



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Statistics & Probability Letters 70 (2004) 37–48

STATISTICS &
PROBABILITY
LETTERS

www.elsevier.com/locate/stapro

An application of the double Edgeworth expansion to a filtering model with Gaussian limit

Hiroki Masuda^{a,*}, Nakahiro Yoshida^b

^aGraduate School of Mathematics, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan

^bGraduate School of Mathematical Sciences, University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan

Received 29 August 2003; received in revised form 3 June 2004

Abstract

In a class of continuous-time filtering models with Gaussian limit, we provide a practical scheme of an approximation of a conditional expectation given finite-dimensional observations, in the light of the double Edgeworth expansion. Simple and explicit expressions up to the second order are given, so that we can easily write a computer program.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Conditional expectation; Double Edgeworth expansion; Filtering problem

1. Introduction

Let $(\Omega, \mathcal{F}, \mathbf{F} = (\mathcal{F}_t)_{t \in \mathbf{R}_+}, P)$ be a complete stochastic basis endowed with an r_w -dimensional standard Wiener process $w = (w^\alpha)_{\alpha=1}^{r_w}$, an r_L -dimensional pure jump Lévy process $L = (L^\beta)_{\beta=1}^{r_L}$ with mean zero, and one-dimensional càdlàg process $\theta^\varepsilon = (\theta_t^\varepsilon)_{t \in \mathbf{R}_+}$ independent of (w, L) . For simplicity, we suppose that a non-random limit process $\theta^0 = \lim_{\varepsilon \downarrow 0} \theta^\varepsilon$ exists, however, (see Remark A.2). On this basis, we consider the two-dimensional process $(X^\varepsilon, Y^\varepsilon) = \{(X_t^\varepsilon, Y_t^\varepsilon)\}_{t \in \mathbf{R}_+}$, $\varepsilon \in (0, 1]$,

*Corresponding author.

E-mail address: hiroki@math.kyushu-u.ac.jp (H. Masuda).

described by the following system of partially linear stochastic differential equations:

$$dX_t^\varepsilon = \{a(\theta_t^\varepsilon) + A_0^X(\theta_t^\varepsilon)X_t^\varepsilon + A_0^Y(\theta_t^\varepsilon)Y_t^\varepsilon\} dt + \sum_{\alpha=1}^{r_w} A_\alpha(\theta_t^\varepsilon) dw_t^\alpha + \varepsilon \sum_{\beta=1}^{r_L} \tilde{A}_\beta(\theta_{t-}^\varepsilon) dL_t^\beta,$$

$$X_0^\varepsilon = x_0 \tag{1}$$

and

$$dY_t^\varepsilon = \{b(\theta_t^\varepsilon) + B_0^X(\theta_t^\varepsilon)X_t^\varepsilon + B_0^Y(\theta_t^\varepsilon)Y_t^\varepsilon\} dt + \sum_{\alpha=1}^{r_w} B_\alpha(\theta_t^\varepsilon) dw_t^\alpha + \varepsilon \sum_{\beta=1}^{r_L} \tilde{B}_\beta(\theta_{t-}^\varepsilon) dL_t^\beta,$$

$$Y_0^\varepsilon = y_0, \tag{2}$$

where x_0 and y_0 are (known) constants independent of ε .

Let $m, n \in \mathbb{N}$, and put $\mathbf{X}^\varepsilon = (X_{t_1}^\varepsilon, X_{t_2}^\varepsilon, \dots, X_{t_m}^\varepsilon)^\top$ and $\mathbf{Y}^\varepsilon = (Y_{s_1}^\varepsilon, Y_{s_2}^\varepsilon, \dots, Y_{s_n}^\varepsilon)^\top$, where $(t_j)_{j=1}^m$ and $(s_j)_{j=1}^n$ are non-random time points such that $0 < t_1 < \dots < t_m$ and $0 < s_1 < \dots < s_n$; no restriction is imposed on the magnitude relation between t_m and s_n . The purpose of this note is, apart from full technical exposition, to provide a *practical scheme* for computation of the conditional expectation

$$\Pi^\varepsilon(y; g) := P[g(\mathbf{X}^\varepsilon) | \mathbf{Y}^\varepsilon = y]$$

for a reasonable class of $g : \mathbb{R}^m \rightarrow \mathbb{R}$, with a view to compute simulations. The essential tool we apply here is the *double Edgeworth expansion (DEE)*, cf. Yoshida (2003), which leads to an approximate value of $\Pi^\varepsilon(y; g)$ in terms of a power series of ε with coefficients depending on y and g . To the best of our knowledge, there are no existing results in this direction.

The model $(X^\varepsilon, Y^\varepsilon)$ may be regarded as a Gaussian filtering model randomly perturbed by (L, θ^ε) . The magnitude of the perturbation determined by the parameter ε : (X^0, Y^0) is nothing but a well investigated Gaussian system. By choosing (L, θ^ε) variously, we can flexibly express $(X^\varepsilon, Y^\varepsilon)$'s non-Gaussianity, (auto)correlation structure, and so on. To say nothing of the classical continuous-time Kalman–Bucy filter model, our model includes, for instance, a continuous-time analog of partial non-Gaussian state space (conditionally Gaussian state space models); see Shephard (1994), Carter and Kohn (1996), and Doucet and Andrieu (2001), where several Markov chain Monte Carlo algorithms were studied.

Applicability of the DEE to such stochastic differential equations in a slightly simpler form than (1) and (2) has been already mentioned in Yoshida (2003, Example 4), though neither detailed computation nor practical scheme was given there. We should note that since \mathbf{Y}^ε here is typically continuously distributed and non-Gaussian for each $\varepsilon \in (0, 1]$, we cannot write down $\Pi^\varepsilon(y; g)$ explicitly in general. Moreover, when we try to carry out the Monte Carlo procedure with a small target window around the conditioning value y as in Yoshida (2003, Example 2), it may not work well and is very time-consuming different from the DEE (see Section 3). In diffusion-based filtering models, Del Moral et al. (2001) discussed an approximation algorithm of the filter via the Monte Carlo procedure and the Euler scheme and gave some precise error bounds in L^1 -sense.

