, as shown in Blumenthal and Getoor (1968), one can construct (possibly on an enlarged state space) the $\beta$-sub-process of $X_t$, which is a Markov process $X_t^\beta$ satisfying

$$T_t^\beta f(x) = \mathbb{E}_x[f(X_t^\beta)].$$

As in §3 above, the extended generator of $X^\beta$ is an operator $A^\beta$ such that

$$f(X_t^\beta) - f(x) - \int_0^t A^\beta f(X_s^\beta) ds$$

is a martingale for $f \in C_0^\infty(\mathbb{S})$, and from (4.7) this is equivalent to saying that

$$C_t^{\beta f} := \beta_t f(X_t) - f(x) - \int_0^t \beta_s A^\beta f(X_s) ds$$

is a martingale. The latter formulation has however the advantage that it does not involves the $\beta$-sub process or condition (4.8) (which is not satisfied in any of the applications we have in mind). We thus define the extended generator of $T_t^\beta$ as an operator $A^\beta$ such that $C_t^{\beta f}$ given by (4.9) is a martingale for all $f \in C_0^\infty(\mathbb{S})$. Here then are the types of multiplicative functional.

1. **Gauge transformation type.** Suppose $a \in C_0^\infty(\mathbb{S})$ and $a(x) > 0$ for all $x \in \mathbb{S}$. Define

$$\beta_t = \frac{a(X_t)}{a(X_0)}.$$  

This is clearly a MF, and from (4.8),

$$T_t^\beta(x) = \frac{1}{a(x)} T_t(a)(x).$$

2. **Feynman-Kac type.** Define

$$\beta_t = \exp \left( \int_0^t \gamma(x) \, ds \right),$$

where $\gamma(x)$ is a $\beta$-sub-process of $X_t$. Then $\beta_t$ is a MF and

$$T_t^\beta f(x) = \mathbb{E}_x[f(X_t^\beta)].$$

3. **Girsanov type.** Suppose there exists a Markov process $Y_t$ such that $\beta_t = \exp \left( \int_0^t \gamma(x) \, ds \right)$, where $\gamma(x)$ is the generator of $Y_t$. Then $\beta_t$ is a MF and

$$T_t^\beta f(x) = \mathbb{E}_x[f(Y_t)].$$
Using the signal equation (3.2) with $f = ag$ we see that $C^bg$ is a martingale if

$$A^bg(x) = \frac{1}{a(x)} A(ag)(x)$$

and this is therefore the generator of $T^b_t$.

(γ) Feynman-Kac type. For a given $v \in C^\infty_b(S)$, define

$$\gamma_t = \exp \left( - \int_0^t v(X_s)ds \right).$$

(4.11)

Computing the product of the semimartingles $f(X_t)$ and $\gamma_t$ using (3.2) and the Ito formula shows immediately that

$$A^7f(x) = Af(x) - v(x)f(x).$$

(δ) Girsanov type. The above transformations can be applied to any Markov process but this one is specifically tied to the model (3.1). Fix $g \in C^\infty_b(S)$ and define

$$\delta_t = \exp \left( - \int_0^t V_jg(X_u)dW^j_u - \frac{1}{2} \sum_j \int_0^t (V_jg(X_u))^2 du \right).$$

(4.12)

A standard application of the Girsanov theorem shows that we can define a new measure $\mathbb{P}^\delta$ by taking $d\mathbb{P}^\delta/d\mathbb{P} = \delta_T$, and that under $\mathbb{P}^\delta$

$$d\tilde{W}^j_u := dW^j_u + V_jg(X_u)du, \quad u \leq t$$

is a standard Brownian motion for $j = 1, 2, \ldots, r$. Thus (3.1) becomes

$$df(X_u) = (V_0f(X_u) - V_jg(X_u)V_jf(X_u))du + V_jf(X_u) \circ d\tilde{W}^j_u.$$ 

Now

$$\mathbb{E}_x[f(X_t)|\delta_t] = \mathbb{E}^\delta_x[f(X_t)]$$

and it follows that

$$A^\delta = A - (V_jg)V_j.$$ 

(4.13)

