Stochastics, 1982, Vol. 8, pp. 1–16
0090-9491/82/0801-0001 $06.50/0

Explicit Filters for Diffusions with Certain Nonlinear Drifts†

D. L. OCONE,‡ J. S. BARAS§ and S. I. MARCUS¶

‡*Mathematics Department and the Mathematics Research Center, University of Wisconsin-Madison, Madison, WI 53706.*
§*Electrical Engineering Department, University of Maryland, College Park, MD 20742.*
¶*Department of Electrical Engineering, University of Texas at Austin, Austin, TX 78712.*

(*Accepted for publication December 9, 1981*)

Let $x(t)$ be a diffusion satisfying the stochastic differential equation $dx(t)=f(x(t))dt+db(t)$, where $f'(x)+f^2(x)=ax^2+bx+c$, $a\geqq 0$. V. Beneš gave an explicit formula for the conditional density of $x(t)$ given $y(s)$, $0\leqq s\leqq t$, where $y(s)=\int_0^t x(s)ds+w(t)$, when $w(\cdot)$ is a Brownian process independent of $x(\cdot)$. This result is extended and then applied to derive recursive filtering equations for estimating conditional moments $E\{x^n(t)|y(s),\ 0\leqq s\leqq t\}$, for estimating polynomial functionals of $x(\cdot)$, and for smoothing.

1. INTRODUCTION

Let $f(x)$ be a real-valued function defined on all of R and satisfying the Riccati equation

$$f'(x)+f^2(x)=ax^2+bx+c. \tag{1.1}$$

It is assumed that f has no singularities. Note that this implies $a\geqq 0$, for otherwise f explodes at some finite x. This paper considers the filtering

†Sponsored by the United States Army under Contract No. DAAG29-80-C-0041. The work of the first author was supported by the National Science Foundation under Grant No. MCS-7927062. The work of the second author was supported in part by the Office of Naval Research under Contract N00014-79-C-0808. The third author was supported in part by the Joint Services Electronics Program under Contract F4 9620-77-C-0101.

problem

$$dx(t) = f(x(t))dt + db(t)$$

$$x(0) = x \in R \tag{1.2}$$

$$dy(t) = x(t)dt + dw(t) \tag{1.3}$$

in which $b(\cdot)$ and $w(\cdot)$ are independent Wiener processes, $x(\cdot)$ is the signal, and $y(\cdot)$ the observation of $x(\cdot)$.

For the system (1.1)–(1.3), Beneš [1] recently derived an explicit formula for the conditional density of $x(t)$ given $\mathscr{F}_t^y$, where $\mathscr{F}_t^y := \sigma$-algebra generated by $\{y(s)|0 \leqq s \leqq t\}$. This result is interesting because the class of functions satisfying (1.1) includes nonlinear f, whereas conditional densities for (1.2)–(1.3) had been computed previously only for linear f: for examples and an extension to the multidimensional case, see Beneš [1].

Besides conditional densities one also wants to calculate filters $E\{\varphi | \mathscr{F}_t^y\}$ of $x(\cdot)$-dependent statistics φ. Given a random process $\varphi(t)$ we shall say that $E\{\varphi(t)|\mathscr{F}_t^y\}$ is *finite dimensionally computable* (FDC) if it can be expressed as the output of finite dimensional system of stochastic differential equations driven by $y(\cdot)$. When the signal dynamics in (1.2) are linear, many examples of FDC estimates are known. For example, $E\{x^n(t)|\mathscr{F}_t^y\}$ is FDC for any integer n, since the Kalman–Bucy equations calculate $E[x(t)|\mathscr{F}_t^y\}$ and the conditional variance, and higher order moments are derived from these by virtue of the normality of the conditional density. A more subtle class of examples consists of estimates $E\{\eta(t)|\mathscr{F}_t^y\}$ where $\eta(t)$ is any polynomial functional of $x(\cdot)$ in the form

$$\eta(t) = \int_0^t \dots \int_0^{s_{n-1}} \gamma(s_1, \dots, s_n) x^{k_1}(s_1) \dots x^{k_n}(s_n) ds_n \dots ds, \quad 1 \leqq i \leqq n,$$

in which γ is a separable function and the $\{k_i\}$ are non-negative integers (Marcus, Willsky [7]; Marcus, Mitter, Ocone [6]). Formulae and recursive systems for the smoothed estimate $E\{x(s)|\mathscr{F}_t^y\}$ are also well known (see [3]).

In this note, we extend the linear theory by showing that these same statistics are FDC for the general model (1.1)–(1.3). The strategy, as in the linear case, is to derive finite dimensional systems by using the explicit form of the conditional density to truncate formally infinite dimensional systems of moment equations. The material is organized as follows. In Section 2 we calculate conditional joint densities of $x(\cdot)$ given $\mathscr{F}_t^y$. As a consequence, we show that the conditional law of the process $\{x(s)|0 \leqq s \leqq t\}$ given $\mathscr{F}_t^y$ *and* $x(t)$ is Gaussian. This is precisely the feature that makes it possible to handle polynomial functionals. In Section 3, we prove FDC of conditional moments, smoothers, and polynomial functionals.