In Section 2, we shall briefly present a general framework of DEEs (without truncation); in this note we deal only with sufficiently regular situations and do not go for the weakest possible assumptions (see Section 2.2 for some technical remarks). The simple but illustrative numerical examples which exhibit usefulness of the DEE will be given in Section 3. The Appendix is devoted to derivation of the explicit DEE up to the second order.

2. Double Edgeworth expansion

2.1. The formula

Hereafter we shall suppose that the conditions of Theorem 5 of Yoshida (2003) are fulfilled: precisely, we suppose that

- (A1) a Malliavin operator \mathcal{L} exists on \mathbf{B} and in connection with \mathcal{L} , \mathbf{X}^ε and \mathbf{Y}^ε admit a *smooth stochastic expansion*, say $\mathbf{X}^\varepsilon \sim \zeta_0 + \varepsilon\zeta_1 + \varepsilon^2\zeta_2 + \dots$ and $\mathbf{Y}^\varepsilon \sim f_0 + \varepsilon f_1 + \varepsilon^2 f_2 + \dots$, as $\varepsilon \downarrow 0$ in the sense of Yoshida (2003, Definition 2), where the coefficients (ζ_j, f_j) are random vectors given by the formal Taylor expansion around $\varepsilon = 0$ (in particular, $\zeta_0 = (X_{t_1}^0, X_{t_2}^0, \dots, X_{t_m}^0)^\top$, $\zeta_1 = (X_{t_1}^{[1]}, X_{t_2}^{[1]}, \dots, X_{t_m}^{[1]})^\top$, $f_0 = (Y_{s_1}^0, Y_{s_2}^0, \dots, Y_{s_n}^0)^\top$ and $f_1 = (Y_{s_1}^{[1]}, Y_{s_2}^{[1]}, \dots, Y_{s_n}^{[1]})^\top$, where $F^{[k]} := (\partial_\varepsilon^k)_0 F^\varepsilon$, $k \in \mathbb{N}$, for a random function $\varepsilon \mapsto F^\varepsilon$);
- (A2) the determinant of the Malliavin covariance of $(\mathbf{X}^\varepsilon, \mathbf{Y}^\varepsilon)$, say Δ_ε , satisfies $\limsup_{\varepsilon \downarrow 0} P[\Delta_\varepsilon^{-p}] < \infty$ for every $p > 1$.

Let $\mathcal{S}'(\mathbb{R}^m)$ stand for the set of Schwartz tempered distributions on \mathbb{R}^m . Also, let $p^F(z)$ denote the density of a random variable F . Then, uniformly on every compact subsets of $\{y \in \mathbb{R}^n : p^{f_0}(y) > 0\}$, we have the DEE

$$\Pi^\varepsilon(y; g) \sim c_0(y; g) + \varepsilon c_1(y; g) + \varepsilon^2 c_2(y; g) + \dots$$

as $\varepsilon \downarrow 0$, where $\{c_j(y; g)\}_{j \geq 0}$ are described in terms of $\{(\zeta_j, f_j)\}_{j \geq 0}$, in particular,

$$c_0(y; g) = P[g(\zeta_0) | f_0 = y], \tag{3}$$

$$\begin{aligned} c_1(y; g) = & (p^{f_0}(y))^{-1} \left[\int_{\mathbf{R}^m} g(x) (-\partial_x) \cdot \{P[\zeta_1 | (\zeta_0, f_0) = (x, y)] p^{\zeta_0, f_0}(x, y)\} dx \right. \\ & + \int_{\mathbf{R}^m} g(x) (-\partial_y) \cdot \{P[f_1 | (\zeta_0, f_0) = (x, y)] p^{\zeta_0, f_0}(x, y)\} dx \\ & \left. - c_0(y; g) (-\partial_y) \cdot \{P[f_1 | f_0 = y] p^{f_0}(y)\} \right] \tag{4} \end{aligned}$$

with the dot denoting the divergence (see Yoshida (2003, Section 6) for details). We are concerned here with derivation of reduced forms of $c_0(y; g)$ and $c_1(y; g)$ together with numerical implementation: $c_0(y; g)$ is easily obtained due to Gaussianity of (ζ_0, f_0) , however, for $c_1(y; g)$, we should take account of the randomness of $\theta^{[1]}$. The details are deferred to the Appendix. In this note, we do not consider the higher order terms.

2.2. Some remarks on the regularity conditions

Denoting $\Phi_t^\varepsilon = (X_t^\varepsilon, Y_t^\varepsilon)^\top$, we can rewrite (1) and (2) as

$$\begin{aligned} d\Phi_t^\varepsilon &= (\Theta_t^{1,\varepsilon} + \Theta_t^\varepsilon \Phi_t^\varepsilon) dt + \Theta_t^{2,\varepsilon} dw_t + \varepsilon \Theta_{t-}^{3,\varepsilon} dL_t, \\ \Phi_0^\varepsilon &= \eta, \end{aligned}$$

where Θ_t^j and $\Theta_t^{j,\varepsilon}$, $j = 1, 2, 3$, are matrix-valued random functions (of θ_t^ε), and $\eta = (x_0, y_0)^\top$, from which we easily see that

$$\Phi_t^\varepsilon = \Psi_t^\varepsilon \left\{ \eta + \int_0^t (\Psi_s^\varepsilon)^{-1} (\Theta_s^{1,\varepsilon} ds + \Theta_s^{2,\varepsilon} dw_s + \varepsilon \Theta_{s-}^{3,\varepsilon} dL_s) \right\} \quad (5)$$