The three transformations are related by the $X_t$ equation written in Ito form (3.2); indeed, from (3.2)

$$\int_s^t V_jg(X_u)dW^j_u = g(X_t) - g(X_s) - \int_s^t Ag(X_u)du,$$

and inserting this in (4.12) we see that $\delta_t$ factors in form

$$\delta_t = \beta_t\gamma_t$$

where $\beta, \gamma$ are given by (4.10) and (4.11) respectively with

$$a(x) = e^{-g(x)}$$

$$v(x) = -Ag(x) + \frac{1}{2} \sum_j (V_jg(x))^2.$$
Applying $\beta$ and $\gamma$ successively (the order is immaterial) we conclude that
\[ A^\delta f = e^g A(e^{-g}f) + (Ag - \frac{1}{2} \sum_j(V_j g)^2)f. \] (4.14)

But this is just a disguised form of the Baker-Campbell-Hausdorff formula: using the expression (4.13) for $A^\delta$ and the relations (3.4), equation (4.14) becomes
\[ e^g A e^{-g} = A - (V_j g) V_j - Ag + \frac{1}{2} \sum_j(V_j g)^2 = A + ad_y A + \frac{1}{2} ad_y^2 A. \] (4.15)

(Recall that $ad_y^k A = 0$ for $k > 2$.)

4.3 The generator of $T^y_{s,t}$

Recall from §4.1 that the MF appearing in the Kallianpur-Striebel formula is
\[ \alpha^s_t(y) = \exp \left( -\int_s^t y(u) dh(X_u) - \frac{1}{2} \int_s^t h^2(X_u) du \right). \]

Using (3.2) we can factor this into the product of a Girsanov MF and a Feynman-Kac MF as follows:
\[ \alpha^s_t(y) = \exp \left( -\int_s^t y(u) Ah(X_u) du \right) \exp \left( \int_s^t \frac{1}{2} y^2(u) \sum_j(V_j h(X_u))^2 du \right). \]

It follows immediately that the corresponding generator is
\[ A^g_{s,t} f = Af - y(s)(V_j h)V_j f \]
\[ + \left( \frac{1}{2} y^2(s) \sum_j(V_j h)^2 - y(s)Ah - \frac{1}{2} h^2 \right) f \] (4.16)
\[ = Af + y(s)(ad_h A)f + \frac{1}{2} y^2(s)(ad_h^2 A)f - \frac{1}{2} h^2 f \]
\[ = e^{y(s)h} A(e^{-y(s)h} f) - \frac{1}{2} h^2 f, \] (4.17)

the last equality being an application of (4.15) with $g = y(s)h$.

It is clear a priori that (4.17) must be the right formula: in (4.16) the calculation is done for an arbitrary function $y(\cdot)$ but the result depends only on $y(s)$. Therefore $A^g_{s,t} = A^g_{s,s}$ where $\bar{y}$ is the constant function
\[ \bar{y}(u) = y(s) \quad \text{for } u \geq s. \]

But
\[ \alpha^s_t(\bar{y}) = \exp(-y(s)h(X_t) + y(s)h(X_s)) \exp \left( -\frac{1}{2} \int_s^t h^2(X_u) du \right). \]
This factors $\alpha_s(\bar{y})$ into the product of a gauge MF and a Feynman-Kac MF, and (4.17) is immediate for $\bar{y}(\cdot)$. But of course some extra work has to be done to show that the same formula works for non-constant $y(\cdot)$

Note from (4.16) that $A^Y_s$ is of the form

$$A^Y_s f(x) = \frac{1}{2} \sum_j V_j^2 f(x) + \tilde{V}_0(y(s)) f(x) + \psi(x, y(s)) f(x),$$

where $\tilde{V}_0$ is a $y$-dependent vector field, i.e. the second-order part of $A^Y_s$ is the same as that of $A$, and the effect of the MF transformation is only to add $y$-dependent ‘drift’ and ‘potential’ terms. Thus essentially the same conditions that ensure smooth solutions of the Fokker-Planck equation of the signal process also ensure smooth solutions of (4.5) (except that these conditions must allow for continuous, but not differentiable, $t$-dependence in the coefficients).

