*Stochastics*, Vol. 23, pp. 241–275
Photocopying permitted by license only

Printed in Great Britain



# A Conditionally Almost Linear Filtering Problem with Non-Gaussian Initial Condition*

U. G. HAUSSMANN
*Mathematics Department, University of British Colombia, Vancouver, Canada V6T 1Y4*

and

E. PARDOUX
*U.E.R. de Mathématiques, Université de Provence, 13331 Marseille Cedex 3, France*

*(Received 17 February 1986; in final form 4 August 1987)*

We derive a formula for computing explicitly the optimal non-linear filter for a class of problems which admit finite dimensional filters. The result includes all known results of this kind as special cases.

KEY WORDS: Non-linear filtering, conditionally Gaussian, finite-dimensional filters.

*This work was supported by the N.S.E.R.C. of Canada under grant A8051, and was carried out while the second author was visiting the University of British Columbia.

# 1. INTRODUCTION

Since recent results (cf. Hazewinkel and Marcus [5], Chaleyat-Maurel and Michel [4], Ocone [12], among others) tend to indicate that there are very few non-linear filtering problems with optimal filters which depend only on a finite number of statistics, those classes of problems which admit such computable finite dimensional filters are very important in practice since they might be implemented without further approximation.

Essentially three such classes have been discovered to date, and they are all modifications of the classical linear-Gaussian filtering problem solved by Kalman and Bucy [8]:

$$\begin{aligned} dX_t &= AX_t\,dt + B\,dW_t, \\ dY_t &= HX_t\,dt + L\,dW_t, \end{aligned} \tag{1.1}$$

where $\{X_t:t\geqq 0\}$ is the unobserved process which is to be filtered, $\{Y_t:t\geqq 0\}$ is the observed process, $A$, $B$, $H$, $L$ are matrices, possibly depending on $t$, $\{W_t:t\geqq 0\}$ is a standard multidimensional Wiener process independent of the Gaussian random vector $X_0$, while $Y_0=0$.

The first class is the "conditionally Gaussian" problem of Liptser and Shiryayev [10], where $A$, $B$, $H$, $L$ in (1.1) are allowed at each time $t$ to depend on the past of $\{Y_s\}$ up to time $t$.

The second class if the "Beneš problem", cf. Beneš [1, 2], in which the linear drift $AX_t$ is replaced by $f(X_t)$ where $f$ satisfies a particular condition, and in addition the signal $\{X_t\}$ and the observation noise are uncorrelated, i.e. $B=(B_1,0)$, $L=(0,L_2)$.

The last class consists of problems with the linear dynamics (1.1), but with non-Gaussian initial condition. It was studied by Beneš and Karatzas [3] and by Makowski [11], again in the uncorrelated case.

It has become apparent very recently that one can obtain new classes of explicitly solvable non-linear filtering problems by mixing the above cases. Kolodziej and Mohler [9] have considered "conditionally linear" filtering problems with non-Gaussian initial condition (see also a particular case of this situation in Rishel [13]) and Shukhman [14] has treated the "Beneš problem" with non-Gaussian initial condition. Note that Shukhman [14] as well as Zeitouni and Brobrovsky [16] have generalized the "Beneš problem" to include the original Kalman–Bucy filter as a special case.

The aim of our work is to generalize and to give a unified treatment of all of the above-mentioned results. In Section 2 we generalize the conditionally Gaussian result of Liptser and Shiryayev [10]. In Section 3 we add to the system considered in the previous section a non-linear drift of the "Beneš type", and in Section 4 we allow in addition a non-Gaussian initial condition.

The class of problem which we can treat is rather particular and not as general as this introduction might lead one to think; in particular, as noted already by Zeitouni and Bobrovsky [16] and by Shukhman [14], in the multidimensional case it does not seem possible to add an arbitrary linear drift to a non-linear one satisfying Beneš' condition and to obtain an explicit finite dimensional filter. Nevertheless we do generalize all previous work known to us, in particular because *we allow correlation between the signal* $\{X_t\}$ *and the observation noise.* The essential tool which permits us to do so is our generalized conditionally Gaussian filtering theorem, where in the $\{X_t\}$ dynamics we allow a term consisting of a linear function of $X_t$ multiplied by $dY_t$. Our results do include for instance the conditionally linear filtering problem with non-Gaussian initial condition and with correlation between the signal $\{X_t\}$ and the observation noise.



## 2. CONDITIONAL GAUSSIAN PROCESSES

In this section we show under reasonable hypotheses on the data that if $\{X_t\}$, $\{Y_t\}$ are two processes satisfying

$$dX_t = [A(t,Y)X_t + a(t,Y)]\,dt + B(t,Y)\,dW_t + \sum_{j=1}^{d} [G^j(t,Y)X_t + g^j(t,Y)]\,dY_t^j, \tag{2.1}$$

$$dY_t = [H(t,Y)X_t + h(t,Y)]\,dt + dU_t, \qquad Y_0 = 0, \tag{2.2}$$

and $X_0$ is Gaussian, then $X_t$ is conditionally Gaussian given $\mathscr{Y}_t$, the $\sigma$-algebra generated by $\{Y_s : s \leqq t\}$. We also compute equations satisfied by the conditional mean $m_t \equiv E\{X_t | \mathscr{Y}_t\}$ and by the conditional covariance $R_t \equiv E\{(X_t - m_t)(X_t - m_t)^* | \mathscr{Y}_t\}$. Note that * denotes transpose.

Let us be more precise. We assume:

($A_1$) $\{X_t : t \geqq 0\}$ is an adapted a.s. continuous $\mathbb{R}^N$-valued process, $\{Y_t : t \geqq 0\}$ is an adapted, a.s. continuous $\mathbb{R}^d$-valued process and $\{W_t : t \geqq 0\}$, $\{U_t : t \geqq 0\}$ are given independent $\mathbb{R}^M$ and $\mathbb{R}^d$-valued (respectively) standard Wiener processes on a given filtered probability space $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t \geqq 0}, P)$, such that (2.1) and (2.2) hold a.s. Moreover $X_0$ is Gaussian with distribution $N(m_0, R_0)$.

We write $C(\mathbb{R}^+, \mathbb{R}^d)$ for the space of continuous functions $[0, \infty) \to \mathbb{R}^d$ under the topology of uniform convergence on compact sets. Let $\{\mathscr{G}_t\}_{t \geqq 0}$ denote the canonical Borel filtration on $C(\mathbb{R}^+; \mathbb{R}^d)$.

($A_2$) $A$, $a$, $B$, $G^j$, $g^j$, $H$, $h$ are all defined on $\mathbb{R}^+ \times C(\mathbb{R}^+; \mathbb{R}^d)$ and assume values (respectively) in $\mathbb{R}^N \otimes \mathbb{R}^N$, $\mathbb{R}^N$, $\mathbb{R}^N \otimes \mathbb{R}^M$, $\mathbb{R}^N \otimes \mathbb{R}^N$, $\mathbb{R}^N$, $\mathbb{R}^d \times \mathbb{R}^N$, $\mathbb{R}^d$. Moreover they are $\mathscr{G}_t$ progressively measurable.

($A_3$) If $\alpha$ is any one of $|A|^2$, $|a|$, $|B|^2$, $|G^j|^2$, $|H|$, $|h|$, then $\alpha(\cdot, y)$ is in $L^1_{\text{loc}}(\mathbb{R}^+)$ for each $y$ in $C(\mathbb{R}^+; \mathbb{R}^N)$ where $|\cdot|$ denotes the norm in the appropriate space.

($A_4$) For each $T$ in $\mathbb{R}^+$, $E\Lambda_T^{-1} = 1$ where

$$\Lambda_T = \exp\left\{\int_0^T [H(s, Y)X_s + h(s, Y)]^* \, dY_s - \frac{1}{2}\int_0^T |H(s, Y)X_s + h(s, Y)|^2 \, ds\right\}. \tag{2.3}$$

It follows that $E_T \Lambda_T = E1 = 1$, where $E_T$ is expectation under $P_T$ and $P_T$ is defined by

$$P_T(A) = \int_A \Lambda_T^{-1} \, dP, \qquad A \in \mathscr{F}_T.$$

*Remark 2.1* One might consider replacing $dU_t$ by $L(t, Y)\, dU_t$ in (2.2). However to make the method work, $L(t, Y)$ would have to be non-singular and Lipschitz in $Y$. But in that case no generality is gained for we could define

$$\tilde{Y}_t = \int_0^t L(s, Y)^{-1} \, dY_s.$$

Then $\tilde{Y}$ satisfies an equation of the form (2.2). Moreover from the definition of $\tilde{Y}$ we see that $\tilde{\mathscr{Y}}_t \subset \mathscr{Y}_t$, and that

$$dY = L(s, Y)\, d\tilde{Y}.$$

The Lipschitz continuity of $L$ implies that $Y$ is a strong solution of this equation, hence $\mathscr{Y}_t \subset \tilde{\mathscr{Y}}_t$ and we conclude that $\tilde{\mathscr{Y}}_t = \mathscr{Y}_t$.

*Remark 2.2* We might take $Y_0$ random and replace the last part of $(A_1)$ by:

The conditional law of $X_0$ given $\mathscr{Y}_0$ is Gaussian, $N(m_0(Y_0), R_0(Y_0))$. This extension is trivial and we leave it to the reader.

*Remark 2.3* Putting $dY$ in (2.1) is just a convenient way of expressing correlation between the noise in (2.1) and (2.2). In this form it is obvious that (2.1) is conditionally linear given $Y$. Moreover, even if no such term is present in the model of the next two sections, the method of solution will reduce the original system to one of the form (2.1), (2.2) with non-zero $G$ and $g$.

Let us now give two examples where the condition $(A_4)$ is satisfied.

*Example 2.1* Assume $(A_1)$–$(A_3)$ and

$$X = \begin{pmatrix} x^1 \\ x^2 \end{pmatrix}, \quad A = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}, \quad G^j = \begin{pmatrix} 0 & 0 \\ G^j_{21} & G^j_{22} \end{pmatrix}, \quad H = (H_1, 0),$$

$$B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}, \quad a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \quad g^j = \begin{pmatrix} g^j_1 \\ g^j_2 \end{pmatrix}$$

and for all $T < \infty$

$$\sup_{0 \leqq t \leqq T} \sup_y |\alpha(t, y)| < \infty \tag{2.4}$$

where $\alpha$ is any one of $|H_1|, |h|$ or any of the coefficients appearing in the equation for $X^1$. Now the equations for $(X^1, Y)$ can be rewritten as

$$dX^1 = \left[\left(A_{11} + \sum_j g^j_1 H_{1j}\right) X^1 + (a_1 + \sum_j g^j_1 h^j)\right] dt + B_1\, dW + \sum_j g^j_1\, dU^j,$$

$$dY = [H_1 X^1 + h]\, dt + dU$$

where $H_{1j}$ is the $j$th column of $H_1$. They have drift with linear growth and bounded diffusion coefficients. Because of the special form of $H$, $\Lambda_T^{-1}$ is determined solely by $(X^1, Y)$, so that Corollary 7.2.2 of Kallianpur [7] implies that $(A_4)$ holds.