with $\Psi_s^\varepsilon = \exp(\int_0^s \Theta_u^\varepsilon du)$. Once the coefficients and the process θ^ε are concretely given, we can verify conditions (A1) and (A2) in the context of the (partial) Malliavin calculus for processes with jumps, cf. Bichteler et al. (1987). In terms of the explicit expression (5) of the solution, we primarily need sufficient integrability of the integrands of (5) (the functionals of θ^ε) as well as their sufficient smoothness in ε , under existence of higher order moments of the Lévy measure of L outside neighborhoods of the origin: especially, if θ^ε itself is described by a stochastic differential equation driven by a Lévy process independent of (w, L) , then it is possible to list a set of easily verifiable conditions. Also, if the diffusion part of Φ^ε is completely non-degenerate, then it may be possible to validate (A2) without utilizing the jump part. Further, we should note that we can actually treat more general situations, using a *truncation functional*, cf. Yoshida (2003, Theorem 4). By virtue of the truncation, not complete but local non-degeneracy of $(\mathbf{X}^\varepsilon, \mathbf{Y}^\varepsilon)$ may be sufficient to induce the validity of the DEE; indeed, the truncation technique often turns out to be inevitable when the coefficients of (1) and (2) possess certain singularity. Nevertheless, how to choose the truncation functional essentially depends on the concrete structure of θ^ε and the coefficients of (1) and (2). For such reasons, we shall not go into the greatest generality of argument in this note. One dimensionality of the processes X^ε , Y^ε and θ^ε is just for notational simplicity; we can treat cases where X^ε , Y^ε and θ^ε are multi-dimensional without any substantial change. Further, we can treat cases where the coefficients of (1) and (2) may depend on ε and t (such as $A_0^X(\theta_t^\varepsilon, \varepsilon, t)$), regarding $(\theta^\varepsilon, \varepsilon, t)$ as θ^ε ; in this sense, that L is pure jump type and has mean zero is not essential. In addition, let us remark that our Gaussian limit setting is *ad hoc* to avoid the technical problem of conditional expectations associated with (multiple) Lévy integrals, which was mentioned in Yoshida (2003, Example 3).

3. Numerical examples

For some simple cases, we shall compare the DEE with Monte Carlo simulations, where *the Monte Carlo trials are counted only when all of $(Y_{s_j}^\varepsilon)_{j=1}^n$ simultaneously hit the target windows $([y_{s_j} - 0.025, y_{s_j} + 0.025])_{j=1}^n$ for given conditioning values $(y_{s_j})_{j=1}^n$* ; in each simulation, the number of actual trials and the corresponding consumed time on 3.2 GHz PC are reported. In the first two examples, theoretical values of $\Pi^\varepsilon(y; g)$ can be written down, while this is not the case for the last two examples. Let us stress that our DEE can evaluate an approximate value of $\Pi^\varepsilon(y; g)$ in a flash,

and it is obvious that our method would work more effectively for higher-dimensional conditioning cases. The DEE formulae in all the examples as well as the theoretical expressions for $\Pi^\varepsilon(y; g)$ in the first two examples can be computed without difficulty. Hence we do not give those expressions to save space.

Example 1. Let $g(x) = x$; $m = 1$, $n = 2$ with $t_1 = s_2 = 1$ and $s_1 = 0.5$; $a(\theta) = 1.3\theta$, $b(\theta) = -0.5\theta$, all of A_0^X , A_0^Y , B_0^X , B_0^Y are null; $r_w = 2$ with $A_1(\theta) = 0.5$, $A_2(\theta) = B_1(\theta) = 1$, and $B_2(\theta) = 0.8$; $L \equiv 0$; $\varepsilon = 0.2$; $y_{0.5} = 0$ and $y_1 = 0.05$. Here θ^ε is the drifted Wiener process given by $\theta_t^\varepsilon = \varepsilon(t + 2w_t^*)$, where w^* is a standard Wiener process independent of w . This is a two-dimensional conditioning version of the numerical example in Yoshida (2003, Example 2).

Example 2. As mentioned in Section 2.2, the processes may be multi-dimensional; here is such an example. Let $g(x) = x$; $m = n = 1$ with $t_1 = s_1 = 1$; $a(\theta) = 2.5\theta$ and $b(\theta) = 1.6\theta$ and all of A_0^X , A_0^Y , B_0^X , B_0^Y are null; $r_w = 3$ with $A_1(\theta) = 1$, $A_2(\theta) = -0.5$, $A_3(\theta) = 0.7$, $B_1(\theta) = 0.2$, $B_2(\theta) = 0.6$, $B_3(\theta) = -1$; $L \equiv 0$; $\varepsilon = 0.2$. Let θ^ε be given by $\theta_t^\varepsilon = \varepsilon(0.5t + w_t^*)$ with w^* as above. Now, we consider another observed process \tilde{Y}^ε given by

$$d\tilde{Y}_t^\varepsilon = 1.5\theta_t^\varepsilon dt + 1.2 dw_t^1 + 0.5 dw_t^2 - 0.2 dw_t^3, \quad \tilde{Y}_0^\varepsilon = 0,$$

and regard $(Y^\varepsilon, \tilde{Y}^\varepsilon)$ as Y^ε . Here we set the conditioning values as $y_1 = 0.02$ and $\tilde{y}_1 = 0.1$ (Table 1).

Example 3. Let us consider a case of non-Gaussian θ^ε . Let $g(x) = x$; $m = 1$, $n = 1$ with $t_1 = s_1 = 1$; $a(\theta) = 2\theta$, $b(\theta) = 1.5\theta$, all of A_0^X , A_0^Y , B_0^X , B_0^Y are null; $r_w = 2$ with $A_1(\theta) = 3$, $A_2(\theta) = 2$, $B_1(\theta) = 0.7$, and $B_2(\theta) = 0.8$; $L \equiv 0$; $\varepsilon = 0.1$; $y_1 = 0.1$. Let L^* be a normal inverse Gaussian Lévy motion with parameters $(\alpha, \beta, \delta, \mu) = (5, 1, 1, 0)$ and define θ^ε by $\theta_t^\varepsilon = \varepsilon L_t^*$; see Barndorff-Nielsen (1998) for details of the normal inverse Gaussian distribution $NIG(\alpha, \beta, \delta, \mu)$.