Lie algebraic techniques from geometric control theory have been introduced recently into filtering, especially as regards finite dimensional computability, and they have been worked out successfully for known FDC problems in which f is linear. (For a survey of these ideas, see Brockett [8].) The main results of this paper, in particular Proposition 3.7, were suggested by a Lie algebraic analysis of (1.1)–(1.3). Consequently, the full range of the Lie theory for linear drifts extends to the general case (1.1). Since our methods here are not algebraic, we do not pursue the issue further, but refer instead to Ocone [7] for further discussion.

2. CONDITIONAL JOINT DENSITIES

Let $x(\cdot)$ and $y(\cdot)$ be given from (1.1)–(1.3), and let $t=s_0>s_1>\ldots>s_n\geqq 0$, $\mathbf{z}=(z_0,z_1,\ldots,z_n)^T$. The expression $p(z_0,t;\ z_1,s_1,\ldots,\ z_n,s_n|\mathscr{F}_t^y)$ shall denote the joint density of $(x(t),x(s_1),\ldots,x(s_n))$ conditioned on $\mathscr{F}_t^y$, that is, for any bounded, Borel $\psi:\mathbf{R}^{n+1}+\mathbf{R}$

$$E\{\psi(x(t),\ldots,x(s_n))|\mathscr{F}_t^y\}=\int_{R^{n+1}}\psi(\mathbf{z})p(z_0t;z_1,s_1;\ldots;z_n,s_n|\mathscr{F}_t^y)d\mathbf{z}.$$

In Theorem 2.1 of this section, we employ a method of Beneš [1] to compute an explicit formula for this conditional density. From this we then derive $p(z_1,s_1;\ldots;z_n,s_n|\mathscr{F}_t^y,x(t))$, the conditional density of $(x(s_1),\ldots,x(s_n))$ given $\mathscr{F}_t^y$ and $x(t)$.

The results are stated in terms of an auxiliary process $\xi(t)$, evolving in R^3 and defined by

$$d\xi(t)=A(t)\xi(t)dt+\begin{bmatrix}1\\ y(t)\\ 0\end{bmatrix}dB(t)$$

$$\xi(0)=(x,0,0)^T$$

where

$$A(t)=\begin{bmatrix}-\kappa & 0 & 0\\ 0 & 0 & 0\\ \kappa y(t)-\frac{1}{2}b & 0 & 0\end{bmatrix},$$

$$\kappa=(a+1)^{1/2}$$

and $B(\cdot)$ is a Brownian motion independent of the signal and observation noises, $b(\cdot)$ and $w(\cdot)$.

Let

$$\Xi_0=(\xi^T(t),\xi^T(s_1),\ldots,\xi^T(s_n))^T$$

$$\Xi=(\xi_1(t),\xi_1(s_1),\ldots,\xi_1(s_n))^T$$

Then the following conditional moments are needed:

$$m(t):=E\{\xi(t)|\mathscr{F}_t^y\}$$

$$R(t,s):=\operatorname{cov}(\xi(t),\xi(s)|\mathscr{F}_t^y)$$

$$R(t)=[r_{ij}(t)]_{1\leq i,j\leq 3}=R(t,t)$$

$$M(t,s_1,\ldots,s_n):=E\{\Xi|\mathscr{F}_t^y)$$

$$=(m_1(t),\ldots,m_1(s_n))^T$$

$$P_0(t,s_1,\ldots,s_n):=\operatorname{Var}(\Xi_0|\mathscr{F}_t^y)$$

$$P(t,s_1,\ldots,s_n):=\operatorname{Cov}(\Xi,\Xi_0|\mathscr{F}_t^y)$$

$$=E\{(\Xi-M)(\Xi_0-E\{\Xi_0|\mathscr{F}_t^y\})^T|\mathscr{F}_t^y\}$$

$$Q(t,s_1,\ldots,s_n):=\operatorname{Var}(\Xi|\mathscr{F}_t^y).$$

To simplify later expressions, we shall often drop the $(t,s_1,\ldots,s_n)$ dependence and write only M, P_0, P and Q. In addition, let

$$v=(0,1,-1,0,\ldots,0)^T\in R^{3(n+1)}.$$

The random vectors P_0v and Pv play an important role in Theorem 2.1 and are related by

$$Pv=((P_0v)_1,(P_0v)_4,\ldots,(P_0v)_{3n+1})^T.$$

Note that $(Pv)_k=(P_0v)_{3k-2}=\operatorname{cov}\xi_1(s_{k-1}),\xi_2(t)-\xi_3(t)|\mathscr{F}_t^y)$.

It is important to observe that all these conditional moments are properly thought of as functionals (on $C[0,t]$) of $y(\cdot)$. Indeed, these functionals are easily calculated by solving for every $\bar{y}(\cdot)\in C[0,t]$ the system

$$d\xi_y(s)=A_{\bar{y}}(s)\xi_{\bar{y}}(s)ds+\begin{bmatrix}1\\ \bar{y}(s)\\ 0\end{bmatrix}dB(s)$$

$$\xi_{\bar{y}}(0)=(x,0,0)^T.$$

Then, if $\bar{m}(t,\bar{y}(\cdot))=E\xi_{\bar{y}}(t), m(t)(\omega)=\bar{m}(t,y(\cdot)(\omega))$, and similarly for R, P_0, etc.