*Example 2.2* Assume $(A_1)$–$(A_3)$ and

$$X=\begin{pmatrix} X^1 \\ X_2 \end{pmatrix}, \quad A=\begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}, \quad B=\begin{pmatrix} B_{11} & 0 \\ 0 & B_{22} \end{pmatrix}, \quad W=\begin{pmatrix} W^1 \\ W^2 \end{pmatrix},$$

$$G^j=\begin{pmatrix} 0 & 0 \\ G^j_{21} & 0 \end{pmatrix}, \quad g^j=\begin{pmatrix} 0 \\ g^j_2 \end{pmatrix}$$

and that all coefficients in the equation for $X^1$ are constant in the second variable $(Y)$. Then $X^1$ does not explode. Assume also that $|H|$, $|h|$ and the coefficients in the equation for $X^2$ satisfy (2.4). Let $\mathscr{X}^1_t$ be the $\sigma$-algebra generated by the past of $X^1$ and let $\mathscr{W}^1_t$ be that generated by $W^1$. Now $\{X^1_t\}$ is only driven by $\{W^1_t\}$ so $\mathscr{X}^1_t \subset \mathscr{W}^1_t \vee \mathscr{X}^1_0$ and hence $\{(W^2_t, U_t):0 \leqq t \leqq T\}$ (respectively $\{W^2_t, Y_t):0 \leqq t \leqq T\}$) remains a Wiener process on $(\Omega, \mathscr{F}, P)$ (respectively on $(\Omega, \mathscr{F}, P_T)$) given $\mathscr{X}^1_T$. But given $\mathscr{X}^1_T$, the equations for $(X^2, Y)$ have affine drift and bounded diffusion coefficients a.s., so for each sample path of $X^1$ we have

$$E\{\Lambda_T^{-1} | \mathscr{X}^1_T\} = 1,$$

and hence $(A_4)$ holds by integrating out $X^1$.

Observe that this example includes the case when $X^1$ is a parameter, i.e.

$$dX^1 = 0$$

$$dX^2 = (A_{21}X^1 + A_{22}X^2 + a)\,dt + B\,dW + \sum_j (G^j_{21}X^1 + g^j)\,dY^j.$$

We will give now the main result of this section.

THEOREM 2.1 *Assume* $(A_1)$–$(A_3)$. *Then for any* $T<\infty$ *the conditional distribution of* $X_T$ *given* $\mathscr{Y}_T$ *is Gaussian.*

*Proof* We need to show that for any $z$ in $\mathbb{R}^N$

$$E\{\exp(iz^*X_T) | \mathscr{Y}_T\} = \exp[i\alpha^* z - z^*\beta z] \quad \text{a.s.} \tag{2.5}$$

where $\alpha$ and $\beta$ are $\mathscr{Y}_T$ measurable random variables assuming values in $\mathbb{R}^N$ and in the $N \times N$ symmetric positive semi-definite matrices respectively. It will be easier to work with $P_T$ so observe that

$$E(e^{iz^*X_T} | \mathscr{Y}_T) = E_T(e^{iz^*X_T}\Lambda_T | \mathscr{Y}_T) E_T(\Lambda_T | \mathscr{Y}_T)^{-1},$$

and hence we need to show that

$$E(e^{iz^*X_T}\Lambda_T | \mathscr{Y}_T) = k \exp(i\alpha^* z - z^* \beta z) \quad \text{a.s.} \tag{2.6}$$

for some $\mathscr{Y}_T$ measurable scalar random variable $k$. The left side of (2.6) is $E_T(\xi | \mathscr{Y}_T)$ with

$$\xi = \exp\left\{ iz^*X_T + \int_0^T [H(s, Y)X_s + h(s, Y)]^* \, dY_s - \tfrac{1}{2}\int_0^T |H(s, Y)X_s + h(s, Y)|^2 \, ds \right\}.$$

Now let $\Phi(t)$ be the unique strong matrix valued solution of

$$d\Phi = A(t, Y)\Phi \, dt + \sum_j G^j(t, Y)\Phi \, dY_t^j, \qquad 0 \leqq t \leqq T,$$

$$\Phi(0) = I.$$

It exists on $(\Omega, \mathscr{F}, P_T)$ by the result of Jacod [6], and is invertible since

$$\det \Phi(t) = \det \Phi(0) \exp\left\{ \int_0^t \operatorname{tr}\left[ A(s, Y) - \tfrac{1}{2}\sum_j G^j(s, Y)^2 \right] ds + \sum_j \int_0^t \operatorname{tr}[G^j(s, Y)] \, dY_s^j \right\}$$

$$\neq 0,$$

where $\operatorname{tr} A$ is the trace of $A$ and $\det \Phi$ is the determinant of $\Phi$. It follows that

$$X_t = \Phi(t)\left\{X_0 + \int_0^t \Phi(s)^{-1} B(s, Y)\, dW_s + \int_0^t \Phi(s)^{-1} a(s, Y)\, ds + \sum_j \int_0^t \Phi(s)^{-1} g^j(s, Y)\, dY_s^j\right\}.$$

If we define

$$\eta_t = X_0 + \int_0^t \Phi(s)^{-1} B(s, Y)\, dW_s,$$

$$\gamma_t = \int_0^t \Phi(s)^{-1} a(s, Y)\, ds + \sum_j \int_0^t \Phi(s)^{-1} g^j(s, Y)\, dY_s^j,$$

then we have

$$X_t = \Phi(t)\{\eta_t + \gamma_t\}.$$

If we substitute this expression for $X_t$ into $\xi$ we see that we can factor $\xi$ into two parts, $\xi = \xi_1 \xi_2$, where $\xi_1$ is $\mathscr{Y}_T$ measurable. Specifically if

$$Q(s) = \Phi(s)^* H(s, Y)^* H(s, Y) \Phi(s)$$

then

$$\xi_1 = \exp\left\{iz^*\Phi(T)\gamma_T + \int_0^T [H(x, Y)\Phi(s)\gamma_s + h(s, Y)]^*\, dY_s - \tfrac{1}{2}\int_0^T |H(s, Y)\Phi(s)\gamma_s + h(s, Y)|^2\, ds\right\},$$

$$\xi_2 = \exp\left\{iz^*\Phi(T)\eta_T + \int_0^T [H(s, y)\Phi(s)\eta_s]^*\, dY_s - \int_0^T [H(s, Y)\Phi(s)\gamma_s + h(s, Y)]^*[H(s, Y)\Phi(s)\eta_s]\, ds\right.$$

$$-\tfrac{1}{2}\int_0^T \eta_s^* Q(s)\eta_s\,ds\Big\},$$

$$=\exp\Big\{iz^*\Phi(T)\eta_T+\int_0^T [H(s,Y)\Phi(s)\eta_s]^*\,dY_s$$

$$-\int_0^T [\gamma_s^* Q(s)\eta_s+h(s,y)^* H(s,Y)\Phi(s)\eta_s+\tfrac{1}{2}\eta_s^* Q(s)\eta_s]\,ds\Big\}.$$

Upon substituting for $\eta$ in this last expression (except in the term which is quadratic in $\eta$), interchanging the order of integration $dW$, $dY$ in $\int (H\Phi\eta)^*\,dY$ which is permitted since $W$ and $Y$ are independent under $P_0$ (hence the integrands remain non-anticipating), and interchanging the order of integration $dW$, $ds$, we obtain

$$\xi_2=\exp\Big\{iz^*\Phi(T)\Big[X_0+\int_0^T \Phi(s)^{-1}B(s,Y)\,dW_s\Big]$$

$$+X_0^*\Big[\int_0^T \Phi(s)^* H(s,Y)^*[dY_s-h(s,Y)\,ds]-\int_0^T Q(s)\gamma_s\,ds\Big]$$

$$+\int_0^T\Big(\int_s^T \Phi(u)^* H(u,Y)^*[dY_u-h(u,Y)\,du]-\int_s^T Q(u)\gamma_u\,du\Big)^*$$

$$\times\,\Phi(s)^{-1}B(s,Y)\,dW_s-\tfrac{1}{2}\int_0^T \eta_s^* Q(s)\eta_s\,ds\Big\}$$

$$=\exp\Big\{[i\Phi(T)^*z+\alpha_1(T)]^*X_0+(i\Phi(T)^*z)^*\int_0^T \Phi(s)^{-1}B(s,Y)\,dW_s$$

$$+\int_0^T \alpha_2(T,s)^*\,dW_s-\tfrac{1}{2}\int_0^T \eta_s^* Q(s)\eta_s\,ds\Big\}$$

where $\alpha_1(T)$, $\alpha_2(T,s)$ are $\mathscr{Y}_T$ measurable random vectors. We can now apply Lemma 1 of the Appendix to obtain

$$E_T(\xi\,|\,\mathscr{Y}_T)=\xi_1 E_T(\xi_2\,|\,\mathscr{Y}_T)$$

$$=\xi_1 k_1\exp[\alpha_3(T)^* b(T)+b(T)^*\beta_1(T)b(T)]$$

where the $2N+1$ dimensional random vector $b(T)$ is

$$b(T)=\begin{bmatrix} i\Phi^*(T)z+\alpha_1(T) \\ i\Phi^*(T)z \\ 1 \end{bmatrix}.$$

Hence

$$E_T(\xi|\mathscr{Y}_T)=\xi_1 k_2 \exp\{i[\Phi(T)\alpha_4(T)]^* z - z^*\Phi(T)\beta_2(T)\Phi(T)^* z\}$$

where $\beta_2(T)=(I,I,0)\beta_1(T)(I,I,0)^*\geqq 0$ and $I$ is the $N\times N$ identity matrix. Since $\xi_1$ has the form

$$\xi_1=\exp\{i[\Phi(T)\alpha_5(T)]^* z+\alpha_6(T)\}$$

then the result follows. □

COROLLARY 2.1 *As functions of $T$, $\alpha$ and $\beta$ are a.s. locally bounded.*

*Proof* It is readily seen that all the $\alpha$'s and $\beta$'s in the above proof are continuous in $T$, hence the result follows. □

*Remark 2.4* We have only shown that $X_T$ given $\mathscr{Y}_T$ is Gaussian since this is all that we require, but the same proof shows that the distribution of $X_{t_1}$, $X_{t_2},\ldots,X_{t_n}$ given $\mathscr{Y}_T, t_1<t_2<\cdots<t_n<T$, is Gaussian. □

Next we wish to compute the conditional mean $m(t)$ and conditional covariance $R(t)$ of $X_t$—they are of course $\alpha(t)$ and $2\beta(t)$ of Theorem 2.1, but we want to obtain recursive formulae for $m$ and $R$, i.e. we want to derive equations driven by the observation $Y$, which are satisfied by $m$ and $R$. To do so we employ the Kushner–Stratonovich equation (cf. Liptser and Shiryayev [10, Theorem 8.1]). However we must add new hypotheses:

($A_5$) For each $T<\infty$, $\alpha(t, Y(\omega))$ is in $L^2((0,T)\times\Omega, dt\times dP)$ where $\alpha$ is any one of

$$|a|,\quad |g^j|,\quad |h|,\quad |h||g^j|,\quad |A|,\quad |B|,\quad |G^j|^2,$$

$$|H|^2,\quad |h||G^j|,\quad |g^j||H|.$$

($A_6$) For each $T$ in $\mathbb{R}^+$ $|H(t, Y(\omega))X_t(\omega)+h(t, Y(\omega))|$ is in $L^2([0, T]\times\Omega, dt\times dP)$.