Example 4. Let $g(x) = x^2$; $m = 1$, $n = 1$ with $t_1 = s_1 = 1$; $a(\theta) = b(\theta) = 0$; $A_0^X(\theta) = 2\theta$, $B_0^X(\theta) = 1.5\theta$, and A_0^Y and B_0^Y are null; $r_w = 2$ with $A_1(\theta) = 1$, $A_2(\theta) = 0$, $B_1(\theta) = 0$, and $B_2(\theta) = 0.5$; $L \equiv 0$; $\varepsilon = 0.2$; $y_1 = 0.1$. Let θ^ε be given by $\theta_t^\varepsilon = \varepsilon(t + 0.7w_t^*)$ with w^* as before (Table 2).

In each example, it could be said that the second-order term $c_1(y; g)$ reasonably improves the accuracy of the approximation.

Table 1
Results of Examples 1 and 2 with 10^7 Monte Carlo trials

	Example 1	Example 2
Theoretical value	0.205018	0.210021
Actually counted trials ($/10^7$)	4875	2808
Consumed time	2 h 23 min 56 s	2 h 46 min 54 s
Monte Carlo	0.204624	0.200184
DEE 1st order	0.039634	0.0709198
DEE 2nd order	0.169918	0.14059
DEE 1st + 2nd	0.209552	0.21151

Table 2

Results of Examples 3 and 4 with 10^6 Monte Carlo trials, where “Difference rate” = (DEE – MC)/MC

	Example 3	Example 4
Actually counted trials (/10 ⁶)	18935	38582
Consumed time	20 min 34 s	18 min 7 s
Monte Carlo	0.298084	0.876365
DEE 1st order	0.327434	1
DEE 2nd order	–0.0297154	–0.133333
DEE 1st+2nd	0.297718	0.866667
Difference rate	–0.123%	–1.106%

Appendix: Explicit computations up to the second order

Here we obtain reduced forms of $c_0(y; g)$ and $c_1(y; g)$ of (3) and (4) in order to run a computer program successfully through a PC. Note that we presuppose that the coefficients of (5) and θ^ε be smooth in ε (Section 2.1).

A.1. Some preliminaries

Put $Z^{(1)} = (X^0, Y^0)^\top$, $Z^{(2)} = (X^{[1]}, Y^{[1]})^\top$ and $Z = (Z^{(1)\top}, Z^{(2)\top})^\top$. Then

$$\begin{aligned} dZ_t &= K_t Z_t dt + (c_t dt + G_t dw_t + J_{t-} dL_t), \\ Z_0 &= (\eta^\top, 0, 0)^\top, \end{aligned} \quad (6)$$

where

$$\begin{aligned} c_t &= ((c_t^{(1)})^\top | (c_t^{(2)})^\top)^\top = (a(\theta_t^0), b(\theta_t^0) | \partial a(\theta_t^0) \theta_t^{[1]}, \partial b(\theta_t^0) \theta_t^{[1]})^\top, \\ K_t &= \begin{pmatrix} K_t^{(1)} & 0 \\ K_t^{(2)} & K_t^{(1)} \end{pmatrix} = \begin{pmatrix} A_0^X(\theta_t^0) & A_0^Y(\theta_t^0) & 0 & 0 \\ B_0^X(\theta_t^0) & B_0^Y(\theta_t^0) & 0 & 0 \\ \partial A_0^X(\theta_t^0) \theta_t^{[1]} & \partial A_0^Y(\theta_t^0) \theta_t^{[1]} & A_0^X(\theta_t^0) & A_0^Y(\theta_t^0) \\ \partial B_0^X(\theta_t^0) \theta_t^{[1]} & \partial B_0^Y(\theta_t^0) \theta_t^{[1]} & B_0^X(\theta_t^0) & B_0^Y(\theta_t^0) \end{pmatrix}, \\ G_t &= \begin{pmatrix} G_t^{(1)} \\ G_t^{(2)} \end{pmatrix} = \begin{pmatrix} A_1(\theta_t^0) & \cdots & A_{r_w}(\theta_t^0) \\ B_1(\theta_t^0) & \cdots & B_{r_w}(\theta_t^0) \\ \partial A_1(\theta_t^0) \theta_t^{[1]} & \cdots & \partial A_{r_w}(\theta_t^0) \theta_t^{[1]} \\ \partial B_1(\theta_t^0) \theta_t^{[1]} & \cdots & \partial B_{r_w}(\theta_t^0) \theta_t^{[1]} \end{pmatrix} \end{aligned}$$

and

$$J_t = \begin{pmatrix} 0 \\ J_t^{(1)} \end{pmatrix} = \begin{pmatrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \\ \tilde{A}_1(\theta_t^0) & \cdots & \tilde{A}_{r_L}(\theta_t^0) \\ \tilde{B}_1(\theta_t^0) & \cdots & \tilde{B}_{r_L}(\theta_t^0) \end{pmatrix}.$$

Note that $c^{(1)}$, $K^{(1)}$, $G^{(1)}$ and $J^{(1)}$ are deterministic, while $c^{(2)}$, $K^{(2)}$ and $G^{(2)}$ are random owing to $\theta^{[1]}$. Solving (6), we see that $Z^{(1)}$ and $Z^{(2)}$ are given by

$$Z_t^{(1)} = Q_t^{(1)}\eta + \int_0^t H_s^{(t,c^{(1)})} ds + \int_0^t H_{s-}^{(t,G^{(1)})} dw_s, \tag{7}$$

$$Z_t^{(2)} = Q_t^{(2)}\eta + \int_0^t M_s^{(t,c^{(1)},c^{(2)})}(\theta^{[1]}) ds + \int_0^t M_s^{(t,G^{(1)},G^{(2)})}(\theta^{[1]}) dw_s + \int_0^t M_s^{(t,0,J^{(1)})} dL_s, \tag{8}$$

where

$$Q_t = \begin{pmatrix} Q_t^{(1)} & 0 \\ Q_t^{(2)} & Q_t^{(1)} \end{pmatrix} = \exp\left(\int_0^t K_s ds\right),$$

$Q^{(1)}$ (resp. $Q^{(2)}$) being non-random (resp. random), and

$$H_s^{(t,i^{(1)})} = Q_t^{(1)}(Q_s^{(1)})^{-1}i_s^{(1)},$$

$$M_s^{(t,i^{(1)},i^{(2)})} = \{Q_t^{(2)}(Q_s^{(1)})^{-1} - Q_t^{(1)}(Q_s^{(1)})^{-1}Q_s^{(2)}(Q_s^{(1)})^{-1}\}i_s^{(1)} + Q_t^{(1)}(Q_s^{(1)})^{-1}i_s^{(2)}$$

for $s \in [0, t]$ and functions $i^{(1)}$ and $i^{(2)}$ on $[0, t]$; in (8), the integrands attached by the argument $(\theta^{[1]})$ are random. We shall denote by $H_{\kappa;s}^{(t,i^{(1)})}$ the κ th component of $H_s^{(t,i^{(1)})}$, in the same manner for $M_s^{(t,i^{(1)},i^{(2)})}$.

