Conditional Moment Generating Functions for Integrals and Stochastic Integrals: Maximum-Likelihood Estimation

Charalambos D. Charalambous*
Department of Electrical Engineering
McGill University
3480 University Street, Montréal, P.Q. Canada H3A 2A7
email: chadcha@cim.mcgill.ca

Abstract

In this paper we present two methods for computing filtered estimates for moments of integrals and stochastic integrals of continuous-time nonlinear systems. The first method utilizes recursive stochastic partial differential equations. The second method utilizes conditional moment generating functions. For the case of Gaussian systems the recursive computations involve integrations with respect to Gaussian densities, while the moment generating functions involve differentiations of parameter dependent ordinary stochastic differential equations. The second method is applied in the expectation maximization algorithm.

1. Introduction

This paper discusses the following problem. We are given noisy observations $\{y_s; 0 \leq s \leq t\}$ of the system state process $\{x_s; 0 \leq s \leq t\}$, and we wish to derive filtered estimates for moments of integrals and stochastic integrals.

Specifically,

$$dx_t = f(t, x_t)dt + \sigma(t, x_t)dw_t,\ x(0) \in \mathbb{R}^n, \quad (1)$$

$$dy_t = h(t, x_t)dt + \alpha_t dw_t + N_t^{1/2} db_t,\ y(0) = 0 \in \mathbb{R}^n, \quad (2)$$

where $x_t \in \mathbb{R}^n, y_t \in \mathbb{R}^d$ and $\{w_s; 0 \leq s \leq t\}$, $\{b_s; 0 \leq s \leq t\}$ are independent standard Wiener processes; $x(0)$ is a random variable independent of the Wiener processes.

We are interested in conditional expectations (filtered estimates) of moments of integrals and stochastic integrals

$$L_{0,t}^{\kappa,1} = \left(\int_0^t f_1(s, x_s)ds\right)^{\kappa}, \quad L_{0,t}^{\kappa,2} = \left(\int_0^t f_2(s, x_s)dw_s\right)^{\kappa}, \quad (3)$$

$$L_{0,t}^{\kappa,3} = \left(\int_0^t f_3(s, x_s)db_s\right)^{\kappa}, \quad \kappa \geq 1.$$

Aside from their mathematical value, these estimates are important, for example, in least-squares estimation/filtering, Volterra series expansions of nonlinear realization theory [1], Wiener Chaos expansions (of nonlinear filtering) [2], Maximum Likelihood Estimation.

The first method, Theorem 3.2, utilizes a recursive system of stochastic partial differential equations (SPDE's). The second method, Theorem 3.9, utilizes conditional moment generating functions of $L_{0,t}^{1,j}, j = 1, 2, 3$. That is, for a test function $\Phi : \mathbb{R}^n \to \mathbb{R}$, we use measure-valued conditional moment generating functions

$$\widetilde{\beta}_t^{\theta,j}(\Phi) = \widetilde{\mathrm{E}}[\Phi(x_t) \exp\left(\theta L_{0,t}^{1,j}\right) |\mathcal{F}_{0,t}^y],\ \theta = i\,\omega, \quad i = \sqrt{-1},$$

for $j = 1, 2, 3$. We show that when $\widetilde{\beta}_t^{\theta,j}(\Phi)$ have density functions, $\widetilde{\beta}^j(x,t), j = 1, 2, 3$, then

$$\lim_{\theta \to 0} \frac{d^\kappa}{d\theta^\kappa} \widetilde{\beta}_t^{\theta,j}(\Phi) = \widetilde{\mathrm{E}}[\Phi(x_t) L_{0,t}^{\kappa,j} |\mathcal{F}_{0,t}^y] \text{ w.p.1}, \ \kappa \geq 0, \quad (4)$$

$$\lim_{\theta \to 0} \frac{d^\kappa}{d\theta^\kappa} \widetilde{\beta}_t^{\theta,j}(1) = \widetilde{\mathrm{E}}[L_{0,t}^{\kappa,j} |\mathcal{F}_{0,t}^y] \text{ w.p.1}, \ \kappa \geq 0. \quad (5)$$