Note that ($A_6$) holds for our two examples if all coefficients satisfy (2.4). Moreover ($A_5$) can easily be guaranteed by requiring that all the terms appearing there satisfy (2.4). We write $H^j(t, Y)$ for the $j$th row of $H(t, Y)$.

THEOREM 2.2 *Assume* ($A_1$)–($A_6$). *Then the equations*

$$d\tilde{m}_t=\left\{A(t, Y)\tilde{m}_t+a(t, Y)-\tilde{R}_tH(t, Y)^*[H(t, Y)\tilde{m}_t+h(t, Y)]\right.$$

$$\left.+\sum_j G^j(t, Y)\tilde{R}_tH^j(t, Y)^*\right\}dt$$

$$+\sum_j \{G^j(t, Y)\tilde{m}_t+g^j(t, Y)+\tilde{R}_tH^j(t, Y)^*\}\,dY_t^j \tag{2.7}$$

$$d\tilde{R}_t=\left\{B(t, Y)B(t, Y)^*+A(t, Y)\tilde{R}_t+\tilde{R}_tA(t, Y)^*\right.$$

$$\left.+\sum_j G^j(t, Y)\tilde{R}_tG^j(t, Y)^*-\tilde{R}_tH(t, Y)^*H(t, Y)\tilde{R}_t\right\}dt$$

$$+\sum_j \{G^j(t, Y)\tilde{R}_t+\tilde{R}_tG^j(t, Y)^*\}\,dY_t^j \tag{2.8}$$

*with* $\tilde{m}_0=m_0$, $\tilde{R}_0=R_0$, *cf.* ($A_1$), *have an a.s. continuous unique strong solution which is a modification of* $(m_t, R_t)$.

*Remark 2.5* Since the pair $(X_., Y_.)$ is *not* Gaussian in general it is not surprising that the conditional covariance matrix $R_t$ is random, just as in the conditionally Gaussian case treated by Liptser and Shiryayev [10, Chapter 12]; however in contradistinction to that case, now (2.8) contains the stochastic differential $dY$. On the other hand, a feature of the Kalman–Bucy filter which is preserved is the fact that (2.8) does not involve $\tilde{m}_t$. □

The proof of the theorem is divided into several lemmas. First we show that (2.7), (2.8) have a unique solution: so if $m, R$ satisfy these equations then this fact identifies them uniquely. Then in Lemma 2.2

we apply non-linear filtering theory to find the Kushner–Stratonovich equation satisfied by $E\{\psi(x_t)|\mathcal{Y}_t\}$ when $\psi$ is a smooth function of compact support. Next a limiting argument is used to obtain this equation for the cases $\psi(x)=x_i$, $\psi(x)=x_ix_j$, cf. (2.15). The last step, which is left to the reader, is to use the fact that for Gaussian random variables (and by Theorem 2.1 we know that $X_t$ is conditionally Gaussian) the third moments which appear in (2.15) can be expressed in terms of the first two moments. Now algebraic manipulation allows (2.15) to be reduced to (2.7), (2.8).

LEMMA 2.1 *Assume* $(A_1)$ *and* $(A_6)$. *Then* (2.7), (2.8) *have a unique strong solution* $\{(\tilde{m}_t, \tilde{R}_t): t \geqq 0\}$.

*Proof* Fix $M<\infty$. In Eq. (2.8) replace the term $\tilde{R}_tH^*H\tilde{R}_t$ for $|\tilde{R}_t|>M$ by $\tilde{R}_tH^*H\tilde{R}_tM/|\tilde{R}_t|$. Now a result of Jacod [6] allows us to conclude that the modified Eq. (2.8) has a unique solution. Letting $M\to\infty$ we find that (2.8) has a unique strong solution on $[0,\tau_R)$ where $\tau_R$ is the explosion time of $\tilde{R}_t$. We shall see shortly that this solution is the conditional covariance of $X_t$, hence $\tau_R=+\infty$ by Corollary 2.1.

Now the result of Jacod [6] can be applied to (2.7) to conclude again that a unique strong solution exists on $[0,\infty)$. □

Let $C(\mathbb{R}^N)$ denote the continuous real valued functions defined on $\mathbb{R}^N$ and let $C_0^2(\mathbb{R}^N)$ be the subset of those functions which have compact support and which are twice continuously differentiable. Note that if $\phi$ is in $C(\mathbb{R}^N)$ and if there exist constants $c$ and $k$ such that

$$|\phi(x)|\leqq c(1+|x|^k) \tag{2.9}$$

then by Theorem 2.1 we have

$$E\{|\phi(X_t)||\mathcal{Y}_t\}<\infty \quad \text{a.s.}$$

and we can define $\hat{\phi}_t := E\{\phi(X_t)|\mathcal{Y}_t\}$. This is also true if for each $x$, $\phi(x)=\phi_t(x)$ is a $\mathcal{Y}_t$ measurable random variable. The constant $c$ in (2.9) may now depend on $Y$.

LEMMA 2.2 *Assume* $(A_1)$, $(A_6)$. *If* $\phi\in C_0^2(\mathbb{R}^N)$ *then*

$$\hat{\phi}_t=\hat{\phi}_0+\int_0^t \widehat{(L\phi)}_s\, ds+\sum_j \int_0^t [\widehat{(K_j\phi)}_s-\hat{\rho}_s^j\hat{\phi}_s][dY_s^j-\hat{\rho}_s^j\, ds]$$

*where* $\rho_t^j(x)=\sum_i H_{ji}(t, Y)x_i+h_j(t, Y)$ *and*

$$(L\phi)_t(x)=\tfrac{1}{2}\sum_{ij}\Big\{B(t, Y)B(t, Y)^*+\sum [G^j(t, Y)x+g^j(t, Y)]$$

$$\times [G^j(t, Y)x+g^j(t, Y)]^*\Big\}_{ij}\frac{\partial^2\phi(x)}{\partial x_i\,\partial x_j}$$

$$+\sum_i\Big\{A(t, Y)x+a(t, Y)$$

$$+\sum_j \rho_t^j(x)[G^j(t, Y)x+g^j(t, Y)]\Big\}_i\frac{\partial\phi(x)}{\partial x_i}$$

$$(K_j\phi)_t(x)=\sum_i [G^j(t, Y)x+g^j(t, Y)]_i\frac{\partial\phi(x)}{\partial x_i}+\rho_t^j(x)\phi(x).$$

*Proof* The result is an immediate consequence of Liptser and Shiryayev [10, Theorem 8.1], since $\phi(X_t)$ is a bounded semimartingale and since $(A_5)$, $(A_6)$ and the fact that $\phi$ has compact support imply that the coefficients of $dt$ in the semimartingale decomposition of $\phi(X_t)$ and of $Y_t$ belong to $L_2((0, T)\times\Omega, dt\times dP)$ for all $T<\infty$. □

We complete the proof of Theorem 2.2 with the following lemma.

LEMMA 2.3 *Assume* $(A_1)$–$(A_6)$. *A modification of* $(m_t, R_t)$ *satisfies* (2.7), (2.8).

*Proof* For $i,j=1,2,\ldots,N$ introduce sequences $\{\phi^{ni}\}\subset C_0^2(\mathbb{R}^N)$, $\{\phi^{nij}\}\subset C_0^2(\mathbb{R}^N)$ such that $|\phi^{ni}(x)|\leqq|x|$, $|\phi_x^{ni}(x)|\leqq|x|$, $|\phi_{xx}^{ni}(x)|\leqq|x|$, $|\phi^{nij}(x)|\leqq|x|^2$, $|\phi_x^{nij}(x)|\leqq|x|^2$, $|\phi_{xx}^{nij}(x)|\leqq|x|^2$ and $\phi^{ni}(x)\rightarrow x_i$, $\phi^{nij}(x)\rightarrow x_ix_j$ as $n\rightarrow\infty$. According to Lemma 2.2

$$\hat{\phi}_t^{ni}=\hat{\phi}_0^{ni}+\int_0^t \widehat{(L\phi^{ni})}_s\, ds+\sum_k\int_0^t \widehat{(K_k\phi^{ni})}_s-\hat{\rho}_s^k\hat{\phi}_s^{ni}][dY_s^k-\hat{\rho}_s^k\, ds]$$

$$\hat{\phi}_t^{nij}=\hat{\phi}_0^{nij}+\int_0^t \widehat{L\phi^{nij})}_s\, ds+\sum_k\int_0^t \widehat{(K_k\phi^{nij})}_s-\hat{\rho}_s^k\hat{\phi}_s^{nij}][dY_s^k-\hat{\rho}_s^k\, ds]$$

which we rewrite as

$$\hat{\phi}_t^{ni} = \phi_0^{ni} + \int_0^t \alpha_s^{ni}\, ds + \sum_k \int_0^t \beta_{ks}^{ni}\, dY_s^k,$$

$$\hat{\phi}_t^{nij} = \phi_0^{nij} + \int_0^t \alpha_s^{nij}\, ds + \sum_k \int_0^t \beta_{ks}^{nij}\, dY_s^k, \tag{2.10}$$

where each of $\alpha^{ni}$, $\alpha_k^{ni}$, $\alpha^{nij}$, $\beta_k^{nij}$ is a finite sum of terms of the form $M(s, Y)P(\hat{X}_s)\hat{\psi}_s^n$ with $M\{\mathscr{Y}_s\}$—progressively measurable, $M$ in $L^1((0, T) \times \Omega, dt \times dP)$ for all $T < \infty$ if the term occurs in $\alpha$, $M$ in $L^2((0, T) \times \Omega, dt \times dP)$ for all $T < \infty$ if it occurs in $\beta$, and with $P(x)$ a monomial of order 0, 1 or 2 in the $x_i$. Moreover

$$|\psi^n(x)| \leqq c(1 + |x|^4) \qquad \forall n, \quad \forall x,$$

$$\psi^n(x) \to \psi(x) \qquad \forall x. \tag{2.11}$$

We wish to pass to the limit in (2.10) as $n \to \infty$.

Let us define a sequence of $\mathscr{Y}_t$-stopping times

$$\tau_m = \inf\{t \geqq 0: E\{|X_t|^4 | \mathscr{Y}_t\} \geqq m\}, \qquad m = 1, 2, \dots$$

with the convention that $\tau_m(\omega) = +\infty$ if $\sup_{t \geqq 0} E\{|X_t|^4 | \mathscr{Y}_t\}(\omega) < m$. It follows from Theorem 2.1 that

$$E\{|X_t|^4 | \mathscr{Y}_t\} = F(\alpha(t), \beta(t)) \quad \text{a.s.}$$

for some continuous function $F$, and hence from Corollary 2.1 that $\tau_m \to \infty$ a.s. as $m \to \infty$.