Remark A.1. Clearly Q satisfies $dQ_t = K_t Q_t dt$ with $Q_0 = I_4$, or, equivalently,

$$dQ_t^{(1)} = K_t^{(1)} Q_t^{(1)} dt, \quad Q_0^{(1)} = I_2,$$

$$dQ_t^{(2)} = (K_t^{(2)} Q_t^{(1)} + K_t^{(1)} Q_t^{(2)}) dt, \quad Q_0^{(2)} = 0.$$

Here I_k stands for the k -dimensional identity matrix. If we can simulate sample paths of $\theta^{[1]}$ in a certain way, then, using them, it is straightforward to simulate sample paths of Q ; apply, e.g., Runge–Kutta method after generating a trajectory of $\theta^{[1]}$ over the time interval $[0, t_m \vee s_n]$.

A.2. The first-order $c_0(y; g)$

Put $\mathcal{L}(\zeta_0, f_0) =: N_{m+n}(\mu^0, \Sigma^0)$, where we implicitly assume that Σ^0 is non-degenerate. Write $Q_t^{(1)} = (q_{ij}^{(1)}(t))_{i,j=1}^2$, then, it follows from (7) that $\mu^0 = (\mu_l^0)_{l=1}^{m+n}$ is given by

$$\mu_l^0 = \begin{cases} q_{11}^{(1)}(t_l)x_0 + q_{12}^{(1)}(t_l)y_0 + \int_0^{t_l} H_{1;s}^{(t_l,c^{(1)})} ds, & (l = 1, 2, \dots, m), \\ q_{21}^{(1)}(s_{l-m})x_0 + q_{22}^{(1)}(s_{l-m})y_0 + \int_0^{s_{l-m}} H_{2;s}^{(s_{l-m},c^{(1)})} ds, & (l = m + 1, m + 2, \dots, m + n), \end{cases}$$

and that all the components of Σ^0 are fully described by

$$\begin{aligned} \text{Cov}[X_{t_i}^0, X_{t_j}^0] &= \sum_{\alpha=1}^r \int_0^{t_i \wedge t_j} H_{(1,\alpha):s}^{(t_i, G^{(1)})} H_{(1,\alpha):s}^{(t_j, G^{(1)})} ds \quad (i, j = 1, 2, \dots, m), \\ \text{Cov}[Y_{s_i}^0, Y_{s_j}^0] &= \sum_{\alpha=1}^r \int_0^{s_i \wedge s_j} H_{(2,\alpha):s}^{(s_i, G^{(1)})} H_{(2,\alpha):s}^{(s_j, G^{(1)})} ds \quad (i, j = 1, 2, \dots, n), \\ \text{Cov}[X_{t_i}^0, Y_{s_j}^0] &= \sum_{\alpha=1}^r \int_0^{t_i \wedge s_j} H_{(1,\alpha):s}^{(t_i, G^{(1)})} H_{(2,\alpha):s}^{(s_j, G^{(1)})} ds \quad (i = 1, \dots, m, j = 1, \dots, n). \end{aligned}$$

Decomposing μ^0 and Σ^0 as

$$\mu^0 = \begin{matrix} 1 \\ m \\ n \end{matrix} \begin{pmatrix} \mu^{0,1} \\ \mu^{0,2} \end{pmatrix}, \quad \Sigma^0 = \begin{matrix} m & n \\ m & n \end{matrix} \begin{pmatrix} \Sigma_{11}^0 & \Sigma_{12}^0 \\ \Sigma_{21}^0 & \Sigma_{22}^0 \end{pmatrix},$$

we have $\mathcal{L}(\zeta_0 | f_0 = y) = N_m(\mu_{\zeta_0 | f_0}^0(y), \Sigma_{\zeta_0 | f_0}^0)$ with

$$\begin{aligned} \mu_{\zeta_0 | f_0}^0(y) &= \mu^{0,1} + \Sigma_{12}^0 (\Sigma_{22}^0)^{-1} (y - \mu^{0,2}), \\ \Sigma_{\zeta_0 | f_0}^0 &= \Sigma_{11}^0 - \Sigma_{12}^0 (\Sigma_{22}^0)^{-1} \Sigma_{21}^0. \end{aligned}$$

Therefore, the first-order coefficient is given by

$$c_0(y; g) = \int_{\mathbb{R}^m} g(x) \phi(z; \mu_{\zeta_0 | f_0}^0(y), \Sigma_{\zeta_0 | f_0}^0) dx, \tag{9}$$

where $\phi(\cdot; \mu_{\zeta_0 | f_0}^0(y), \Sigma_{\zeta_0 | f_0}^0)$ denotes the Gaussian density with mean vector $\mu_{\zeta_0 | f_0}^0(y)$ and covariance matrix $\Sigma_{\zeta_0 | f_0}^0$. For any $g \in \mathcal{S}'(\mathbb{R}^m)$, $c_0(y; g)$ can be easily evaluated through iterative sampling from $N_m(\mu_{\zeta_0 | f_0}^0(y), \Sigma_{\zeta_0 | f_0}^0)$.

A.3. The second-order $c_1(y; g)$

We prepare a simple lemma. Let $P^{\chi_1 | \chi_2}(\cdot)$ stand for a regular conditional distribution of χ_1 given $\sigma(\chi_2)$; we drop χ_2 in this notation if $\sigma(\chi_2)$ is trivial.