As a final bit of notation, set $F(z_0)=\int_0^{z_0} f(s)\,ds$.

THEOREM 2.1 *Let* $t>s_1>s_2>\ldots>s_n$. *Then*

$$p(z_0,t;z_1,s_1;\ldots;z_n,s_n|\mathcal{F}_t^y)$$

$$=(1/\Psi)\exp\left\{F(z_0)+z_0y(t)+x\kappa(z_0-x)+\kappa\left[\frac{(z_0-x)^2}{2}-t\right]\right\}$$

$$\times\exp\left\{-\tfrac{1}{2}\langle\mathbf{z}-M+Pv,\ Q^{-1}(\mathbf{z}-M+Pv)\rangle\right\}$$

where $\Psi=\Psi(t,s_1,\ldots,s_n,x)$ *is a normalizing factor.*

Proof The demonstration is analogous to Beneš' proof in [1] of the case $p(z_0,t|\mathcal{F}_t^y)$, and so we shall only sketch the main steps. The Kallianpur–Striebel formula for conditional estimation in system (1.2)–(1.3) implies that

$$p(z_0,t;\ldots;z_n,s_n|\mathcal{F}_t^y)dz_0\ldots dz_n$$

$$\propto\tilde{E}\left\{\prod_{j=0}^{n}1_{\{x(s_j)\in dz_j\}}\exp\int_0^t x(s)dy(s)-\tfrac{1}{2}\int_0^t x^2(s)ds\right\} \tag{2.2}$$

In expression (2.2), we think of $y(t)$ as a fixed function and of $\tilde{E}$ as an expectation against the measure μ induced by (1.2) on the space of continuous sample paths $x(\cdot)$. '$\propto$' here means proportional up to a normalization factor that does not depend on $\mathbf{z}$. To evaluate (2.2) we follow Beneš and apply a sequence of Girsanov measure transformations. Let $\mu_{B,x}$ denote the measure on $C[0,t]$ induced by $x+B(\cdot)$. Then $\mu\ll\mu_{B,x}$ and

$$\frac{d\mu}{d\mu_{B,x}}(x(\cdot))=\exp\left\{\int_0^t f(x(s))dx(s)-\tfrac{1}{2}\int_0^t f^2(x(s))ds\right\}.$$

By using this to change measures in the expectation term of (2.2) and noting, by Ito's rule, that

$$\int_0^t x(s)dy(s) = x(t)y(t) - \int_0^t y(s)dx(s)$$

and

$$F(x+B(t)) - F(x) = \int_0^t f(x+B(s))dB(s) - \tfrac{1}{2}\int_0^t f'(x+B(s))ds,$$

we derive

$$p(\ldots)\,d\mathbf{z} \propto \exp\{F(z_0) - F(x) + z_0 y_t\}$$

$$\times E\left(\prod_{j=0}^{n} 1_{\{x+B(s_j)\in dz_j\}} \exp - \int_0^t y(s)dB(s) - \tfrac{1}{2}\int_0^t V(x+B(s))ds\right) \qquad (2.6)$$

where $V(x) = (a+1)x^2 + bx + c$. To evaluate this last expectation, we treat the quadratic part of V as arising from the Radon–Nikodym derivative of $x + B(\cdot)$ with respect to the Ornstein–Uhlenbeck process $\xi_1(\cdot)$; the linear terms in the exponent in (2.6) can then be re-expressed in terms of $\xi_2(\cdot)$ and $\xi_3(\cdot)$. The result is

$$\exp\left\{x\kappa(z_0 - x) + \kappa\left[\frac{(z_0-x)^2}{2} - t\right]\right\}$$

$$\times E\left\{\prod_{j=0}^{n} 1_{\{\xi_1(s_j)\in dz_j\}} \exp -(\xi_2(t) - \xi_3(t)) \middle| \mathscr{F}_t^y\right\}. \qquad (2.7)$$

Given $\mathscr{F}_t^y$, $\xi(\cdot)$ is a Gaussian process, and thus the expectation in (2.7) may be written, up to a normalizing factor, as

$$\int_{R^{2n+2}} d\bar{\zeta} \exp\{-\tfrac{1}{2}\langle \zeta - E_y\Xi_0 + P_0 v, P_0^{-1}(\zeta - E_y\Xi_0 + P_0 v)\rangle\} \exp \tfrac{1}{2}\langle v, P_0 v\rangle$$

where $E_y\Xi_0 = E\{\Xi_0 | \mathscr{F}_t^y\}$ and where $d\bar{\zeta}$ signifies that $\zeta_1 = z_0$, $\zeta_4 = z_1, \ldots, \zeta_{3n+1} = z_n$ are held fixed and integration is over the remaining variables. But this last expression integrates by standard Gaussian integral formulae to a factor proportional to

$$\exp\{-\tfrac{1}{2}\langle \mathbf{z} - M + Pv,\ Q^{-1}(\mathbf{z} - M + P)\rangle\}. \qquad (2.8)$$