The general conclusion is that computing the conditional distribution of $X_t$ given $Y_t$ is not in any essential way more complicated than computing the unconditional distribution. See Pardoux (1979) and Pardoux (1982) for the case $S = \mathbb{R}^d$.

4.4 Dossing the Zakai equation

There is another way of looking at the basic formula

$$\rho_t(f) = \langle T^y_{0,t}(e^{y(t)h} f), \pi \rangle$$

and that is as a Doss-Sussmann ‘pathwise solution’ Doss (1977), Sussmann (1978) of the Zakai equation (2.4). This was indeed how (4.18) was originally arrived at, and although the MF approach turns out to be more fundamental, the pathwise solution idea is of value in understanding the picture and particularly in unravelling the complexities of the correlated noise case (see §5.3 below).

Let us recall the Doss-Sussmann construction for the simplest type of scalar equation

$$dX_t = f(X_t)dt + g(X_t) \circ dM_t, \quad X_0 = x,$$  

where $M_t$ is a real-valued continuous semimartingale and $f, g$ are smooth functions. (The same basic idea is used with considerably more elaboration in §5.1 below). Let $G(t, x)$ be the flow of $g$, i.e. the solution of the ordinary differential equation

$$\frac{\partial}{\partial t} G(t, x) = g(G(t, x)) \quad G(0, x) = x.$$  

Then the solution of (4.19) takes the form

$$X_t = G(M_t, \eta_t)$$  

where $\eta$ is the solution of another ODE, parametrized by the sample path $(M_t)$. Indeed, defining $X_t$ by (4.20) we have

$$dX_t = g(X_t) \circ dM_t + G_x(M_t, \eta_t) \eta_t dt$$

But

$$G_x(t, x) = \exp \left( \int_0^t g_x(G(s, x)) ds \right),$$
so that (4.19) and (4.21) agree as long as
\[ \dot{\eta}_t = \exp \left( - \int_0^{M_t} g_x(G(s, \eta_t)) \, ds \right) f(G(M_t, \eta_t)), \quad \eta_0 = x. \] (4.23)

This is an ordinary differential equation for \( \eta \), parametrized by the sample path \((M_t)\), and shows that the solution of (4.19) can be calculated separately for each sample path of \((M_t)\): first solve (4.23) and then evaluate (4.20). The same construction works for \( X_t \in \mathbb{R}^n \), except for the explicit expression (4.22), but not in general, for vector \( M_t \) (see below). Things are particularly simple in the bilinear case: \( f(x) = Ax \), \( g(x) = Hx \). Then (4.23) and (4.20) become respectively
\[ \dot{\eta}_t = e^{-H M_t} A e^{H M_t} \eta_t, \quad \eta_0 = x, \]
\[ X_t = e^{H M(t)} \eta_t. \]

Let us now apply the same argument to the bilinear measure-valued Zakai equation (2.4). In Stratonovich form this is
\[ d\rho_t(f) = \rho_t((A - \frac{1}{2} h^2)f) \, dt + \rho_t(h f) \circ dY_t, \quad \rho_0 = \pi. \] (4.24)

If the drift term in (4.24) were absent then the solution would be, as is easily checked,
\[ \rho_t(f) = \langle e^{Y(t)h f}, \pi \rangle. \]

An argument exactly analogous to the above shows that the solution with the drift term is given by (4.18), if \( T_{0,t}^Y \) is a semigroup with generator
\[ A^Y_t f = e^{Y(t)h} A(e^{-Y(t)h} f) - \frac{1}{2} h^2 f. \] (4.25)

But we saw in (4.14) above that this precisely is the generator of the semigroup given by the Kallianpur-Striebel formula. Thus the two approaches lead to the same result. If, however, one starts with the Zakai equation, one has somehow to show that there exists a semigroup whose generator is (4.25). The only way to do this that I know of is through a probabilistic argument Davis (1979) which leads straight back to the Kallianpur-Striebel formula. This is why I describe the MF approach as ‘more fundamental’.