Lemma A.1. *Let \mathbf{X}, \mathbf{Y} and \mathbf{Z} be random elements taking values in some measurable spaces $(S, \mathcal{S}), (T, \mathcal{T})$ and (U, \mathcal{U}) , respectively. Assume that S and T are Borel spaces, that \mathbf{X} and \mathbf{Y} are independent and that \mathbf{Z} is $\sigma(\mathbf{X})$ -measurable. Further, let $\varphi : (S \times T, \mathcal{S} \otimes \mathcal{T}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a measurable function such that*

$$\int_{S \times T} |\varphi(x, y)| P^{(\mathbf{X}, \mathbf{Y})}(dx, dy) < \infty. \tag{10}$$

Then we have almost surely

$$\int_{S \times T} \varphi(x, y) P^{(X, Y) | Z}(dx, dy) = \int_T \int_S \varphi(x, y) P^{X | Z}(dx) P^Y(dy). \tag{11}$$

Proof. Clearly $P^{(X, Y) | Z}(\cdot)$, $P^{X | (Y, Z)}(\cdot)$ and $P^{Y | Z}(\cdot)$ exist. Since X and Y are conditionally independent given $\sigma(Z)$ under the assumptions, we have $P^{X | (Y, Z)}(\cdot) = P^{X | Z}(\cdot)$ a.s. Also $P^{Y | Z}(\cdot) = P^Y(\cdot)$ a.s. because Y and Z are independent. Therefore, the disintegration argument yields

$$\begin{aligned} \int_{S \times T} \varphi(x, y) P^{(X, Y) | Z}(dx, dy) &= \int_T \int_S \varphi(x, y) P^{X | \{Y=y, Z\}}(dx) P^{Y | Z}(dy) \\ &= \int_T \int_S \varphi(x, y) P^{X | Z}(dx) P^Y(dy), \end{aligned}$$

which ends the proof. \square

We shall apply Lemma A.1 by allocating (w, L) , $\theta^{[1]}$ and (ζ_0, f_0) to X , Y and Z , respectively; specifically, (11) implies that we may compute $c_1(y; g)$ by regarding $\theta^{[1]}$ as a deterministic process at first and then integrating it over the corresponding path-space. Under suitable integrability conditions associated with (10), the path-integration can be carried out numerically through the Monte Carlo procedure by generating sample paths of $\theta^{[1]}$ repeatedly.

Remark A.2. It should be noted that, by applying Lemma A.1, it is possible to deal with the case where θ^0 is random; then (ζ_0, f_0) (resp. (ζ_1, f_1)) turns out to be nothing but a θ^0 -conditionally (resp. $(\theta^0, \theta^{[1]})$ -conditionally) Gaussian random vector.

We proceed step by step.

Step 1. Recalling (4), we should begin with computing

$$\begin{aligned} \lambda_j^X(x, y) &= P[X_{t_j}^{[1]} | (\zeta_0, f_0) = (x, y)] \quad (j = 1, 2, \dots, m), \\ \lambda_k^Y(x, y) &= P[Y_{s_k}^{[1]} | (\zeta_0, f_0) = (x, y)] \quad (k = 1, 2, \dots, n), \\ \gamma_k^Y(y) &= P[Y_{s_k}^{[1]} | f_0 = y] \quad (k = 1, 2, \dots, n). \end{aligned}$$

Put $Q_t^{(2)} = (q_{ij}^{(2)}(t; \theta^{[1]}))_{i,j=1}^2$. Then, from (8), $X_{t_j}^{[1]}$ and $Y_{s_k}^{[1]}$ are given by

$$\begin{aligned} X_{t_j}^{[1]} &= \left\{ q_{11}^{(2)}(t_j; \theta^{[1]})x_0 + q_{12}^{(2)}(t_j; \theta^{[1]})y_0 + \int_0^{t_j} M_{1;s}^{(t_j, c^{(1)}, c^{(2)})}(\theta^{[1]}) ds \right. \\ &\quad \left. + \sum_{\alpha=1}^{r_w} \int_0^{t_j} M_{(1, \alpha);s}^{(t_j, G^{(1)}, G^{(2)})}(\theta^{[1]}) dw_s^\alpha \right\} + \sum_{\beta=1}^{r_L} \int_0^{t_j} M_{(1, \beta);s^-}^{(t_j, 0, J^{(1)})} dL_s^\beta \\ &= X_{t_j}^{[1],1}(\theta^{[1]}) + X_{t_j}^{[1],2}, \quad \text{say,} \end{aligned} \tag{12}$$

and

$$\begin{aligned}
 Y_{s_k}^{[1]} &= \left\{ q_{21}^{(2)}(s_k; \theta^{[1]})x_0 + q_{22}^{(2)}(s_k; \theta^{[1]})y_0 + \int_0^{s_k} M_{2:s}^{(s_k, c^{(1)}, c^{(2)})}(\theta^{[1]}) ds \right. \\
 &\quad \left. + \sum_{\alpha=1}^{r_w} \int_0^{s_k} M_{(2,\alpha):s}^{(s_k, G^{(1)}, G^{(2)})}(\theta^{[1]}) dw_s^\alpha \right\} + \sum_{\beta=1}^{r_L} \int_0^{s_k} M_{(2,\beta):s-}^{(s_k, 0, J^{(1)})} dL_s^\beta \\
 &= Y_{s_k}^{[1],1}(\theta^{[1]}) + Y_{s_k}^{[1],2} \quad \text{say.}
 \end{aligned} \tag{13}$$

Suppose that L_1 admits moments of any order; recall that $P[L_1] = 0$ is presupposed. Independence between (ζ_0, f_0) and L , (12), (13), and Lemma A.1 together imply that

$$\begin{aligned}
 \lambda_j^X(x, y) &= \int \int X_{t_j}^{[1],1}(a) dP^{w|\{(\zeta_0, f_0)=(x, y)\}} P^{\theta^{[1]}}(da) + \int X_{t_j}^{[1],2} dP^L, \\
 \lambda_k^Y(x, y) &= \int \int Y_{s_k}^{[1],1}(a) dP^{w|\{(\zeta_0, f_0)=(x, y)\}} P^{\theta^{[1]}}(da) + \int Y_{s_k}^{[1],2} dP^L, \\
 \gamma_k^Y(y) &= \int \int Y_{s_k}^{[1],1}(a) dP^{w|\{f_0=y\}} P^{\theta^{[1]}}(da) + \int Y_{s_k}^{[1],2} dP^L,
 \end{aligned}$$

and it is easy to see that all the second terms on the right-hand side vanish: we implicitly assume that the deterministic process $(M_s^{(t, 0, J^{(1)})})_{s \in \mathbb{R}_+}$ is locally square-integrable. We shall write $P^{\theta^{[1]}}[F] = \int F(a)P^{\theta^{[1]}}(da)$ in the sequel.