By combining (2.6)–(2.8), we arrive at the desired result. □

The conditional density of the process $x(\cdot)$ thus consists of a Gaussian factor multiplied by $\exp F(z_0)$. Further conditioning on $x(t)$ will remove $\exp F(z_0)$ and leave only the normal part. Indeed, let

$$\Xi^{(2)}=(\xi_1(s_1),\ldots,\xi_1(s_n))^T$$

$$Q(t,s_1,\ldots,s_n)=\begin{bmatrix} r_{11}(t) & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}$$

$$Q_{22}(t,s_1,\ldots,s_n)=\mathrm{Var}\,(\Xi^{(2)}|\mathscr{F}_t^y)$$

$$Pv=\begin{bmatrix} r_{12}(t)-r_{13}(t) \\ Pv^{(2)} \end{bmatrix}, \qquad M=\begin{bmatrix} m_1(t) \\ M^{(2)} \end{bmatrix}$$

COROLLARY 2.9 *The conditional law of* $(x(s_1),\ldots,x(s_n))$ *given* $\mathscr{F}_t^y$ *and* $x(t)$ *is normal with mean*

$$M^{(2)}-Pv^{(2)}+r_{11}^{-1}(t)Q_{21}(x(t)-m_1(t)+Pv_1)$$

and variance $Q_{22}-r_{11}^{-1}(t)Q_{21}Q_{12}$.

Proof Use (2.1) to conclude that

$$p(z_1,s_1;\ldots;z_n,s_n|x_0(t)=z_0,\mathscr{F}_t^y)=p(z_0,t;\ldots;z_n,s_n|\mathscr{F}_t^y)/p(z_0,t|\mathscr{F}_t^y)$$

$$\propto\exp\{-\tfrac{1}{2}\langle \mathbf{z}-M+Pv,Q^{-1}(\mathbf{z}-M+Pv)\rangle\}$$

$$\times\exp\left\{\tfrac{1}{2}\frac{1}{r_{11}(t)}(z_0-m_1(t)+(Pv)_1)^2\right\}.$$

But this is just the conditional density of $\eta_1,\ldots,\eta_n$ given η_0 where $(\eta_0,\ldots,\eta_n)$ is a normal random vector with mean $M-Pv$ and variance Q. The result then follows from the standard formula for conditioning one part of a normal random vector upon another. □

This corollary demonstrates how closely those diffusions defined by (1.1)–(1.2) are related to Gaussian processes. In fact, the steps above can be repeated to calculate joint densities of the process $x(\cdot)$; (2.1) yields the correct expression if $y(\cdot)$ is replaced by 0. Then, in the same way, it follows that $(x(s_1),\ldots,x(s_n)|x(t))$ is normal, that is, the process $\{x(s)|\leq s\leq t]$ conditioned on the endpoint $x(t)$ is Gaussian. This conditional normality is key to the filtering results of Section 3.

3. FILTERING EQUATIONS

3.1 Conditional moments

Let $\widehat{x^n}(t) := E\{x^n(t)|\mathscr{F}_t^y\}$. $\widehat{x^n}(t)$ satisfies the equation

$$d\widehat{x^n} = [(n(n-1))/2\widehat{x^{n-2}} + n\widehat{f(x)x^{n-1}}]dt + [\widehat{x^{n+1}} - \hat{x}\widehat{x^n}][dy - \hat{x}dt] \quad (3.1)$$

(Fujisaki, Kallianpur, Kunita [2]). To calculate $\widehat{x^n}$ one must therefore also find $\widehat{x^{n+1}}$, $\widehat{f(x)x^{n-1}}$, etc. These in turn also satisfy stochastic differential equations that introduce yet other quantities to be estimated. Continuing in this manner, if we begin with $\hat{x}$, we arrive at an infinite, coupled system of conditional moment equations for $\hat{x}, \widehat{x^2}, \ldots, \widehat{x^n}, \ldots, \widehat{f(x)x}$, etc. We will show that $\widehat{x^n}$ is FDC by deriving expressions for $\widehat{f(x)x^{n-1}}$ and $\widehat{x^{n+1}}$ in (3.1) as functions of $\hat{x}, \ldots, \widehat{x^n}$ and t. The infinite system of moment equations therefore truncates without approximation to a finite dimensional system that computes $\widehat{x^n}$. Actually, FDC of the moments could be argued on general grounds using the fact that the FDC process $\mu(t)$ (see immediately below) characterizes $p(z,t|\mathscr{F}_t^y)$. However, the approach here leads to explicit filtering equations.
Straightforward analysis of (2.1) shows that

$$p(z,t|\mathscr{F}_t^y) = \frac{1}{N(t,x)} \exp\{F(z) - (z-\mu(t))^2/2\sigma(t)\}$$

where $N(t,x)$ is a normalizing factor, and

$$\dot{\sigma} = 1 - \kappa^2\sigma^2 \qquad \sigma(0) = 0 \quad (3.2)$$

$$d\mu = [-\kappa^2\sigma\mu - \tfrac{1}{2}b\sigma]dt + \sigma dy \qquad \mu(0) = x \quad (3.3)$$