Now from (2.10) we obtain

$$\hat{\phi}_{t \wedge \tau_m}^{ni} = \hat{\phi}_0^{ni} + \int_0^{t \wedge \tau_m} \alpha_s^{ni}\, ds + \sum_k \int_0^{t \wedge \tau_m} \beta_{ks}^{ni}\, dY_s^k$$

$$\hat{\phi}_{t \wedge \tau_m}^{nij} = \hat{\phi}_0^{nij} + \int_0^{t \wedge \tau_m} \alpha_s^{nij}\, ds + \sum_k \int_0^{t \wedge \tau_m} \beta_{ks}^{nij}\, dY_s^k. \tag{2.12}$$

We shall first pass to the limit in (2.12) as $n \to \infty$. Note that by (2.11) and Hölder's inequality

$$|M(s,Y)P(\hat{X}_s)\hat{\psi}_s^n 1_{\{s\leqq\tau_m\}}|\leqq|M(s,Y)|(1+\sqrt{m})c(1+m) \tag{2.13}$$

since $|P(x)|\leqq 1+|x|^2$, and

$$\hat{\psi}_s^n 1_{\{s\leqq\tau_m\}}=E\{\psi^n(X_s)1_{\{s\leqq\tau_m\}}|\mathscr{Y}_s\},$$

$$|\psi^n(X_s)1_{\{s\leqq\tau_m\}}|\leqq c(1+|X_s|^4)1_{\{s\leqq\tau_m\}}, \tag{2.14}$$

$$E\{|X_s|^4 1_{\{s\leqq\tau_m\}}|\mathscr{Y}_s\}\leqq m.$$

Now (2.11) and (2.14) allow us to apply the dominated convergence theorem for conditional expectations to conclude that as $n\to\infty$

$$\hat{\psi}_s^n 1_{\{s\leqq\tau_m\}}\to\hat{\psi}_s 1_{\{s\leqq\tau_m\}} \quad \text{a.s., a.e.}$$

But this fact and (2.13) now imply similarly that

$$M(s,Y)P(\hat{X}_s)\hat{\psi}_s^n 1_{\{s\leqq\tau_m\}}\to M(s,Y)P(\hat{X}_s)\hat{\psi}_s 1_{\{s\leqq\tau_m\}}$$

in $L^1((0,t)\times\Omega,\ dt\times dP)$ if the term occurs in an $\alpha^n$, and in $L^2((0,t)\times\Omega,\ dt\times dP)$ if it occurs in a $\beta^n$. Hence we can pass to the limit in (2.12).

We can now pass to the limit as $m\to\infty$ to obtain

$$E\{X_t^i|\mathscr{Y}_t\}=E\{X_0^i\}+\int_0^t\widehat{(Lx_i)}_s\,ds+\sum_k\int_0^t[\widehat{(K_kx_i)}_s+\hat{\rho}_s^k(\hat{x}_i)_s][dY_s^k-\hat{\rho}_s^k\,ds]$$

$$E\{X_t^iX_t^j|\mathscr{Y}_t\}=E\{X_0^iX_0^j\}+\int_0^t\widehat{(Lx_ix_j)}_s\,ds \tag{2.15}$$

$$+\sum_k\int_0^t[\widehat{(K_kx_ix_j)}_s-\hat{\rho}_s^k\widehat{(x_ix_j)}_s][dY_s^k-\hat{\rho}_s^k\,ds].$$

Note that again $\widehat{(x_i)}_s$, $\widehat{(x_ix_j)}_s$, etc. are all functions of $\alpha(t)$, $\beta(t)$ and hence are locally bounded in $t$ a.s. by Corollary 2.1, hence the passage to the limit as $m\to\infty$ is justified by the fact that a.s. $M(\cdot,Y)$ is in $L^1_{\text{loc}}(\mathbb{R}^+)$ or in $L^2_{\text{loc}}(\mathbb{R}^+)$ as the case may be.

Some tedious but elementary algebra allows us to deduce from (2.15) that $(m_t, R_t)$ satisfy (2.7), (2.8). Note that since the conditional law of $X_t$ given $\mathscr{Y}_t$ is Gaussian then $(\widehat{x_i x_j x_k})_t = R_{ij}(\hat{x}_k)_t + R_{jk}(\hat{x}_i)_t + R_{ki}(\hat{x}_j)_t + (\hat{x}_i)_t(\hat{x}_j)_t(\hat{x}_k)_t$. □

## 3. SYSTEMS WITH "BENEŠ" TYPE NON-LINEARITIES

In the preceding section we saw that the conditional distribution of $X_t$ given $\mathscr{Y}_t$ is Gaussian, and we computed the two sufficient statistics $m_t, R_t$. This was done under the assumption that $(X_t, Y_t)$ satisfied (2.1), (2.2), specifically that, given $Y$, the drift of $X_t$ is linear in $X_t$. We shall now relax this linearity hypothesis to a certain extent, following the ideas of Beneš [1] and Shukhman [14]: in essence we shall transform the non-linear problem into one to which Theorems 2.1 and 2.2 can be applied. Since we cannot allow correlation between the observation noise and the noise in that part of the signal which involves the non-linear dynamics, we split the signal $X_t$ into two parts, $X_t^1$ and $X_t^2$. We assume

($B_1$) On some underlying filtered space $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}, P)$, $\{X_t : t \geqq 0\}$ is an adapted, a.s. continuous $\mathbb{R}^N$-valued process, $\{Y_t : t \geqq 0\}$ is an adapted, a.s. continuous $\mathbb{R}^d$-valued process, $\{W_t : t \geqq 0\}$, $\{V_t : t \geqq 0\}$, $\{U_t : t \geqq 0\}$ are independent $N_1$, $M$ and $d$ dimensional (respectively) standard Wiener processes such that $X_t^* = (X_t^{1*}, X_t^{2*})$, $X_t^1 \in \mathbb{R}^{N_1}$, $X_t^2 \in \mathbb{R}^{N_2}$, and

$$dX_t^1 = [A(t, Y)X_t^1 + a(t, Y) + f(t, Y, X_t^1)]\, dt + B(t, Y)\, dW_t \tag{3.1}$$

$$dX_t^2 = [C(t, Y)X_t + c(t, Y)]\, dt + D(t, Y)\, dV_t + \sum_i [G^i(t, Y)X_t + g^i(t, Y)]\, dY_t^i \tag{3.2}$$

$$dY_t = [H(t, Y)X_t + h(t, Y)]\, dt + dU_t,\ Y_0 = 0. \tag{3.3}$$

($B_2$) $A$, $a$, $B$, $C$, $c$, $D$, $G^i$, $g^i$, $H$, $h$ are all defined on $\mathbb{R}^+ \times C(\mathbb{R}^+; \mathbb{R}^d)$ and are progressively measurable and bounded, and $f$ is defined on $\mathbb{R}^+ \times C(\mathbb{R}^+; \mathbb{R}^d) \times \mathbb{R}^{N_1}$ and is $\mathscr{P} \otimes \mathscr{B}^{N_1}$ measurable ($\mathscr{P}$ is the $\sigma$-algebra of progressively measurable subsets of $\mathbb{R}^+ \times C(\mathbb{R}^+; \mathbb{R}^d)$ and $\mathscr{B}^{N_1}$ is the Borel $\sigma$-algebra on $\mathbb{R}^{N_1}$). Moreover $f(\cdot, y, x)$ is in $L^1_{\text{loc}}(\mathbb{R}^+)$ for each $(y, x)$.

($B_3$) $H(t,y)=(H_1(t,y)\ H_2(t,y))$ and $H_1^*H_2\equiv 0$.

($B_4$) There exists a $\mathscr{P}\otimes\mathscr{B}^{N_1}$ measurable function $\Phi$: $\mathbb{R}^+\times C(\mathbb{R}^+;\mathbb{R}^d)\times\mathbb{R}^{N_1}\to\mathbb{R}$ which is $C^{1,2}$ in $(t,x)$ for each $y$, such that for each $y,x$, $\Phi(t,y,x)=\Phi(0,y,x)+\int_0^t\dot{\Phi}(s,y,x)\,ds$, and that $(1+|x|)^{-1}\Phi_x(t,y,x)$ is bounded, and there exist progressively measurable, bounded functions $\Sigma$, $\Gamma$, $\delta$ defined on $\mathbb{R}^+\times C(\mathbb{R}^+;\mathbb{R}^d)$ with values in $\mathbb{R}^{N_1}\otimes\mathbb{R}^{N_1}$, $\mathbb{R}^{N_1}$, $\mathbb{R}$ with $\Sigma$ symmetric such that $(1+|x|)^{-1}\Phi_x(t,y,x)$ is bounded and such that for all $(t,y,x)$

$$BB^*\Phi_x=f \tag{3.4}$$

$$2\dot{\Phi}+\operatorname{tr}(BB^*\Phi_{xx})+2\Phi_x^*(Ax+a)+|B^*\Phi_x|^2=x^*\Sigma x+\Gamma^*x+\delta, \tag{3.5}$$

$$\Sigma+H_1^*H_1\geqq 0, \tag{3.6}$$

where $(\Phi_x)_i=\partial\Phi/\partial x_i$, $(\Phi_{xx})_{ij}=\partial^2\Phi/\partial x_i\,\partial x_j$ and $\dot{\Phi}=\partial\Phi/\partial t$. Note that since $\Phi$ is adapted then $\Phi(0,y,x)$ is a function of $x$ only which we indicate by $\Phi_0(x)$.

The main idea is to perform a Girsanov transformation which changes $f$ into $BB^*JX^1$ where $J$ is a matrix valued process to be defined below. Unfortunately in doing so we introduce a Radon–Nikodym density into the conditional expectation, but the hypothesis ($B_4$) allows us to express this density as a function of a new variable $X^3$ such that $(X^1,X^2,X^3)$ satisfy a conditionally linear system to which the results of Section 2 can be applied.

In the following examples ($B_4$) is satisfied.