Finally, let us note that all of the above results extend without difficulty to the case of vector observations
\[ dY^i_t = h^i(X_t) \, dt + dW^0_t, \quad i = 1, \ldots, m, \]
if \( X_t \) and \( W^0_t \) are independent for all \( i \). The Zakai equation is
\[ d\rho_t(f) = \rho_t(Af) \, dt + \sum_{i=1}^m \rho_t(h^i f) dY^i_t, \]
and a pathwise solution is constructed from this (or from the Kallianpur-Striebel formula) as before. The reason this ‘works’ is that the operators of ‘multiplication by \( h^i \)’, which appear in the diffusion term of the Zakai equation, commute: \( h^i h^j f(x) = h^j h^i f(x) = h^i(x) h^j f(x) \).

Recall that the condition under which the Doss-Sussmann construction for (4.19) can be extended to multiple inputs \( \sum g^i(X_t) \circ dM_t^i \) is precisely that the vector fields \( g^i \) commute.
5 The correlated noise case

We now wish to consider the filtering problem given by (2.1) and (3.1) as before, but allowing for possible correlation between the signal noise \((W^1, \ldots, W^r)\) and the observation noise \(W^0\).

We assume the simplest form of correlation; it will be obvious how to extend the results to more general cases. Specifically, we suppose

(i) \(W_t^i\) is a standard Brownian motion (i.e. \(W_0^i = 0\) and \(<W^i>_t = t, \; i = 0, 1, \ldots, r\)).

(ii) \(W^i, W^j\) are independent for \(i \neq j \neq 0\)

(iii) \(<W^i, W^0>_t = \alpha_i t\) for constants \(\alpha_i\) such that \(\sum_i \alpha_i^2 < 1\).

The first two of these are the same as before, while the condition in (iii) ensures a well-defined and strictly positive definite covariance matrix. We have in particular

\[ \mathbb{E}[W_t^i W_s^0] = \alpha_i t \land s. \]

The Kalman-Striebel formula is no longer valid in the form (2.3); we shall derive the correct form in §5.2 below. As regards the Zakai equation, it follows directly from the general filtering equation of Mitter (1979) that (2.4) should be amended to

\[ d\rho_t(f) = \rho_t(Af)dt + \rho_t(Df)dY_t \]

where \(D = Z + h\), with

\[ Z = \alpha_j V_j \]

and \(h\) denoting the zeroth order operator \(hf(x) = h(x)f(x)\).

5.1 Decomposition of the signal equation

To get the appropriate form of the Kalman-Striebel formula, introduce a measure \(\mathbb{P}_0\) via the Girsanov transformation

\[ \frac{d\mathbb{P}_0}{d\mathbb{P}} = \exp \left( -\int_0^T h(X_s) dW_s^0 - \frac{1}{2} \int_0^T h^2(X_s) ds \right) \]

and for \(i = 1, 2, \ldots, r\) define

\[ dv^i := dW^i + \alpha_i h(X_t) dt \]

Then, under \(\mathbb{P}_0\),

(i) \(Y_t\) and \(v^i_t, i = 1, \ldots, r\) are standard Brownian motions

(ii) \(v^i, v^j\) are independent for \(i \neq j\)

(iii) \(<v^i, Y>_t = \alpha_i t\).

Now project the \(v^j\) onto \(Y\), i.e. define

\[ b^i_t = v^i_t - \alpha_i Y_t. \]

Then each \(b^i\) is an unnormalized Brownian motion, which is uncorrelated with, and hence independent of, \(Y\). Denote \(b'_t = (b^i_t, \ldots, b^r_t)\), \(\alpha' = (\alpha_1, \ldots, \alpha_r)\) and let \(I\) denote the \(r \times r\) identity matrix. Then

\[ <b>_t = (I - \alpha \alpha') t. \]
Now $I - \alpha \alpha'$ is positive definite and can be factored into a product $\Xi \Xi'$ of positive definite matrices. Defining

$$B_t := \Xi^{-1} b_t$$

we find that

$$< B >_t = It,$$

i.e. $B^1, ..., B^r$ are (under measure $\mathbb{P}_0$) independent standard Brownian motions, independent of $Y$. Using (5.3) - (5.5), the signal equation (3.1) becomes