(See Beneš [1].) Note that (3.2) implies that $1 - \kappa^2\sigma^2(t) > 0 \ \forall t \geqq 0$, and hence that

$$a < \sigma^{-1}(t) \qquad \forall t \geqq 0.$$

As a consequence, $\widehat{x^n}(t)$ is well-defined for any n. If $a > 0$, (1.1) implies $f(x) \sim a^{1/2}x$ as $|x| \to \infty$ and thus that $F(z) \sim (a^{1/2}/2)z^2$ as $|z| \to \infty$. Therefore

$$F(z) - (z-\mu)^2/2\sigma \sim \tfrac{1}{2}(a^{1/2} - \sigma^{-1})z^2, \qquad |z| \to \infty$$

and so $P(z,t|\mathscr{F}_t^y)$ decays as $\exp[-\delta z^2]$, $\delta = \tfrac{1}{2}(\sigma^{-1} - a^{1/2}) > 0$. If $a = 0$, then b

$=0$, $c>0$ is necessary in order that (1.1) have a solution f on R without singularities. Then f will be bounded, and $F(z)$ will grow at most linearly.

LEMMA 3.4 *For* $n \geqq 0$

$$(\sigma^{-1}-a)\widehat{x^{n+2}} = (b+2\mu\sigma^{-2})\widehat{x^{n+1}} + (c+((2n+1)-\mu^2)\sigma^{-1})\widehat{x^n}$$
$$+2\mu n\sigma^{-1}\widehat{x^{n-1}} - n(n-1)\widehat{x^{n-2}}.$$

Proof Integrating by parts, we have

$$\frac{1}{N(t,x)}\int_{\infty}^{\infty} dz \exp F(z) \frac{d^2}{dz^2}(z^n \exp[-(z-\mu)^2/2\sigma])$$

$$= \int_{-\infty}^{\infty} dz\,[f'(z)+f^2(z)]z^n\, p(z,t|\mathscr{F}_t^y)$$

$$= a\widehat{x^{n+2}} + b\widehat{x^{n+1}} + c\widehat{x^n}.$$

Calculation of $d^2/dz^2(z^n \exp[-(z-\mu)^2/2\sigma])$ in the first integral leads to the desired identity. □

PROPOSITION 3.5 $\widehat{x^n}(t)$ *is FDC for all n.*

Proof Since $\sigma^{-1}(t)-a>0$, Lemma 3.1 implies that $\widehat{x^m}(t)$, $m \geqq 2$ can be expressed as a linear combination with FDC coefficients of lower order conditional moments. Thus it suffices to prove that $\hat{x}(t)$ is FDC. For $\hat{x}(t)$, (3.1) becomes

$$d\hat{x} = \widehat{f(x)}dt + (\widehat{x^2} - \hat{x}^2)(dy - \hat{x}dt), \qquad \hat{x}(0) = x.$$

Now

$$\widehat{f(x)} = \frac{1}{N(t,x)}\int dz[d/dz \exp F(z)] \exp[-(z-\mu)^2/2\sigma]$$

$$= -\int dz(z-\mu)\sigma^{-1} p(z,t|\mathscr{F}_t^y)$$

$$= -(\hat{x}-\mu)\sigma^{-1}$$

and

$$\widehat{x^2} = [\hat{x}(b+2\mu\sigma^{-2}) + c + (1-\mu^2)\sigma^{-1}](\sigma^{-1}-a)^{-1}.$$

Thus

$$d\hat{x}=(\mu-\hat{x})\sigma^{-1}dt$$

$$+[(\hat{x}(b+2\mu\sigma^{-2})+c+(1-\mu^2)\sigma^{-1})(\sigma^{-1}-a)^{-1}-\hat{x}^2][dy-\hat{x}dt]$$

together with (3.2)–(3.3) constitutes a finite-dimensional system for $\hat{x}(t)$. □

Remark It is also clear that $\widehat{f(x(t))x^n(t)}$ will be FDC for any n; simply eliminate $f(x(t))$ in favor of polynomials in $x(t)$ by integration by parts as in the proof above. In the same way, filters may be constructed for any conditional estimate in the infinite set of moment equations generated by starting with $\hat{x}(t)$ and using (3.1).

3.2 Smoothing

PROPOSITION 3.6 *Let $s<t$. Then*

$$E\{x(s)|\mathscr{F}_t^y\}=\frac{\sinh\kappa s}{\sinh\kappa t}[\hat{x}(t)-xe^{-\kappa t}+(P(t,s)v)_1]+xe^{-\kappa s}-(P(t,s)v)_2.$$

Proof This is immediately consequent from Corollary 2.9 once it is noted that

$$r_{11}^{-1}(t)Q_{21}=r_{11}^{-1}(t)\operatorname{cov}(\xi_1(t),\xi_1(s))=\frac{\sinh\kappa s}{\sinh\kappa t}$$

and

$$m_1(t)=xe^{-\kappa t}.$$

Explicit formulae for $(P(t,s)v)_1$ and $(P(t,s)v)_2$ are easily found and will be given in the next section. □

3.3 Polynomial functionals

Let $\eta(t)$ be any non-anticipating functional of the signal process of the form

$$=(t)\eta\int_0^t\cdots\int_0^{s_{n-1}}\gamma(s_1,\ldots,s_n)x^{k_1}(s_1)\ldots x^{k_n}(s_n)ds_n\ldots ds_1$$

where $k_1,\ldots,k_n$ are arbitrary, non-negative integers and $\gamma(s_1,\ldots,s_n)$ is a separable function. Also let $\hat{\eta}(t)=E\{\eta(t)|\mathscr{F}_t^y\}$.