*Example 3.1* Let $B=I$ and $\alpha_1(t,y)$, $\alpha_2(t,y)$ be progressively measurable functions mapping $\mathbb{R}^+\times C(\mathbb{R}^+;\mathbb{R}^d)$ into $\mathbb{R}$, which are continuously differentiable in $t$ in the sense that $\alpha_i(t,y)=\alpha_i(0,y)+\int_0^t\dot{\alpha}_i(s,y)\,ds$, $\dot{\alpha}_i(\cdot,y)$ continuous, and are bounded on $[0,T]\times C(\mathbb{R}^+;\mathbb{R}^d)$ for any $T<\infty$, such that $\alpha_2(t,y)\alpha_1(t,y)>0$ for all $(t,y)$. Let $p_0$, $p_1$, $p_2$ be similarly measurable, differentiable and bounded with $p_0$ assuming values in $\mathbb{R}$, $p_1$ in $\mathbb{R}^{N_1}$ and $p_2$ in $\mathbb{R}^{N_1}\otimes\mathbb{R}^{N_1}$. We suppose that they satisfy

$$\dot{p}_2(t,y)+p_2(t,y)A(t,y)+A(t,y)^*p_2(t,y)=0,$$

$$\dot{p}_1(t,y)+A(t,y)^*p_1(t,y)+[p_2(t,y)+p_2(ty)^*]a(t,y)=0,$$

$$\dot{p}_0(t,y)+\mathrm{tr}[p_2(t,y)+p_1(t,y)^*a(t,y)=\frac{1}{2}\frac{d}{dt}\ln\left(\frac{\alpha_2(t,y)}{\alpha_1(t,y)}\right).$$

Now suppose

$$f(t,y,x)=\left[\frac{\alpha_1(t,y)\,e^{P(x)}-\alpha_2(t,y)\,e^{-P(x)}}{\alpha_1(t,y)\,e^{P(x)}+\alpha_2(t,y)\,e^{-P(x)}}\right]\nabla P(x),$$

where

$$P(x)=x^*p_2(t,y)x+p_1(t,y)^*x+p_0(t,y),$$

$$\nabla P(x)=[p_2(t,y)+p_2(t,y)^*]x+p_1(t,y).$$

Then ($\mathbf{B}_4$) holds with

$$\Phi(t,y,x)=\ln[\alpha_1(t,y)\,e^{P(x)}+\alpha_2(t,y)\,e^{-P(x)}],$$

$$\Sigma(t,y)=[p_2(t,y)+p_2(t,y)^*]^2,$$

$$\Gamma(t,y)=2[p_2(t,y)+p_2(t,y)^*]p_1(t,y),$$

$$\delta(t,y)=|p_1(t,y)|^2+\frac{d}{dt}\ln[\alpha_1(t,y)\alpha_2(t,y)].$$

The above example is a generalization of those found in [1, 2] where $f(x)=\tanh x$, $A=0$, $a=0$ so that $\alpha_1=\alpha_2=1$, $p_2=p_0=0$, $p_1=1$.

*Example 3.2* Suppose $B=I$ and

$$\Phi(t,y,x)=\ln[M(t,y,x)+k(t,y)]$$

where $(1+|x|)^{-1}|M_x||M+k|^{-1}$ is bounded and

$$2\dot{M}+2\dot{k}+\nabla^2 M+2M_x^*(Ax+a)=0,$$

(for example if $M$ is a suitable quadratic in $x$), then ($\mathrm{B}_4$) holds with

$$f(t, y, x) = M_x(t, y, x)[M(t, y, x) + k(t, y)]^{-1}$$

and $\Sigma = 0$, $\Gamma = 0$, $\delta = 0$. □

We return to the general case. For $y$ in $C(\mathbb{R}^+; \mathbb{R}^d)$ let $J(t, Y)$ denote the unique symmetric, positive semi-definite solution of

$$\dot{J} + JA + A^*J + J(BB^*)J - (\Sigma + H_1^*H_1) = 0, \qquad J(0) = 0. \tag{3.7}$$

Let us partition $C = (C_1 \ C_2)$ corresponding to $X^1$, $X^2$, and similarly $G^j = (G_1^j \ G_2^j)$. Also we write $H_1^i$ for the $i$th row of $H_1$. Now let

$$X_t^3 = -\Phi_0(X_0^1) + \int_0^t [J(s, Y)a(s, Y) - H_1(s, Y)^*h(s, Y) - \tfrac{1}{2}\Gamma(s, Y)]^* X_s^1 \, ds$$

$$+ \int_0^t [H_1(s, Y)X_s^1]^* \, dY_s$$

and write

$$\tilde{X}_t = \begin{bmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \end{bmatrix}, \quad \tilde{A} = \begin{bmatrix} A + BB^*J & 0 & 0 \\ C_1 & C_2 & 0 \\ a^*J - h^*H_1 - \tfrac{1}{2}\Gamma^* & 0 & 0 \end{bmatrix}, \quad \tilde{a} = \begin{bmatrix} a \\ c \\ 0 \end{bmatrix}$$

$$\tilde{G}^j = \begin{bmatrix} 0 & 0 & 0 \\ G_1^j & G_2^j & 0 \\ H_1^j & 0 & 0 \end{bmatrix}, \quad \tilde{g}^j = \begin{bmatrix} 0 \\ g^j \\ 0 \end{bmatrix}, \quad \tilde{B} = \begin{bmatrix} B & 0 \\ 0 & D \\ 0 & 0 \end{bmatrix}$$

$\tilde{H} = (0 \ H_2 \ 0)$. We write $\mu(\cdot; m, R)$ for the distribution of a Gaussian random $N+1$ dimensional vector, mean $m$ and covariance $R$. Now we have

THEOREM 3.1 *Assume* $(B_1)$–$(B_4)$ *and one of*

i) $H_2 = 0$,

ii) $G^j(t, y) = 0, \forall t, y, j = 1, \ldots, d$,

iii) $a$, $A$, $B$, $f$, $J$ *are constant in the second variable* $y$, *and* $G_2^j = 0$.

*Then an unnormalized conditional distribution of $X_T$ given $\mathscr{Y}_T$ and $X_0 = x_0$ is given by*

$$\rho(X_T \in S | \mathscr{Y}_T, X_0 = x_0)$$

$$= \iiint_{S \times \mathbb{R}} \exp\{\Phi(T, Y, x^1) - \tfrac{1}{2} x^{1*} J(T, Y) x^1 + x^3\} \mu(dx^1 dx^2 dx^3; m_T, R_T)$$

*where $S$ is a Borel set in $\mathbb{R}^N$, $x^* = (x^{1*}, x^{2*}, x^3)$ and $m, R$ satisfy*

$$dm_t = \Big\{ \tilde{A}(t, Y) m_t + \tilde{a}(t, Y) - R_t \tilde{H}(t, Y)^* [\tilde{H}(t, Y) m_t + h(t, Y)]$$

$$+ \sum_j \tilde{G}^j(t, Y) R_t \tilde{H}^j(t, Y)^* \Big\} dt$$

$$+ \sum_j \{\tilde{G}^j(t, Y) m_t + \tilde{g}^j(t, Y) + R_t \tilde{H}^j(t, Y)^*\} \, dY_t^j,$$

$$m_0 = \begin{bmatrix} x_0 \\ -\Phi_0(x_0) \end{bmatrix},$$

$$dR_t = \Big\{ \tilde{B}(t, Y) \tilde{B}(t, Y)^* + \tilde{A}(t, Y) R_t + R_t \tilde{A}(t, Y)^*$$

$$+ \sum_j \tilde{G}^j(t, Y) R_t \tilde{G}^j(t, Y)^* - R_t \tilde{H}(t, Y)^* \tilde{H}(t, Y) R_t \Big\} dt$$

$$+ \sum_j \{\tilde{G}^j(t, Y) R_t + R_t \tilde{G}^j(t, Y)^*\} \, dY_t^j,$$

$$R_0 = 0.$$

Observe that the above equations for $m$, $R$ are just (2.7), (2.8) for the process $\tilde{X}$.

*Proof* The first step is to change the non-linearity $f$ into a linear term using a Girsanov transformation. We let

$$\tilde{\Lambda}_t = \exp\Big\{ \int_0^t [\Phi_x(s, Y, X_s^1) - J(s, Y) X_s^1]^* \, dX_s^1 - \tfrac{1}{2} \int_0^t [\Phi_x(s, Y, X_s^1)$$

$$-J(s, Y)X_s^1]^*[B(s, Y)B(s, Y)^*(\Phi_x(s, Y, X_s^1)+J(s, Y)X_s^1)$$

$$+2(A(s, Y)X_s^1+a(s, Y))]\,ds$$

$$+\int_0^t [H_1(x, Y)X_s^1]^*\,dY_s-\tfrac{1}{2}\int_0^t |H_1(s, Y)X_s^1|^2\,ds$$

$$-\int_0^t h(s, Y)^*H_1(s, Y)X_s^1\,ds\bigg\}.$$

Under our hypotheses we may define $\tilde{P}_T$ by $d\tilde{P}_T=\tilde{\Lambda}_T^{-1}\,dP$ as a probability measure, and for some independent standard Wiener processes $\bar{W}$, $V$, $\tilde{U}$ on $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\leq T}, \tilde{P}_T)$ $X^1$ and $Y$ satisfy

$$dX_t^1=\{[A(t, Y)+B(t, Y)B(t, Y)^*J(t, Y)]X_t^1+a(t, Y)\}\,dt+B(t, Y)\,d\bar{W}_t, \tag{3.8}$$

$$dY_t=\{H_2(t, Y)X_t^2+h(t, Y)]\,dt+d\tilde{U}_t, \qquad t\leqq T. \tag{3.9}$$

Note that

$$d\bar{W}_t=dW_t+B^*(\Phi_x-JX_t^1)\,dt,$$

$$d\tilde{U}_t=dU_t+H_1X_t^1\,dt.$$

The next step is to manipulate $\tilde{\Lambda}_T$ so that it can be written as a functional of the process $\tilde{X}$. From ($B_4$) and Itô's formula (for each fixed sample path of $Y$) it follows that

$$d[\Phi(t, Y, X_t^1)-\tfrac{1}{2}X_t^{1*}J(t, Y)X_t^1]$$

$$=(\Phi_x-JX_t^1)^*\,dX_t^1+\{(\dot{\Phi}-\tfrac{1}{2}X_t^{1*}\dot{J}X_t^1)+\tfrac{1}{2}\operatorname{tr}[(\Phi_{xx}-J)BB^*]\}\,dt$$

$$=(\Phi_x-JX_t^1)^*\,dX_t^1+\tfrac{1}{2}\{-2\Phi_x^*(AX_t^1+a)-|B^*\Phi_x|^2$$

$$+X_t^{1*}\Sigma X_t^1+\Gamma^*X_t^1+\delta-X_t^{1*}\dot{J}X_t^1-\operatorname{tr}(JBB^*)\}\,dt.$$

Using this result to substitute for $\int(\Phi_x-JX^1)^*\,dX^1$ in $\tilde{\Lambda}_T$ we have

$$\log\tilde{\Lambda}_T=\Phi(T, Y, X_T^1)-\Phi_0(X_0^1)-\tfrac{1}{2}X_T^{1*}J(T, Y)X_T^1$$

$$-\tfrac{1}{2}\int_0^T [-2\Phi_x^*(AX_t^1+a)-|B^*\Phi_x|^2+X_t^{1*}\Sigma X_t^1$$

$$+\Gamma^* X_t^1+\delta-X_t^{1*}\dot{J}X_t^1-\operatorname{tr}(JBB^*)]\,dt$$

$$-\tfrac{1}{2}\int_0^T [|B^*\Phi_x|^2-|B^*JX_t^1|^2+2(\Phi_x-JX_t^1)^*(AX_t^1+a)]\,dt$$

$$+\int_0^T (H_1X_t^1)^*dY_t-\int_0^T [\tfrac{1}{2}X_t^{1*}H_1^*H_1X_t^1+h^*H_1X_t^1]\,dt$$

$$=\Phi(T,Y,X_T^1)-\tfrac{1}{2}X_T^{1*}J(T,Y)X_T^1-\Phi_0(X_0^1)$$

$$-\tfrac{1}{2}\int_0^T X_t^{1*}[\Sigma-\dot{J}-JBB^*J-JA-A^*J+H_1^*H_1]X_t^1\,dt$$

$$-\tfrac{1}{2}\int_0^T [(\Gamma-2Ja+2H_1^*h)^*X_t^1+(\delta-\operatorname{tr}(JBB^*))]\,dt$$

$$+\int_0^T (H_1X_t^1)^*dY_t$$

$$=\Phi(T,Y,X_T^1)-\tfrac{1}{2}X_T^{1*}J(T,Y)X_T^1-\Phi_0(X_0^1)+\int_0^T (H_1X_t^1)^*\,dY_t$$

$$-\tfrac{1}{2}\int_0^T [(\Gamma-2Ja+2H_1^*h)^*X_t^1+(\delta-\operatorname{tr}(JBB^*))]\,dt,$$

where we have used (3.7). Now it follows from the definition of $X^3$ that

$$\tilde{\Lambda}_T=\exp\int_0^T \tfrac{1}{2}\{\operatorname{tr}[J(t,Y)B(t,Y)B(t,Y)^*]-\delta(t,Y)\}\,dt\exp\{\Phi(T,Y,X_T^1)$$
$$-\tfrac{1}{2}X_T^{1*}J(T,Y)X_T^1+X_T^3\}. \tag{3.10}$$

The final step is to express an unnormalized conditional distribution of $X_T$ under $P$ as a conditional distribution of $\tilde{X}_T$ under the measure $\tilde{P}_T$, since under $\tilde{P}_T$, $\tilde{X}_.$ is the solution of a conditionally linear system, cf. (3.13), (3.14) below, so the results of Section 2 can be applied.