$$df(X_t) = L_0 f(X_t) dt + Z f(X_t) \circ dY_t + L_j f(X_t) \circ dB_t^j$$

where $Z$ is given by (5.2),

$$L_0 := \mathbb{V}_0 - hZ$$

and, in an obvious notation, the $L_j$ are given by

$$L = \Xi' \mathbb{V}.$$ 

Equation (5.6) is the key formula for the filtering problem, as it expresses $X_t$ under measure $\mathbb{P}_0$ in the form of an equation driven by the observation process $Y_t$ and the other ‘inputs’ $B^1, ..., B^r$ which are independent of $Y$.

To proceed, we follow the approach of Kunita (1981). Essentially, the idea is to use a transformation of the Doss-Sussmann type, as in §4.4, to express the solution of (5.6) sample-pathwise in $Y$.

The tangent space $T_p(S)$ at $p \in S$ consists of the set of derivations, i.e. linear functionals $W_p$ on $C^\infty(U_p)$, where $U_p$ is a neighbourhood of $p$, such that the Leibnitz rule.

$$W_p(fg) = gW_p f + W_p g$$

is satisfied. Now let $\phi : S \to S$ be a diffeomorphism and denote $q = \phi(p)$. Then $\phi$ defines a map $\phi^* : C^\infty(U_q) \to C^\infty(U_p)$ by composition:

$$\phi^* f := f \circ \phi, \quad f \in C^\infty(U_q),$$

and a map $\phi_* : T_p(S) \to T_q(S)$ as follows:

$$(\phi_* W_p) f = W_p (\phi^* f), \quad f \in C^\infty(U_q).$$

Since $\phi^{-1}$ is also a diffeomorphism, $\phi_*^{-1} : T_q(S) \to T_p(S)$ is given likewise by

$$(\phi_*^{-1} W_q) g = W_q (g \circ \phi^{-1}), \quad g \in C^\infty(U_p).$$

If $W$ is a vector field and $W_p$ denotes its restriction to $p \in S$ then this relation defines a mapping, also denoted $\phi_*^{-1}$, between vector fields, which, since $q = \phi(p)$, we can write

$$(\phi_*^{-1} W) g(p) = W(g \circ \phi^{-1})(\phi(p)).$$

Let $\zeta_t(x) = \zeta(t, x)$ denote the flow of the vector field $Z$ defined by (5.2), i.e. the unique solution of the equation

$$\frac{d}{dt} f(\zeta_t(x)) = Z f(\zeta_t(x)), \quad f \in C^\infty(S),$$

$$\zeta_0(x) = x.$$
This is a diffeomorphism for each \( t \geq 0 \). Define
\[
\xi_t(x) = \zeta_{Y(t)}(x).
\]
As is easily checked, \( \xi = \xi_t(x) \) is the solution of
\[
d\xi_t = Z(\xi_t) \circ dY_t
\]
and, obviously, \( \xi_t(\cdot) \) is almost surely a diffeomorphism for each \( t > 0 \). Now consider the equation
\[
df(\eta_t) = \xi^{-1}_t L_0 f(\eta_t) dt + \xi^{-1}_t L_j f(\eta_t) \circ dB^j_t. 
\]
This equation has a unique solution and it follows by applying the Ito formula that
\[
X_t(x) = \xi_t \circ \eta_t(x)
\]
\[
= \zeta(Y(t), \eta_t(x)).
\]
The representation (5.7), (5.8) describes the behaviour of \( X_t \) conditioned on \( Y \) under measure \( \mathbb{P}_0 \). Recall that the map \( \xi^{-1}_t \) is parametrized by \( Y \) and that \( Y, B' \) are independent. Thus, conditional on \( Y, \eta_t \) is a diffusion process whose differential generator is
\[
A^*_t = \xi^{-1}_t L_0 + \sum_j (\xi^{-1}_t L_j)^2
\]
and, for each \( t > 0 \), \( X_t \) is diffeomorphically related to \( \eta_t \) by equation (5.8).