PROPOSITION 3.7 *$\hat{\eta}(t)$ is FDC.*

For simplicity we restrict attention to the case $k_1=k_2=\ldots=k_n=1$; the method of proof extends easily to general choices for the k_i's. Our proof relies on the following identity, presented here in the form that it appears in Marcus and Willsky [7].

LEMMA 3.8 *Let $(u_1,\ldots,u_m)$ be a normal random vector with $e_j=Eu_j$ and $V_{ij}=\operatorname{cov}(u_i,u_j)$. Then*

$$E[u_1\ldots u_m]=\prod_1^m e_j+\sum V_{j_1j_2}e_{j_3}\ldots e_{j_m}+\sum V_{j_1j_2}V_{j_3j_4}e_{j_5}\ldots e_{j_m}+\ldots.$$

The sums are taken over all possible combinations of pairs of indices.

LEMMA 3.9 *Let $s_1>\ldots>s_n$. Then*

$$E\{x(s_1)\ldots x(s_n)|\mathscr{F}_t^y,x(t)\}=\sum_{j=0}^n x^j(t)a_j(t,s_1,\ldots,s_n)$$

for some separable functions $a_j(t,s_1,\ldots,s_n)$, $0\leq j\leq n$, depending on $y(\cdot)$.

Proof Let $l_j=M_j^{(2)}-Pv_j^{(2)}+r_{11}^{-1}(t)Q_{12}(t,\ldots,s_n)[x(t)-m_1(t)+(Pv)_1]$

$$\tilde{Q}=Q_{22}-r_{11}^{-1}(t)Q_{12}Q_{21}$$

and apply Lemma 3.8 and Corollary 2.9. Thus

$$E\{x(s_1)\ldots x(s_n)|x(t),\mathscr{F}_t^y\}=l_1\ldots l_n+\sum\tilde{Q}_{j_1j_2}l_{j_3}\ldots l_{j_n}$$
$$+\sum\sum\tilde{Q}_{j_1j_2}\tilde{Q}_{j_3j_4}l_{j_5}\ldots l_{j_n}+\ldots.$$

This is an nth order polynomial in $x(t)$ since l_j is a linear function of $x(t)$ for each j. Moreover, it is clear that the coefficients a_j will be separable if $P_0(t,s_1,\ldots,s_n)$ and hence $Q(t,s_1,\ldots,s_n)$ are separable (Q is a submatrix of P_0). But

$$P_0(t,s_1,\ldots,s_n)=\begin{bmatrix} R(t,t) & R(t,s_1) & \ldots & R(t,s_n)\\ R(s_1,t) & R(s_1,s_1) & \ldots & R(s_1,s_n)\\ \vdots & & & \\ R(s_n,t) & \ldots & & R(s_n,s_n)\end{bmatrix}$$

and $R(t,s)=\Phi_A(t,s)R(s,s)=\Phi_A(t,0)\Phi_A^{-1}(s,0)R(s)$ where $\Phi_A(t,s)$ is the state transition matrix of $A(t)$. Thus $P_0(t,\ldots,s_n)$ is indeed separable. □

Remark Upon further inspection of the terms $\Phi_A(t,s)$ and $R(s)$, the proof of this lemma demonstrates that a_j may be written

$$a_j(t,s_1,\ldots,s_n)=\sum_i \beta^i_{j,1}(t)\beta^i_{j,2}(s_1)\ldots\beta^i_{j,n}(s_n) \tag{3.10}$$

where each $\beta^i_{j,k}$, $0\leq k\leq n$ is either deterministic or a non-anticipating functional of the observation process $y(\cdot)$.

Proof of proposition By using Lemma 3.9

$$E\{\eta(t)|\mathscr{F}^y_t\}$$

$$=\int_0^t\ldots\int_0^{s_{n-1}}\gamma_1(s_1)\ldots\gamma_n(s_n)E\{E[x(s_1)\ldots x(s_n)|\mathscr{F}^y_t,x(t)]|\mathscr{F}^y_t\}ds_n\ldots ds_1$$

$$=\sum_{j=0}^n \hat{x}^j(t)\int_0^t\int_0^{s_{n-1}}\gamma(s_1,\ldots,s_n)a_j(t,s_1,\ldots,s_n)ds_n\ldots ds_1.$$

To complete the proof, it is only necessary to show that the coefficients of $\hat{x}^j(t)$ are FDC. Each coefficient is a sum of terms of the form

$$u_{n+1}(t)=\alpha_0(t)\int_0^t\ldots\int_0^{s_n}\alpha_1(s_1)\ldots\alpha_n(s_n)ds_n\ldots ds_1$$

which can be computed on-line by the system

$$\begin{aligned}
&\dot{u}_1(t)=\alpha_n(t) && u_1(0)=0\\
&\dot{u}_2(t)=\alpha_{n-1}(t)u_1(t) && u_2(0)=0\\
&\quad\vdots\\
&\dot{u}_n(t)=\alpha_1(t)u_n(t) && u_n(0)=0\\
&\dot{u}_{n+1}(t)=\alpha_0(t)u_n(t) && u_{n+1}(0)=0.
\end{aligned}$$