If $S$ is any Borel set in $\mathbb{R}^N$ we need to compute

$$E\{1_S(X_T)|\mathscr{Y}_T, X_0=x_0\}$$
$$=\tilde{E}_T\{1_S(X_T)\Lambda_T|\mathscr{Y}_T, X_0=x_0\}/\tilde{E}_T\{\Lambda_T|\mathscr{Y}_T, X_0=x_0\} \tag{3.11}$$

or indeed, we need to compute an unnormalized conditional distribution—for example the numerator in the above expression, or again using (3.10), simply

$$\tilde{E}_T\{1_S(X_T)\exp[\psi(T, Y, X_T^1)-\tfrac{1}{2}X_T^{1*}J(T, Y)X_T^1+X_T^3]|\mathscr{Y}_T, X_0=x_0\}. \tag{3.12}$$

But if we set $\tilde{W}=\begin{pmatrix}\bar{W}\\ V\end{pmatrix}$, then

$$d\tilde{X}_t=[\tilde{A}(t, Y)\tilde{X}_t+\tilde{a}(t, Y)]\,dt+\tilde{B}(t, Y)\,d\tilde{W}$$
$$+\sum_i \tilde{G}^i(t, Y)\tilde{X}_t+\tilde{g}^i(t, Y)]\,dY_t^i \tag{3.13}$$

$$dY_t=[\tilde{H}(t, Y)\tilde{X}_t+h(t, Y)]\,dt+d\tilde{U} \tag{3.14}$$

with $\tilde{X}_0=\begin{pmatrix}x_0\\ -\Phi_0(x_0)\end{pmatrix}$. The system (3.13), (3.14) has the form (2.1), (2.2). Moreover $(A_1)$ is satisfied with $m_0=\tilde{X}_0$, $R_0=0$. Our hypothesis $(B_2)$ implies $(A_2)$, $(A_3)$, $(A_5)$. In addition any of the conditions (i), (ii) or (iii) imply that (3.13), (3.14) is a special case of either Example 2.1 or Example 2.2, hence $(A_4)$ and $(A_6)$ also hold. The result now follows from Theorems 2.1 and 2.2. $\square$

*Remark 3.1* $(B_3)$ is satisfied if $H$ has the structure $H=(H_1\;\;H_2)$ with

$$H_1=\begin{pmatrix}H_{11}\\ 0\end{pmatrix},\quad H_2=\begin{pmatrix}0\\ H_{22}\end{pmatrix}.$$

On the other hand if in $(B_4)$ we have $\Sigma \geqq 0$ (which is stronger than (3.6)) then we may omit $(B_3)$. Indeed in (3.7) and in the definition of $\tilde{A}$, $\tilde{G}^j$, $X_t^3$ and $\tilde{\Lambda}$ we replace $H_1$ by 0, and we replace $\tilde{H}$ by $(H_1 \; H_2 \; 0)$.

*Remark 3.2* In connection with (iii) observe that in Examples 3.1 and 3.2, $J$ is constant in $Y$ if $a$, $A$, $B$, $H_1$ are (if $\Sigma \geqq 0$ then only $a$, $A$, $B$ need be constant). Moreover $G_2^j = 0$ simply states that when we transform (3.2) into an Itô equation by substituting for $dY$ from (3.3), the the equation is linear in $X^2$ (but not necessarily in the pair $(X_t^1, X_t^2)$, i.e. in $X_t$).

On the other hand the condition (ii), i.e. $G^j = 0$, implies that the Itô equation for $X_t^2$ is linear in $X_t$. A special case of this situation is treated in [14]. As for (i), in this case $(X_t^1, Y_t)$ constitutes a signal-observation pair with no input from $X_t^2$.

Note also that (i), (ii) or (iii) could be replaced by:

(iv) $\tilde{P}_T(\Omega) = 1$ and the system (3.13), (3.14) satisfies $(A_4)$ and $(A_6)$.

*Remark 3.3* Wong [15] considers a system of the form

$$\begin{aligned} dx_1 &= \bar{f}(x_1, x_2)\, dt + dw_t \\ dx_2 &= g(x_2)\, dt \\ dy &= x_1\, dt + dv_t. \end{aligned} \tag{3.15}$$

If we set $X_t^2 = 0$, $X_t^1 = (x_1(t), x_2(t))$, we obtain the form (3.1)–(3.3) but clearly (3.4) must fail. However, $\bar{f}$ satisfies further hypotheses. In fact in [15] examples of "class $A$" can be put into the form (3.1)–(3.3) with $Y_t = y(t)$, $X_t^2 = 0$, $X_t^{1*} = (\xi_1(t), \xi_2^*(t))$ where $\xi_1(t) = x_1(t)$ and $\xi_2(t)$ is a $2m$ dimensional vector with components

$$\cos \int_0^{x_2(t)} [\lambda_i / g(r)]\, dr, \quad \sin \int_0^{x_2(t)} [\lambda_i / g(r)]\, dr, \qquad i = 1, \ldots, m, \tag{3.16}$$

and the $\lambda_i$ are given. Moreover $f = 0$, $a^* = (\alpha, 0^*)$, $B^* = (1, 0^*)$ and

$$A = \begin{bmatrix} 0 & B^* \\ 0 & A_{22} \end{bmatrix}$$

for a given scalar $\alpha$ and $2m$ dimensional vector $\beta$. Clearly this case is covered by Theorem 3.1.

Examples of the "class $B$" of [15] can be put into the form (3.1)–(3.3) with $Y_t = y(t)$, $X_t^2 = 0$, $X_t^{1*} = \xi_1(t)$, $\xi_2^*(t), \xi_3(t)$ where $\xi_1$ and $\xi_2$ are as above and

$$\xi_3(t) = \int_0^{x_2(t)} \frac{2\beta^*\bar{\xi}_2(r) + 1}{g(r)}\,dr.$$

Here $\bar{\xi}_2(r)$ has components defined by (3.16) with upper limit of integration being $r$ rather than $x_2(t)$. Moreover $a^* = (0, 0^*, 1/2)$, $B^* = (1, 0^*)$,

$$A = \begin{bmatrix} 0 & \beta^* & 0 \\ 0 & A_{22} & 0 \\ 0 & \beta^* & 0 \end{bmatrix}, \qquad f = \begin{bmatrix} \eta(\xi_1 - \xi_3) \\ 0 \\ 0 \end{bmatrix}$$

where $\eta(x) = c\,e^x/(1 + c\,e^x)$, $c > 0$. If we set $\Lambda(x) = \int \eta(x)\,dx$ then $\Phi(t, y, (\xi_1, \xi_2, \xi_3)) = \Lambda(\xi_1 - \xi_3)$ satisfies ($B_4$) since

$$\eta_x(\xi_1 - \xi_3) - \eta(\xi_1 - \xi_3) + |\eta(\xi_1 - \xi_3)|^2 = 0,$$

i.e. we can take $\Sigma = 0$, $\Gamma = 0$, $\delta = 0$. Hence the result is again covered by Theorem 3.1, which thus includes all the results of [15].

## 4. NON-GAUSSIAN INITIAL CONDITION

So far we have only obtained the conditional distribution of $X_T$ given $\mathscr{Y}_T$ and $X_0$. If $(X_0, -\Phi_0(X_0^1))$ is Gaussian, e.g. $X_0$ is Gaussian and $\Phi_0$ is an affine function, then the proof of Theorem 3.1 can be used to give a formula for the conditional distribution of $X_T$ given $\mathscr{Y}_T$ only. In general if $P_0$ is the distribution of $X_0$, one can proceed as in Theorem 3.1 up to (3.11):

$$E\{1_S(X_T)|\mathcal{Y}_T\}=\tilde{E}_T\{1_S(X_T)\Lambda_T|\mathcal{Y}_T\}/\tilde{E}_T\{\Lambda_T|\mathcal{Y}_T\}$$

$$=\frac{\int_{\mathbb{R}^N}\tilde{E}_T\{1_S(X_T)\Lambda_T|\mathcal{Y}_T,X_0=x\}P_0(dx|\mathcal{Y}_T)}{\int_{\mathbb{R}^N}\tilde{E}_T(\Lambda_T|\mathcal{Y}_T,x_0=x)P_0(dx|\mathcal{Y}_T)}$$

but now we do not know $P_0(\cdot|\mathcal{Y}_T)$, the conditional distribution of $X_0$ given $\mathcal{Y}_T$, and we do not wish to find it as the solution of an interpolation problem since this would not give a recursive scheme. Instead we apply the method of Makowski [11], that is we split off the initial condition and find a new measure relative to which $X_0$ and $\mathcal{Y}_T$ are independent so that $P_0(\cdot|\mathcal{Y}_T)=P_0(\cdot)$ (at least under the new measure). Moreover the corresponding Radon–Nikodym density will be a function of the variables $X^1\dots X^5$ which satisfy a conditionally linear system so that Theorem 2.2 can be applied.

The model is again (3.1)–(3.3).

THEOREM 4.1 *Assume* $(B_1)$–$(B_4)$ *and one of*

*i)* $H_2=0$,

*ii)* $G^j(t,y)=0\ \forall\, t,y, j=1,\dots,d$,

*iii)* *$f$, $a$, $A$, $B$, $J$ are constant in the second variable $y$ and* $G^j_2=0$.