Step 2. It follows from the $\theta^{[1]}$ -conditional Gaussianity of $(X_{t_j}^{[1],1}(\theta^{[1]}), \zeta_0, f_0)$ that

$$P^{\theta^{[1]}} \left[\int X_{t_j}^{[1],1}(\cdot) dP^{w|\{(\zeta_0, f_0)=(x, y)\}} \right] = P^{\theta^{[1]}}[\mu_j^{[1],X}] + P^{\theta^{[1]}}[\Sigma_j^{[1],X}](\Sigma^0)^{-1} \left\{ \begin{pmatrix} x \\ y \end{pmatrix} - \mu^0 \right\},$$

where

$$\mu_j^{[1],X}(a) = q_{11}^{(2)}(t_j; a)x_0 + q_{12}^{(2)}(t_j; a)y_0 + \int_0^{t_j} M_{1:s}^{(t_j, c^{(1)}, c^{(2)})}(a) ds$$

and the (l, l) -element of $\Sigma_j^{[1],X}(a) \in \mathbb{R} \otimes \mathbb{R}^{m+n}$ is given by

$$\begin{aligned}
 Cov[X_{t_j}^{[1],1}(a), X_{t_l}^0] &= \sum_{\alpha=1}^{r_w} \int_0^{t_j \wedge t_l} M_{(1,\alpha):s}^{(t_j, G^{(1)}, G^{(2)})}(a) H_{(1,\alpha):s}^{(t_l, G^{(1)})} ds \quad (l = 1, 2, \dots, m), \\
 Cov[X_{t_j}^{[1],1}(a), Y_{s_{l-m}}^0] &= \sum_{\alpha=1}^{r_w} \int_0^{t_j \wedge s_{l-m}} M_{(1,\alpha):s}^{(t_j, G^{(1)}, G^{(2)})}(a) H_{(2,\alpha):s}^{(s_{l-m}, G^{(1)})} ds \\
 &\quad \times (l = m + 1, m + 2, \dots, m + n);
 \end{aligned}$$

note that, at this point, we regard “ a ” as a deterministic function. For notational simplicity, we write

$$\begin{aligned}
 \tilde{\Sigma}_j^{[1],X}(\cdot)(\Sigma^0)^{-1} &=: (\xi_{j:1}^X(\cdot), \dots, \xi_{j:m}^X(\cdot) | \xi_{j:m+1}^X(\cdot), \dots, \xi_{j:m+n}^X(\cdot)) \\
 &=: (\Xi_{j:1}^X(\cdot) | \Xi_{j:2}^X(\cdot)).
 \end{aligned}$$

Then we have

$$\lambda_j^X(x, y) = P^{\theta^{[1]}}[\mu_j^{[1],X}] + P^{\theta^{[1]}}[\Xi_{j:1}^X](x - \mu^{0,1}) + P^{\theta^{[1]}}[\Xi_{j:2}^X](y - \mu^{0,2}), \tag{14}$$

and clearly $\partial_{x_j} \lambda_j^X(x, y) = P^{\theta^{[1]}}[\xi_{j:j}^X]$. As for $\lambda_k^Y(x, y)$ and $\gamma_k^Y(y)$, we can derive similar expressions, say

$$\lambda_k^Y(x, y) = P^{\theta^{[1]}}[\mu_k^{[1],Y}] + P^{\theta^{[1]}}[\Xi_{k:1}^Y](x - \mu^{0,1}) + P^{\theta^{[1]}}[\Xi_{k:2}^Y](y - \mu^{0,2}), \tag{15}$$

$$\gamma_k^Y(y) = P^{\theta^{[1]}}[\mu_k^{[1],Y}] + P^{\theta^{[1]}}[\psi_k^Y](y - \mu^{0,2}). \tag{16}$$

Step 3. Put $(\Sigma^0)^{-1} =: (\sigma_{i;j}^{-1})_{i,j=1}^{m+n}$, and then write $\sigma_{l:(1)}^{-1} = (\sigma_{l:1}^{-1}, \dots, \sigma_{l:m}^{-1})$ and $\sigma_{l:(2)}^{-1} = (\sigma_{l:m+1}^{-1}, \dots, \sigma_{l:m+n}^{-1})$ for $l = 1, 2, \dots, m+n$. It follows from (14) that

$$\begin{aligned} &(-\partial_x) \cdot \{P[\zeta_1 | (\zeta_0, f_0) = (x, y)] p^{\zeta_0 f_0}(x, y)\} (p^{f_0}(y))^{-1} \\ &= p^{\zeta_0 f_0}(x|y) \sum_{j=1}^m \{\lambda_j^X(x, y)(\sigma_{j:(1)}^{-1}(x - \mu^{0,1}) + \sigma_{j:(2)}^{-1}(y - \mu^{0,2})) - P^{\theta^{[1]}}[\xi_{j:j}^X]\}. \end{aligned} \tag{17}$$