5.2 The Kallianpur-Striebel formula and associated multiplicative functional

It follows from a standard formula of conditional expectations that \( \pi_t(f) \) of (2.2) is given in terms of the measure \( \mathbb{P}_0 \) by
\[
\pi_t(f) = \frac{\rho_t(f)}{\rho_t(1)}
\]
where
\[
\rho_t(f) := \mathbb{E}_0 \left[ f(X_t) \exp \left( \int_0^t h(X_s) dY_s - \frac{1}{2} \int_0^t h^2(X_s) ds \right) \bigg| \mathcal{Y}_t \right]. \tag{5.9}
\]
It is immediate from (5.6) that
\[
d < h(X_t), Y >= Z h(X_t) dt
\]
and hence that the Stratonovich version of (5.9) is
\[
\rho_t(f) = \mathbb{E}_0 \left[ f(X_t) \exp \left( \int_0^t h(X_s) \circ dY_s - \frac{1}{2} \int_0^t D h(X_s) ds \right) \bigg| \mathcal{Y}_t \right], \tag{5.10}
\]
where, as in (5.1) above, \( D = Z + h \). Now use (5.8), giving \( X_t \) in the form \( X_t = \zeta(Y_t, \eta_t) \) where \( \eta_t \) is a functional of the independent processes \( Y_t \) and \( B' = (B^1_t, \ldots, B^r_t) \), to express (5.10) in the form
\[
\rho_t(f) = \mathbb{E}^b \left[ \xi^*_t f(\eta_t) \exp \left( \int_0^t \xi^*_s h(\eta_s) \circ dY_s - \frac{1}{2} \int_0^t \xi^*_s D h(\eta_s) ds \right) \right] \tag{5.11}
\]
where $E^b$ means integration over the sample space measure for $B$, i.e. Wiener measure on $C([0,T]; \mathbb{R}^r)$. This is the ‘Kallianpur-Striebel’ formula for the correlated-noise problem. In order to get it in ‘robust’ form we need to calculate the stochastic integral in (5.11) as an explicit functional of $y$. Introduce the function (we write $\zeta_s(x) = \zeta(s,x)\int_0^s \zeta^*_sh(x)ds$)

$$H_t(x) = H(t,x) := \int_0^t \zeta^*_sh(x)ds$$

and calculate $H(Y_t, \eta_t)$ using the Ito formula and (5.8). This gives

$$H(Y_t, \eta_t) = \int_0^t h(X_s) \circ dY_s + \int_0^t (\xi_s^{-1}L_0)H_{Y_s}(\eta_s)ds + \int_0^t (\xi_s^{-1}L_j)H_{Y_s}(\eta_s) \circ dB^j_s. \tag{5.12}$$

The stochastic integral with respect to $B^j$ in (5.12) can be re-expressed in Ito form in the standard way using (5.7). Do this and introduce the notation

$$g_s(x) = \int_0^s \zeta^*_sh(x)du \tag{5.13}$$

$$L^*_j = \xi_s^{-1}L_j$$

$$U_sf(x) = \exp\left(\int_s^0 \zeta^*_sh(x)du\right)\zeta^*_sf(x).$$

Then using (5.12) in (5.10) gives

$$\rho_t(f) = E[U_{Y_t}f(\eta_t)\alpha_t^0(Y)] \tag{5.14}$$

where, for $y \in C([0,T])$,

$$\alpha_t^s(y) = \exp\left(-\int_s^t L^*_jg_u(\eta_u)dB^j_u - \frac{1}{2}\sum_j \int_s^t (L^*_j)^2g_u(\eta_u)du \right. \left. - \int_s^t L^*_jg_u(\eta_u)du - \frac{1}{2}\int_s^t \xi^*_udh(\eta_u)du\right). \tag{5.15}$$

Equation (5.14) is the desired multiplicative functional formula. For each sample path of $y$, $\eta_t$ is a diffusion process governed by vector fields $L^*_j$ as in (5.7), and $\alpha_t^s$ given by (5.15) is a MF of $\eta_t$. The expectation in (5.14) is taken over the distribution of $\eta_t$ for a fixed path $y$.