Thus $u_{n+1}(t)$ will be FDC if each $\alpha_i(t)$ is FDC. However, reasoning from the remark after Lemma 3.9, each $\alpha_i(t)$ will be either deterministic, or a deterministic function multiplied by one of the $y(\cdot)$-dependent $\beta^i_{j,k}(t)$ from (3.10). Now by the proof of Lemma 3.9 these $\beta^i_{j,k}(t)$ come from the $y(\cdot)$-dependent terms in l_i, $1\leq i\leq n$, and $\tilde{Q}$. Actually, inspection reveals that (i)

Q is independent of $y(\cdot)$, since it is a function of the joint covariance $\xi_1(t), \ldots, \xi_1(s_n)$; and (ii) the only y-dependence in the l_i lies in Pv. Recall from Section 2, that a typical element of Pv is $\operatorname{cov}(\xi_1(s), \xi_2(t)-\xi_3(t))$; a simple calculation shows this equals

$$\operatorname{cov}(\xi_1(s), \xi_2(s)-\xi_3(s)) - \kappa^{-1} \sinh \kappa s \int_0^t e^{-\kappa u} g(u)du$$
$$-\kappa^{-1} \operatorname{sink} \kappa s \int_0^s e^{-\kappa u} g(u)du \tag{3.11}$$

where $g(u)=\kappa y(u)-\frac{1}{2}b$. The $y(\cdot)$ functionals in this expression are certainly FDC and thus the components $\beta^i_{j,k}(t)$ are FDC. This completes the proof. □

The proof of Proposition 3.7 is similar to the proof of the linear case due to Marcus and Willsky [7] in its use of Gaussian moment identities. Actually, because conditional Gaussianity obtains in the linear case without first conditioning on $x(t)$, Marcus and Willsky are able to use the general filtering equation and a simpler moment identity than Lemma 3.8 to build a proof by induction on the order of $\eta(t)$. By "general filtering equation" is meant the representation of Fujisaki, Kallianpur, Kunita [2]:

$$d\hat{\eta} = E\left\{x^{k_1}(t)\int_0^t \ldots \int_0^{s_{n-2}} \gamma(t,..)x^{k_2}(s_1)\ldots x^{k_n}(s_{n-1})ds_{n-1}\ldots ds_1 \middle| \mathscr{F}^y_t\right\}dt$$
$$+[\widehat{\eta(t)x(t)} - \hat{\eta}(t)\hat{x}(t)][dy(t)-\hat{x}(t)dt]. \tag{3.12}$$

In our proof, application of (3.12) is superfluous, although, when calculating the filter in a particular example, it can be employed to advantage. The example below will illustrate the possibilities.

Marcus, *et al* [6] give an alternate proof of finite dimensional computability when f is linear by using homogeneous chaos theory and multiple integral expansions. Such an approach might also be possible here by first conditioning on $x(t)$, but this is not pursued, since the calculation would ultimately be like the one here.

Example Consider the model

$$dx(t)=f(x(t))dt+db(t) \qquad x(0)=x_0$$
$$d\eta(t)=x^2(t)dt \qquad \eta(0)=0$$

$$dy(t)=x(t)dt+dw(t) \qquad y(0)=0$$

where the $x(\cdot)$ and $y(\cdot)$ equations are as in (1.1)–(1.2). We will present a finite dimensional system for computing $\hat{\eta}(t)=E\{\int_0^t x^2(s)ds|\mathscr{F}_t^y\}$. This problem was chosen in part by way of comparison to the special case $f\equiv 0$ which is treated in detail in Liu and Marcus [4]. The system given here for $\hat{\eta}(t)$ is, of course, one among many possibilities; our construction was guided by the decision to use Ito equations driven by the innovations $dv(t):=dy(t)-\hat{x}(t)dt$.

Let

$$u(t)=(\sinh \kappa t)^{-2}\int_0^t(\sinh \kappa s)^2 ds=(\sinh \kappa t)^{-1}[(4\kappa)^{-1}\sinh \kappa t-\tfrac{1}{2}t].$$

Then

$$\left.\begin{aligned}
d\hat{\eta}&=\widehat{x^2}dt+u(t)[\widehat{x^3}-\widehat{x^2}\hat{x}]dv\\
&\quad+2[\widehat{x^2}-\hat{x}^2][u(t)(v_2(t)\kappa^{-1}\sinh \kappa t+\frac{\sigma}{1+\kappa\sigma}(\sigma y(t)-\mu))\\
&\quad+v_1(t)\sinh \kappa t-(\kappa \sinh \kappa t)^{-1}v_3(t)]dv\\
\hat{\eta}(0)&=0\\
dv_1&=\sinh \kappa t(\mu-\sigma y(t))\frac{\sigma}{1+\kappa\sigma}dt \qquad v_1(0)=0\\
dv_2&=e^{-\kappa t}[\kappa y(t)-\tfrac{1}{2}b]dt \qquad v_2(0)=0\\
dv_3&=(\sinh \kappa t)^2 v_2(t)dt \qquad v_3(0)=0.
\end{aligned}\right\} \tag{3.13}$$