Similarly, with the obvious notation, we can get

$$\begin{aligned} &(-\partial_y) \cdot \{P[f_1 | (\zeta_0, f_0) = (x, y)] p^{\zeta_0 f_0}(x, y)\} (p^{f_0}(y))^{-1} \\ &= p^{\zeta_0 f_0}(x|y) \sum_{k=1}^n \{\lambda_k^Y(x, y)(\sigma_{m+k:(1)}^{-1}(x - \mu^{0,1}) + \sigma_{m+k:(2)}^{-1}(y - \mu^{0,2})) - P^{\theta^{[1]}}[\xi_{k:m+k}^Y]\}, \end{aligned} \tag{18}$$

and, putting $(\Sigma_{22}^0)^{-1} = (\sigma_k^{(22),-1})_{k=1}^n$ with $\sigma_k^{(22),-1} \in \mathbb{R} \otimes \mathbb{R}^n$ and $\psi_k^Y(\theta^{[1]}) = (\psi_{k:1}^Y(\theta^{[1]}), \dots, \psi_{k:n}^Y(\theta^{[1]}))$, it also follows that

$$(-\partial_y) \cdot \{P[f_1 | f_0 = y] p^{f_0}(y)\} (p^{f_0}(y))^{-1} = \sum_{k=1}^n \{\gamma_k^Y(y) \sigma_k^{(22),-1}(y - \mu^{0,2}) - P^{\theta^{[1]}}[\psi_{k:k}^Y]\}. \tag{19}$$

Step 4. Putting together expressions (14)–(19) and the original formula (4), we now conclude that $c_1(y; g)$ can be written as

$$c_1(y; g) = \sum_{j=1}^m V_j^{(1)}(y; g) + \sum_{k=1}^n V_k^{(2)}(y; g) + c_0(y; g) \left\{ \sum_{j=1}^m V_j^{(3)}(y) + \sum_{k=1}^n V_k^{(4)}(y) \right\}, \tag{20}$$

where each term is given as follows:

$$\begin{aligned} V_j^{(1)}(y; g) &= S_j^{(1)}(y) U_1(y; g) + \sigma_{j:(1)}^{-1} U_2(y; g) P^{\theta^{[1]}}[\Xi_{j:1}^X], \\ V_k^{(2)}(y; g) &= S_k^{(2)}(y) U_1(y; g) + \sigma_{m+k:(1)}^{-1} U_2(y; g) P^{\theta^{[1]}}[\Xi_{k:1}^Y], \\ V_j^{(3)}(y) &= P^{\theta^{[1]}}[\mu_j^{[1],X}] \sigma_{j:(2)}^{-1}(y - \mu^{0,2}) + \sigma_{j:(2)}^{-1}(y - \mu^{0,2}) P^{\theta^{[1]}}[\Xi_{j:2}^X](y - \mu^{0,2}) - P^{\theta^{[1]}}[\xi_{j:j}^X], \\ V_k^{(4)}(y) &= P^{\theta^{[1]}}[\mu_k^{[1],Y}] \sigma_{m+k:(2)}^{-1}(y - \mu^{0,2}) + \sigma_{m+k:(2)}^{-1}(y - \mu^{0,2}) P^{\theta^{[1]}}[\Xi_{k:2}^Y](y - \mu^{0,2}) \\ &\quad - P^{\theta^{[1]}}[\xi_{k:m+k}^Y] - \sigma_k^{(22),-1}(y - \mu^{0,2}) \{P^{\theta^{[1]}}[\mu_k^{[1],Y}] + P^{\theta^{[1]}}[\psi_k^Y](y - \mu^{0,2})\} + P^{\theta^{[1]}}[\psi_{k:k}^Y] \end{aligned}$$

with

$$\begin{aligned}
 S_j^{(1)}(y) &= P^{\theta^{[1]}}[\mu_j^{[1],X}]\sigma_{j:(1)}^{-1} + \sigma_{j:(2)}^{-1}(y - \mu^{0,2})P^{\theta^{[1]}}[\Xi_{j:1}^X] + P^{\theta^{[1]}}[\Xi_{j:2}^X](y - \mu^{0,2})\sigma_{j:(1)}^{-1}, \\
 S_k^{(2)}(y) &= P^{\theta^{[1]}}[\mu_k^{[1],Y}]\sigma_{m+k:(1)}^{-1} + \sigma_{m+k:(2)}^{-1}(y - \mu^{0,2})P^{\theta^{[1]}}[\Xi_{k:1}^Y] + P^{\theta^{[1]}}[\Xi_{k:2}^Y](y - \mu^{0,2})\sigma_{m+k:(1)}^{-1}, \\
 U_1(y; g) &= \int_{\mathbb{R}^m} g(x)(x - \mu^{0,1})p^{\zeta_0/f_0}(x|y) dx, \\
 U_2(y; g) &= \int_{\mathbb{R}^m} g(x)(x - \mu^{0,1})(x - \mu^{0,(1)})^\top p^{\zeta_0/f_0}(x|y) dx.
 \end{aligned}$$

Thus we have seen that, once the coefficients of (1) and (2) meeting (A1) and (A2) are given, all the terms appearing in (9) and (20) can be *automatically* evaluated through the Monte Carlo procedure including path-integrations with respect to $\theta^{[1]}$. By applying some standard numerical techniques, it is easy to write a corresponding computer program.

References

- Barndorff-Nielsen, O.E., 1998. Processes of normal inverse Gaussian type. *Finance Stochastic* 2, 41–68.
- Bichteler, K., Gravereaux, J., Jacod, J., 1987. Malliavin Calculus for Processes with Jumps. in: *Stochastics Monographs*, Vol. 2. Gordon and Breach Science Publishers, New York.
- Carter, C.K., Kohn, R., 1996. Markov chain Monte Carlo in conditionally Gaussian state space models. *Biometrika* 83, 589–601.
- Del Moral, P., Jacod, J., Protter, P., 2001. The Monte Carlo method for filtering with discrete-time observations. *Probab. Theory Related Fields* 3, 346–368.
- Doucet, A., Andrieu, C., 2001. Iterative algorithms for state estimation of jump Markov linear systems. *IEEE Trans. Signal Process.* 49, 1216–1227.
- Shephard, N., 1994. Partial non-Gaussian state space. *Biometrika* 81, 115–131.
- Yoshida, N., 2003. Conditional expansions and their applications. *Stochastic Process. Appl.* 107, 53–81.