It is possible to show that $\rho_t(f)$ given by (5.14) is continuous in $y \in C([0,T])$, i.e. that this is a ‘robust’ version in the sense of (2.6), but we do not give the details here. In outline, one starts with functions $f \in C^\infty_b(S)$ whose support is contained within a single chart of $S$. Then, working in local coordinates, equation (5.7) satisfies the standard Ito conditions and continuous dependence (in the mean square sense) of $f(\eta_t)$ on $y \in C([0,t])$ follows from known results on parametric dependence of solutions of stochastic differential equations; see Theorem 2, §2.7, of Gihman and Skorohod (1972). One completes the argument for $f \in C^\infty_b(S)$ by considering a decomposition of the form

$$\rho_t(f) = \sum_i \rho_t(\lambda_if)$$

where $(\lambda_i)$ is a partition of unity: $\lambda_i \in C^\infty_N$ for all $i$ and $\sum_i \lambda_i(x) = 1$. 
As in §4.3 above, we can compute the generator $A^\alpha_t$ corresponding to the MF $\alpha^\alpha_t(y)$ by factorization. Indeed

$$
\alpha^\alpha_t(y) = \exp \left( - \int_s^t L^*_j g_u(\eta_u) db_u^j - \frac{1}{2} \sum_j \int_s^t (L^*_j g_u(\eta_u))^2 du \right) \times \exp \left( \int_s^t \left( \frac{1}{2} \sum_j (L^*_j g_u(\eta_u))^2 - \frac{1}{2} \sum_j (L^*_j g_u(\eta_u)) - L^*_0 g_u(\eta_u) - \frac{1}{2} \xi^*_u Dh(\eta_u) \right) du \right).
$$

It now follows as before that

$$
A^\alpha_t f = A^*_t f - L^*_j g_t L^*_j f + \left( \frac{1}{2} \sum_j (L^*_j g_t)^2 - A^*_t g_t - \frac{1}{2} \xi^*_t Dh \right) f
$$

(5.16)

where $A^*_t$ is the generator for $\eta_t$, i.e.

$$
A^*_t = L^*_0 + \frac{1}{2} \sum_j (L^*_j)^2.
$$

(5.17)

$A^\alpha_t$ can be expressed in somewhat more explicit form by noting that

$$
L^*_j g_t(x) = \int_0^{y(t)} (\varsigma^{-1}_t L_j \varsigma^*_u h(x) du
$$

$$
= \int_0^{y(t)} L_j \varsigma^*_u h(\varsigma^*_u(x)) du
$$

$$
= \int_0^{y(t)} L_j h(\varsigma^*_u \varsigma^{-1}_t(x)) du
$$

$$
= \int_0^{y(t)} L_j h(\varsigma_t(x)) du
$$

The similarity of (5.16) to (4.16) is obvious (of course, (5.16) reduces to (4.16) if $\alpha = 0$) and similar remarks are pertinent: $A^\alpha_t$ differs from $A^*$ only in the ‘drift’ and ‘potential’ terms and therefore the complexity of computing the conditional distribution is essentially that of computing the distribution of the decomposition $\eta_t$ of the signal process $X_t$.

### 5.3 Solution of the Zakai equation

The Zakai equation for the correlated noise problem was given in (5.1); in Stratonovich form it is

$$
d\rho_t(f) = \rho_t((A - \frac{1}{2} D^2) f) dt + \rho_t(D f) \circ dY_t.
$$

(5.18)

Now

$$
A - \frac{1}{2} D^2 = \frac{1}{2} \sum_j V_j^2 - \frac{1}{2} (\alpha_j V_j + h)^2 + V_0
$$