(3.13) is not the complete system, since equations for $\mu, \sigma, \hat{x}, \widehat{x^2}$, and $\widehat{x^3}$ are also needed. However, these are easily garnered from Section 3. A brief derivation of (3.13) follows. From (3.12)

$$d\hat{\eta}=\widehat{x^2}dt+(\widehat{\eta x}-\hat{\eta}\hat{x})dv. \tag{3.14}$$

However, Corollary 2.9 implies

$$\begin{aligned}
E\{x^2(s)|\mathscr{F}_t^y, x(t)\}&=Q_{22}-r_{11}^{-1}(t)Q_{21}Q_{12}+E^2\{x(s)|\mathscr{F}_t^y, x(t)\}\\
&=F(t,s)+r_{11}^{-2}(t)Q_{21}^2x^2(t)\\
&\quad+2x(t)r_{11}^{-1}(t)Q_{21}[m_1(s)-(Pv)_2-r_{11}^{-1}(t)Q_{21}(m_1(t)-Pv_1)]
\end{aligned} \tag{3.15}$$

where $F(t,s)$ combines those terms not depending on $x(t)$. Now

$$r_{11}^{-1}(t)Q_{21}=\sinh \kappa s/\sinh \kappa t,$$

and

$$Pv_4=Pv_1(s)-\kappa^{-1}\sinh \kappa s\left[\int_0^t e^{-\kappa u}g(u)du-\int_0^s e^{-\kappa u}g(u)du\right].$$

(Recall that $g(u)=\kappa y(u)-1/2\,b$.) Further, it can be shown

$$m_1(t)-Pv_1(t)=\frac{\sigma}{1+\kappa\sigma}(\mu-\sigma y)(t).$$

Using these identities in (3.15) it follows that

$$\widehat{\eta x}-\hat{\eta}\hat{x}=\int_0^t E\{(x(t)-\hat{x}(t))E\{x^2(s)|\mathscr{F}_t^y,x(t)\}|\mathscr{F}_t^y\}\,ds$$

$$=(\widehat{x^3}-\widehat{x^2}\hat{x})u(t)+2(\widehat{x^2}-\hat{x}^2)u(t)[(\sigma y-\mu)\frac{\sigma}{1+\kappa\sigma}$$

$$+\kappa^{-1}\sinh \kappa t\int_0^t e^{-\kappa u}g(u)du$$

$$+2(\widehat{x^2}-\hat{x}^2)\left[(\sinh \kappa t)^{-1}\int_0^t(\sinh \kappa s)[m_1(s)-Pv_1(s)]ds\right.$$

$$\left.-(\kappa\sinh \kappa t)^{-1}\int_0^t(\sinh \kappa s)^2\int_0^s e^{-\kappa u}g(u)duds\right].$$

$$=(\widehat{x^3}-\widehat{x^2}\hat{x})u(t)+2(\widehat{x^2}-\hat{x}^2)\left[u(t)(v_2(t)\kappa^{-1}\sinh \kappa t+\frac{\sigma}{1+\kappa\sigma}(\sigma y(t)-\mu))\right.$$

$$\left.+v_1(t)\sinh \kappa t-(\kappa\sinh \kappa t)^{-1}v_3(t)\right] \tag{3.16}$$

Placing (3.16) in (3.14) one obtains the desired result.

References

[1] V. E. Beneš, Exact finite dimensional filters for certain diffusions with nonlinear drift, *Stochastics* **5** (1981), 65–92.

[2] M. Fujisaki, G. Kallianpur and H. Kunita, Stochastic differential equations for the non linear filtering problem, *Osaka J. Math.* **9** (1972), 19–40.

[3] A. H. Jazewinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.

[4] C.-H. Liu and S. I. Marcus, The Lie algebraic structure of a class of finite dimensional nonlinear filters, in *Lectures in Applied Mathematics* **18** (1980), 277–297.

[5] S. I. Marcus, S. K. Mitter and D. Ocone, Finite dimensional nonlinear estimation for a class of systems in continuous and discrete time, in *Analysis and Optimisation of Stochastic Systems* (Proc. Internat. Conf. Univ. Oxford, Oxford, 1978), 387–406, Academic Press, London, 1980.

[6] S. I. Marcus and A. S. Willsky, Algebra structure and finite dimensional nonlinear estimation, *SIAM J. Math. Anal.* **9** (1978), 312–327.

[7] D. Ocone, Finite dimensionally computable statistics and estimation algebras in nonlinear filtering, *Proceedings, 1981 Int. Sym. Mathematical Theory of Networks and Systems*, Santa Monica, CA, Aug. 5–7, 1981.

[8] R. W. Brockett, Nonlinear Systems and Nonlinear Estimation Theory, in Hazewinkel and Willems, ed., Stochastic Systems: *The Mathematics of Filtering and Identification and Applications*, NATO Advanced Study Institute Series, Reidel, Dordrecht, 1981.