*Let $P_0$ be the distribution of $X_0$. Then an unnormalized conditional distribution of $X$ given $\mathcal{Y}_T$ is given by*

$$\begin{aligned}\rho(X_T\in S|\mathcal{Y}_T)=&\int_{\mathbb{R}^N}\exp\{-\Phi_0(x^1)-\tfrac{1}{2}x^*N(T,Y)x\}\\&\times\int_{\mathbb{R}^{N_1}\times\mathbb{R}^{N_2}\times\mathbb{R}\times\mathbb{R}^{N_1}\times\mathbb{R}^{N_2}}1_S\left[\begin{pmatrix}z^1\\z^2\end{pmatrix}+\Psi(T,Y)x\right]\\&\times\exp\{\Phi(T,Y,z^1+\Psi_{11}(T,Y)x^1)\\&+\tfrac{1}{2}(z^1+\Psi_{11}(T,Y)x^1)^*J(T,Y)(z^1+\Psi_{11}(T,Y)x^1)\\&+z^3+x^{1*}z^4+x^{2*}z^5\}\mu(dz;m_T,R_T)P_0(dx),\end{aligned}$$

*where $S$ is any Borel set in* $\mathbb{R}^N$, $x^*=(x^{1*},x^{2*})$, $z^*=(z^{1*},z^{2*},z^{3*},z^{4*},z^{5*})$,

$$N(T, Y) = \int_0^T [\Psi_{21}(t, Y) \ \Psi_{22}(t, Y)]^* H_2(t, Y)^* H_2(t, Y)$$

$$\times [\Psi_{21}(t, Y) \ \Psi_{22}(t, Y)] \, dt,$$

*and*

$$\Psi(T, Y) = \begin{bmatrix} \Psi_{11}(T, Y) & \Psi_{12}(T, Y) \\ \Psi_{21}(T, Y) & \Psi_{22}(T, Y) \end{bmatrix}$$

*in* $\mathbb{R}^{N_1+N_2} \otimes \mathbb{R}^{N_1+N_2}$ *satisfies*

$$d\Psi(t, Y) = \begin{bmatrix} A(t, Y) + B(t, Y)B(t, Y)^* J(t, Y) & 0 \\ C_1(t, Y) & C_2(t, Y) \end{bmatrix} \Psi(t, Y) \, dt$$

$$+ \sum_j \begin{bmatrix} 0 \\ G^j(t, Y) \end{bmatrix} \Psi(t, Y) \, dY_t^j, \qquad \Psi(0, Y) = I,$$

*and* $m$, $R$ *satisfy*

$$dm_t = \Big\{ \hat{A}(t, Y)m_t + \hat{a}(t, Y) - R_t \hat{H}(t, Y)^* [\hat{H}(t, Y)m_t + h(t, Y)]$$

$$+ \sum_j \hat{G}^j(t, Y) R_t \hat{H}^j(t, Y)^* \Big\} dt$$

$$+ \sum_j \{ \hat{G}^j(t, Y)m_t + \hat{g}^j(t, Y) + R_t \hat{H}^j(t, Y)^* \} \, dY_t^j, \tag{4.1}$$

$$m_0 = 0,$$

$$dR_t = \Big\{ \hat{B}(t, Y)\hat{B}(t, Y)^* + \hat{A}(t, Y)R_t + R_t \hat{A}(t, Y)^*$$

$$+ \sum_j \hat{G}^j(t, Y) R_t \hat{G}^j(t, Y) - R_t \hat{H}(t, Y)^* \hat{H}(t, Y) R_t \Big\} dt \tag{4.2}$$

$$+ \sum_j \{ \hat{G}^j(t, Y) R_t + R_t \hat{G}^j(t, Y)^* \} \, dY_t^j,$$

$$R_0 = 0,$$

$$\hat{A}=\begin{bmatrix} A+BB^*J & 0 & 0 & 0 & 0 \\ C_1 & C_2 & 0 & 0 & 0 \\ (Ja-H_1^*h-\frac{1}{2}\Gamma)^* & 0 & 0 & 0 & 0 \\ 0 & -\Psi_{21}^*H_2^*H_2 & 0 & 0 & 0 \\ 0 & -\Psi_{22}^*H_2^*H_2 & 0 & 0 & 0 \end{bmatrix}$$

$$\hat{a}=\begin{bmatrix} a \\ c \\ 0 \\ \Psi_{11}^*(Ja-H_1^*h-\frac{1}{2}\Gamma)-\Psi_{21}^*H_2^*h \\ -\Psi_{22}^*H_2^*h \end{bmatrix}, \quad \hat{B}=\begin{bmatrix} B & 0 \\ 0 & D \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\hat{G}^j=\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ G_1^j & G_2^j & 0 & 0 & 0 \\ H_1^j & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \hat{g}^j=\begin{bmatrix} 0 \\ g^j \\ 0 \\ (H_1^j\Psi_{11}+H_2^j\Psi_{21})^* \\ (H_2^j\Psi_{22})^* \end{bmatrix}$$

$$\hat{H}=(0, H_2, 0, 0, 0,).$$

*Proof* We define $\xi_t \in \mathbb{R}^{N_1+N_2+1}$, $\bar{X}_t^4 \in \mathbb{R}^{N_1}$, $\bar{X}_t^5 \in \mathbb{R}^{N_2}$ by

$$\xi_t=\begin{bmatrix} \Psi_{11}(t, Y) & \Psi_{12}(t, Y) & 0 \\ \Psi_{21}(t, Y) & \Psi_{22}(t, Y) & 0 \\ \zeta_t^* & 0 & 1 \end{bmatrix} \tilde{X}_0 \tag{4.3}$$

$$\zeta_t=\int_0^t \Psi_{11}(s, Y)^*[J(s, Y)a(s, Y)-H_1(s, Y)^*h(s, Y)-\tfrac{1}{2}\Gamma(s, Y)]\, ds$$

$$+\sum_j \int_0^t \Psi_{11}(s, Y)^*H_1^j(s, Y)^*\, dY_s^j,$$

$$\bar{X}_t^4=\zeta_t-\int_0^t [H_2(s, Y)\Psi_{21}(s, Y)]^*[H_2(s, Y)(X_s^2-\xi_s^2)+h(s, Y)]\, ds$$

$$+\sum_j \int_0^t [H_2^j(s, Y)\Psi_{21}(s, Y)]^*\, dY_s^j$$

$$\bar{X}_t^5 = -\int_0^t [H_2(s, Y)\Psi_{22}(s, Y)]^*[H_2(s, Y)(X_s^2 - \xi_s^2) + h(s, Y)]\, ds$$

$$+\sum_j \int_0^t [H_2^j(s, Y)\Psi_{22}(s, Y)]^*\, dY_s^j.$$

It follows that

$$d\xi_t = \tilde{A}(t, Y)\xi_t\, dt + \sum_j \tilde{G}^j(t, Y)\xi_t\, dY_t^j, \qquad \xi_0 = \tilde{X}_0,$$

and if we define

$$\hat{\Lambda}_T^{-1} = \exp\left\{-\int_0^T [\tilde{H}(t, Y)\xi_t]^*\, d\tilde{U}_t - \tfrac{1}{2}\int_0^T |\tilde{H}(t, Y)\xi_t|^2\, dt\right\},$$

then

$$\hat{\Lambda}_T = \exp\{X_0^{1*}(\bar{X}_T^4 - \zeta_T) + X_0^{2*}\bar{X}_T^5 - \tfrac{1}{2}X_0^* N(T, Y)X_0\}.$$

If we define $\hat{P}_T$ by $d\hat{P}_T = \hat{\Lambda}_T^{-1}\, d\tilde{P}_T$, then under our hypotheses $\hat{P}_T$ is a probability measure under which $Y_t - \int_0^t [\tilde{H}(\tilde{X} - \xi) + h]\, ds$ is a standard Wiener process. Now if $\bar{X}_t^i = \tilde{X}_t^i - \xi_t^i$, $i = 1, 2, 3$, then

$$d\bar{X}_t = [\hat{A}(t, Y)\bar{X}_t + \hat{a}(t, Y)]\, dt + \hat{B}(t, Y)\, d\hat{W}$$

$$+\sum_j [\hat{G}^j(t, Y)\bar{X}_t + \hat{g}^j(t, Y)]\, dY_t^j, \tag{4.4}$$

$$dY_t = [\hat{H}(t, Y)\bar{X}_t + h(t, Y)]\, dt + d\hat{U}_t, \qquad 0 \leqq t \leqq T, \tag{4.5}$$

$$\bar{X} = 0, \quad Y_0 = 0,$$

where $\hat{W}$, $\hat{U}$ are independent standard $\{\mathscr{F}_t\}$ Wiener processes under $\hat{P}_T$. Let $Q_T$ be the probability measure obtained from $\hat{P}_T$ by a Girsanov transformation such that $(\hat{W}, Y)$ is a standard $\{\mathscr{F}_t\}$ Wiener process under $Q_T$. Then $\bar{X}$, the solution of (4.4), is adapted to $(\hat{W}, Y)$, and hence $X_0$ and $(\bar{X}, Y)$ are independent under $Q_T$. It follows from the Girsanov representation for $d\hat{P}_T/dQ_T$ that $X_0$ and $(Y, d\hat{P}_T/dQ_T)$ are independent under $Q_T$ and hence $X_0$ and $Y$ are independent under $\hat{P}_T$.

If $S$ is a Borel set then as in (3.11)

$$E\{1_S(X_T)|\mathscr{Y}_T\}=\tilde{E}_T\{1_S(X_T)\tilde{\Lambda}_T|\mathscr{Y}_T\}/\tilde{E}_T\{\tilde{\Lambda}_T|\mathscr{Y}_T\}$$

and it suffices to compute, cf. (3.12),

$$\tilde{E}_T\{1_S(X_T)\exp[\Phi(T,Y,X_T^1)-\tfrac{1}{2}X_T^{1*}J(T,Y)X_T^1+X_T^3]|\mathscr{Y}_T\}$$

$$=\hat{E}_T\{1_S(X_T)\exp[\Phi(T,Y,X_T^1)-\tfrac{1}{2}X_T^{1*}J(T,Y)X_T^1+X_T^3]\hat{\Lambda}_T|\mathscr{Y}_T\}$$

$$/\hat{E}_T\{\hat{\Lambda}_T|\mathscr{Y}_T\},$$

Let us compute the numerator in the last expression (the denominator is irrelevant). N.B. $\Psi_{12}\equiv 0$.

$$\hat{E}_T\{1_S(X_T)\exp[\Phi(T,Y,X_T^1)-\tfrac{1}{2}X_T^{1*}J(T,Y)X_T^1+X_T^3$$

$$+X_0^{1*}(\bar{X}_T^4-\zeta_T)+X_0^{2*}\bar{X}_T^5-\tfrac{1}{2}X_0^*N(T,Y)X_0]|\mathscr{Y}_T\}$$

$$=\hat{E}_T\left\{1_S\left[\begin{pmatrix}\bar{X}_T^1\\ \bar{X}_T^2\end{pmatrix}+\Psi(T,Y)X_0\right]\exp[\Phi(T,Y,\bar{X}_T^1+\Psi_{11}(T,Y)X_0^1)\right.$$

$$-\tfrac{1}{2}(\bar{X}_T^1+\Psi_{11}(T,Y)X_0^1)^*J(T,Y)(\bar{X}_T^1+\Psi_{11}(T,Y)X_0^1)$$

$$\left.+\bar{X}_T^3-\Phi_0(X_0^1)+X_0^{1*}\bar{X}_T^4+X_0^{2*}\bar{X}_T^5-\tfrac{1}{2}X_0^*N(T,Y)X_0]|\mathscr{Y}_T\right\} \tag{4.4}$$

since $X_T^i=\bar{X}_T^i+\xi_T^i$, $i=1,2,3$ and $\xi_T$ satisfies (4.1). Now (4.4) has the form

$$\hat{E}_T\{F(X_0,\bar{X}_T,Y)|\mathscr{Y}_T\}=\int_{\mathbb{R}^N}\hat{E}_T\{F(x,\bar{X}_T,Y)|\mathscr{Y}_T\}P_0(dx) \tag{4.5}$$

since the distribution of $X_0$ under $\hat{P}_T$ is still $P_0$ and since $X_0$ is independent of $Y$ under $\hat{P}_T$. But Theorems 2.1 and 2.2 applied to $\bar{X}$, $Y$, which satisfy (4.4), (4.5), now allow us to compute the last integrand in (4.5) to obtain the result. □

*Remark 4.1* In the special case $f=0$ we can omit the hypotheses

$(B_3)$ and $(B_4)$ and the variable $X_t^3$. Further simplifications allow Theorem 4.1 to be rephrased as:

Assume $(B_1)$, $(B_2)$, $f=0$, and one of (i), (ii) or
(iii)′ *a, A, B are constant in the variable y and* $G_2^j=0$.