I thank the referee of this chapter for this observation.
and a completion-of-squares calculation shows that
\[ A - \frac{1}{2}D^2 = L_0 + \sum_j L_j^2 - \frac{1}{2}Dh \]  \hspace{1cm} (5.19)
where \( L_0, L_j \) are as in (5.6). The operator \( D = Z + h \) is the generator of the group \( U_t \) on \( C^\infty(\mathbb{S}) \) given by (5.13). An argument analogous to that of §4.4 shows that the Doss-Sussmann solution of (5.18) is
\[ \rho_t(f) = \langle T_{0,t}^Y(U_{Y_t} f), \pi \rangle \]
where \( T_{0,t}^Y \) is the semigroup whose generator is
\[ \tilde{A}_t^Y = U_{Y_t}(A - \frac{1}{2}D^2)U_{-Y_t}. \]  \hspace{1cm} (5.20)
We propose to show, using the Baker-Campbell-Hausdorff formula, that this coincides with \( A_t^Y \) given by (5.16) above. Denote by \( C_t \) the multiplication operator \( C_t f(x) = f(x) \exp \left( \int_0^{Y_t} \xi_t^* h(x) ds \right) \) so that
\[ U_{Y_t} = C_t \xi_t^* \]  \hspace{1cm} (5.21)
and
\[ U_{-Y_t} = U_{-1} = \xi_t^{-1} C_t^{-1}. \]  \hspace{1cm} (5.22)
(Note that \( C_t^{-1} \neq C_{-t} \)) Using (5.19) - (5.22) we can see that
\[ \tilde{A}_t^Y = C_t \xi_t^* (L_0 + \frac{1}{2} \sum_j L_j^2 - \frac{1}{2}Dh)(\xi_t^{-1})^* C_t^{-1} \]
Now
\[ \xi_t^* Dh(\xi_t^{-1})^* f(x) = \xi_t^* Dh(x)f(x) \]
and
\[ \xi_t^* (L_0 + \frac{1}{2} \sum_j L_j^2)(\xi_t^{-1})^* = A_t^* \]
where \( A_t^* \) is given by (5.17). Thus
\[ \tilde{A}_t^Y = C_t A_t^* C_t^{-1} - \frac{1}{2} \xi_t^* Dh. \]  \hspace{1cm} (5.23)
Since \( C_t \) is a multiplication operator, we can expand the right-hand side using the Baker-Campbell-Hausdorff formula (4.15). We obtain
\[ C_t A_t^* C_t^{-1} = A_t^* - (L_t^* g_t) L_t^* - A_t^* g_t + \frac{1}{2} \sum_j (L_j^* g_t)^2. \]
Using this expression in (5.23) we see that \( \tilde{A}_t^Y \) coincides with \( A_t^Y \) given by (5.16). Thus, as claimed in §1, the results are formally analogous to the independent case with the operator \( U_t \) replacing the operator of multiplication by \( \exp(t h(x)) \). However, while it is (with a bit of hindsight) in the independent case fairly obvious from the Kallianpur-Striebel formula that the appropriate generator is (4.17), the interpretation of (5.16) in the correlated case is by no means obvious and it seems essential to look at the Zakai equation to get the full picture.
Finally, if there are vector observations then the last term in (5.18) will be of the form
\[ \sum_i \rho_t(D^i f) \circ dy^i_t \]
where
\[ D^i = Z^i + h^i \]
for some vector fields $Z^i$. There cannot be a pathwise solution of (5.18) unless the $D^i$ commute, but this only happens under extremely artificial conditions. If the $D^i$ do not commute a decomposition of the type (5.8) is still possible, where $\xi_t$ is almost surely a diffeomorphism (Kunita, 1981), but no continuous dependence of $\xi_t$ on $Y$ can be expected. Thus the present results are essentially limited to the scalar case.

Acknowledgments

J.M.C. Clark initiated the whole concept of pathwise filtering and I have had many conversations on these topics with him. Another primary influence was the set of lectures given by H. Kunita at the 1980 London Mathematical Society Durham Symposium on Stochastic Integration. On the algebraic and geometric side, I had helpful discussions with S.K. Mitter and a preview of his paper Mitter (1979), and some patient instruction in the basics of differential geometry from A.J. Krener.

It is extraordinary, but true, that both the original 1981 version of this paper and the present 2009 version were word-processed by Doris Abeysekera (using an IBM golf-ball typewriter and \LaTeX respectively). I take this opportunity to express my gratitude to her for many years of friendly cooperation.

References