Then

$$\rho(X_T\in S|\mathscr{Y}_T)$$

$$=\int_{\mathbb{R}^N}\exp\{-\tfrac{1}{2}x^*N(T,Y)x\}\int_{\mathbb{R}^{N_1}\times\mathbb{R}^{N_2}\times\mathbb{R}^N}1_S\left[\begin{pmatrix}z^1\\z^2\end{pmatrix}+\Psi(T,Y)x\right]$$

$$\times\exp\{x^*z^3\}\mu(dz;m_T,R_T)P_0(dx)$$

where $z^*=(z^{1*},z^{2*},z^{3*})$,

$$N(T,Y)=\int_0^T\Psi(t,Y)^*H(t,Y)^*H(t,Y)\Psi(t,Y)\,dt,$$

$$d\Psi(t,Y)=\begin{pmatrix}A(t,Y) & 0\\ C_1(t,Y) & C_2(t,Y)\end{pmatrix}\Psi(t,Y)\,dt+\sum_j\begin{pmatrix}0\\ G^j(t,Y)\end{pmatrix}\Psi(t,Y)\,dY_t^j,$$

and $m, R$ satisfy (4.1)–(4.2) with

$$\hat{A}=\begin{bmatrix}A & 0 & 0\\ C_1 & C_2 & 0\\ -\Psi^*H^*H_1 & -\Psi^*H^*H_2 & 0\end{bmatrix},\quad \hat{a}=\begin{bmatrix}a\\ c\\ -\Psi^*H^*h\end{bmatrix},$$

$$\hat{B}=\begin{bmatrix}B & 0\\ 0 & D\\ 0 & 0\end{bmatrix},\quad \hat{G}^j=\begin{bmatrix}0 & 0 & 0\\ G_1^j & G_2^j & 0\\ 0 & 0 & 0\end{bmatrix},\quad g^j=\begin{bmatrix}0\\ \hat{g}^j\\ (H^j\Psi)^*\end{bmatrix},$$

$$\hat{H}=(H_1\ \ H_2\ \ 0).$$

## Acknowledgements

The second author is very grateful to Drs. Bobrovsky, Zeitouni and Shukhman for sending early versions of their work relating to this problem. The authors of the present paper want to acknowledge priority to these authors for those results which appear in Shukhman [14] and Zeitouni and Bobrovsky [16].

## References

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

[2] V. E. Beneš, New exact nonlinear filters with large Lie algebras, *Systems & Control Letters* **5** (1985), 217–221.

[3] V. E. Beneš and I. Karatzas, Estimation and control of linear, partially observed systems with non-Gaussian initial distribution, *Stoch. Proc. Applic.* **14** (1983), 233–248.

[4] M. Chaleyat-Maurel and D. Michel, Des résultats de non-existence de filtre de dimension finie, *Stochastics* **13** (1984), 83–102.

[5] M. Hazewinkel and S. I. Marcus, On Lie algebras and finite dimensional filters, *Stochastics* **7** (1982), 29–62.

[6] J. Jacod, Une condition d'existence et d'unicité pour les solutions fortes d'équations différentielles stochastique, *Stochastics* **4** (1980), 23–38.

[7] G. Kallianpur, *Stochastic Filtering Theory*, Springer-Verlag, New York, 1980.

[8] R. E. Kalman and R. S. Bucy, New results in linear filtering and prediction theory, *Trans. ASME J. Basic Engrg.* **83D** (1961), 95–108.

[9] W. J. Kolodziej and R. R. Mohler, State estimation and control of conditionally linear systems, *SIAM J. Control and Optimization* **24** (1986), 497–508.

[10] R. S. Liptser and A. N. Shiryayev, *Statistics of Random Processes*, 2 volumes, Springer-Verlag, New York, 1977.

[11] A. M. Makowski, Filtering formulae for partially observed linear systems with non-Gaussian initial conditions, *Stochastics* **16** (1986), 1–24.

[12] D. Ocone, Stochastic calculus of variations for stochastic partial differential equations, *J. Functional Anal.*, to appear; see also: Probability densities for conditional statistics in the cubic sensor problem, preprint.

[13] R. W. Rishel, Lecture at 3rd Bad Honnef Conference on Stochastic Differential Systems, 1985.

[14] T. Shukhman, Explicit filters for linear systems and certain nonlinear systems with stochastic initial condition, 1985 preprint, Dept. of Electrical Engineering, Technion, Haifa.

[15] W. S. Wong, New classes of finite-dimensional nonlinear filters, *Systems and Control Letters* **3** (1983), 155–164.

[16] O. Zeitouni and B. Z. Bobrovski, Another approach to the equation of nonlinear filtering and smoothing with applications, October 1984 preprint, Dept. of Electrical Engineering, Technion, Haifa.

## Appendix

We establish here a minor variant of a result due to Liptser and Shiryayev [10, Lemma 11.6].

LEMMA 1 *Let* $\{W_t : t \geqq 0\}$ *be a standard Wiener process on* $(\Omega, \mathscr{F}, P)$ *with values in* $\mathbb{R}^d$, *and let* $X_0$ *be an independent Gaussian random* *variable in* $\mathbb{R}^N$. *Let* $B(\cdot)$, $Q(\cdot)$, $\gamma(\cdot)$ *be respectively* $\mathbb{R}^N \otimes \mathbb{R}^d$, $\mathbb{R}^N \otimes \mathbb{R}^N$, $\mathbb{R}^l \otimes \mathbb{R}^d$*-valued measurable functions of* $t$ *such that* $Q(t)$ *is symmetric, positive semi-definite and for* $T < \infty$

$$\int_0^T \{\operatorname{tr}[B(t)B(t)^* + Q(t)] + |\gamma(t)|^2\}\, dt < \infty.$$

*Let* $b(T) = \begin{pmatrix} b_0(T) \\ b_1(T) \end{pmatrix}$ *be a* $\mathbb{C}^{N+l}$*-valued vector and let*

$$I = E \exp\left\{ b_0(T)^* X_0 + b_1(T)^* \beta_T - \int_0^T \eta_t^* Q(t) \eta_t \, dt \right\}$$

*where*

$$\beta_t = \int_0^t \gamma(s)\, dW_s$$

$$\eta_t = X_0 + \int_0^t B(s)\, dW_s.$$

*Then there exists a constant* $k$, *a vector* $\alpha(T)$ *and a positive semi-definite symmetric matrix* $\beta(T)$ *such that*

$$I = k \exp\{\alpha(T)^* b(T) + b(T)^* \beta(T) b(T)\}.$$

*Moreover as functions of* $T$, $\alpha$ *and* $\beta$ *are continuous.*

*Proof* Let $\Gamma(t)$ be the positive semi-definite symmetric solution of

$$\frac{d\Gamma}{dt} = -2Q(t) + \Gamma(t)B(t)B(t)^*\Gamma(t), \qquad \Gamma(T) = 0.$$

Let $\bar{P}$ be defined by

$$d\bar{P} = \exp\left\{ -\int_0^T \eta_s^* \Gamma(s)\, d\eta_s - \tfrac{1}{2} \int_0^T |B(s)^* \Gamma(s) \eta_s|^2\, ds \right\} dP.$$

It is a probability measure as in Liptser and Shiryayev [10, Lemma 11.6], and by Girsanov's theorem

$$\bar{W}_t = W_t + \int_0^t B(s)^*\Gamma(s)\eta_s\,ds$$

is a standard Brownian motion on $(\Omega, \mathscr{F}, \bar{P})$, independent of $X_0$. Observe that

$$\beta_t = \int_0^t \gamma(s)\,d\bar{W}_s - \int_0^t \gamma(s)B(s)^*\Gamma(s)\eta_s\,ds$$

$$\eta_t = X_0 + \int_0^t B(s)\,d\bar{W}_s - \int_0^t B(s)B(s)^*\Gamma(s)\eta_s\,ds,$$

and for any bounded measurable $\phi$

$$\bar{E}\phi(X_0) = E\left\{\phi(X_0)\frac{d\bar{P}}{dP}\right\}$$

$$= E\left\{\phi(X_0)E\left\{\frac{d\bar{P}}{dP}\bigg| X_0\right\}\right\}$$

$$= E\phi(X_0)$$

so that $X_0$ is also Gaussian under $\bar{P}$. Hence under $\bar{P}$, $(\eta, \bar{W})$ are Gaussian and hence so are $(\eta, \bar{W}, \beta)$ and consequently also $(\eta_0, \beta_T)$, i.e. $(X_0, \beta_T) \sim N(m, R)$.

Now

$$\frac{d\bar{P}}{dP} = \exp\left\{-\int_0^T \eta_t^* Q(t)\eta_t\,dt + \delta_T\right\}$$

$$\delta_T = \tfrac{1}{2}\int_0^T \operatorname{tr}[B(t)B(t)^*\Gamma(t)]\,dt + \tfrac{1}{2}X_0^*\Gamma(0)X_0$$

so that

$$I = \bar{E}\exp\{b_0(T)^*X_0 + b_1(T)^*\beta_T - \delta_T\}$$

$$= \exp\left\{-\tfrac{1}{2}\int_0^T \operatorname{tr}[B(t)B(t)^*\Gamma(t)]\,dt\right\} J$$

$$J=\bar{E}\exp\{-\tfrac{1}{2}X_0^*\Gamma(0)X_0+b_0(T)^*X_0+b_1(T)^*\beta_T\}$$

$$=k_0\exp\{\tfrac{1}{2}b^*\tilde{R}^{-1}b+m^*R^{-1}\tilde{R}^{-1}b\}$$

where

$$b=\begin{pmatrix}b_0(T)\\ b_1(T)\end{pmatrix},\quad \tilde{R}=R^{-1}+\begin{pmatrix}\Gamma(0) & 0\\ 0 & 0\end{pmatrix}.$$

The last equality follows by direct computation.

It remains only to establish the continuity. Since $\Gamma(s)$ is continuous as a function of $T$, and since $m, R$, the mean and covariance (under $\bar{P}$) of $(X_0, \beta_T)$, are also continuous in $T$, then this last conclusion also follows.