The unnormalized versions of $\widetilde{\beta}^{\theta,j}(x,t), j = 1, 2, 3$ satisfy linear SPDE's. For the case of Gaussian system models (i.e., $dx_t = Fx_t dt + Gw_t$, $dy_t = Hx_t dt + \alpha dw_t + N^{\frac{1}{2}} b_t$), we employ (5) to derive filtered estimates for

$$\int_0^t x_s' Q x_s ds, \quad \int_0^t x_s' R dw_s, \quad \int_0^t x_s' S db_s. \quad (6)$$

Each filtered estimate is propagated by 4 statistics; the conditional mean and error covariances of x_t given $\{y_s; 0 \leq s \leq t\}$ (Kalman filter), and modified versions of the Kalman filter. These estimates can be used in the Expectation Maximization algorithm derived in [3].

2. The DMZ Equation

Notations 2.1

1. "$'$" denotes transposition of a matrix, I_k denotes $k \times k$ identity matrices, $(\cdot)_i$ denotes the ith component of a vector, and $(\cdot)_{i,j}$ denotes the $ijth$ component of a matrix;

*This work was supported by the Natural Science and Engineering Research Council of Canada under grant OGP0183720.

2. $\mathcal{L}(V_1; V_2)$ denotes the space of linear transformations of a vector space V_1 into a vector space V_2;

3. $D_x = [\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \dots, \frac{\partial}{\partial x_n}]'$, $\quad D_x^2 = \left[\frac{\partial^2}{\partial x_i \partial x_j}\right]$;

4. $\Phi : \mathbb{R}^n \to \mathbb{R}$ denotes an arbitrary test function which is $C_x^2(\mathbb{R}^n)$ and has compact support;

5. $E, \widetilde{E}$ denote expectations with respect to measures $P, \widetilde{P}$, respectively. □

Assumptions 2.2

1. $f : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n, \sigma : [0,T] \times \mathbb{R}^n \to (\mathbb{R}^m; \mathbb{R}^n), h : [0,T] \times \mathbb{R}^n \to \mathbb{R}^d, T > 0$;

2. $N : [0,T] \to (\mathbb{R}^d; \mathbb{R}^d), \alpha : [0,T] \to \mathcal{L}(\mathbb{R}^m; \mathbb{R}^d), N, \alpha$ are continuous in t, and $\exists \beta_1 > 0$, such that $N_t \geq \beta_1 I_d$;

3. $|f(t,x)| + |h(t,x)| + \|\sigma(t,x)\| \leq k(1 + |x|)$. □

Consider the P-martingale $m_t = \int_0^t h'(s, x_s) C_s^{-1} dy_s$, $C_t \doteq \alpha_t \alpha_t' + N_t$, and introduce the exponential martingale

$$\varepsilon^{-1}(m_t) = \exp(-m_t + \frac{1}{2}\langle m, m\rangle_t), \tag{7}$$

where $\langle m, m\rangle_t = \int_0^t |C_s^{-1/2} h(s, x_s)|^2 ds$ is the quadratic variation of $\{m_t; t \in [0,T]\}$. By Assumptions 2.2, we have $\widetilde{E}[\varepsilon^{-1}(m_t)] = 1, \forall t \in [0,T]$, (see [4]). Consequently, we define a measure P through the Radon-Nikodym derivative

$$\Lambda_{0,T}^{-1} \doteq \widetilde{E}\left[\frac{dP}{d\widetilde{P}} | \mathcal{F}_{0,T}\right] = \varepsilon^{-1}(m_T). \tag{8}$$

Since $P(\Omega) = \int_\Omega \Lambda_{0,t}^{-1}(\omega) d\widetilde{P}(\omega) = 1$, $\forall t \in [0,T]$, the Girsanovs Theorem, (see [4]), states that P is a probability measure on $(\Omega, \mathcal{A})$ and that

$$\begin{bmatrix} \overline{w}_t \\ \overline{b}_t \end{bmatrix} = \begin{bmatrix} w_t \\ b_t \end{bmatrix} + \begin{bmatrix} \langle w, m\rangle_t \\ \langle b, m\rangle_t \end{bmatrix} = \begin{bmatrix} w_t \\ b_t \end{bmatrix} + \begin{bmatrix} \int_0^t \alpha_s' C_s^{-1} h(s, x_s) ds \\ \int_0^t N_s^{1/2} C_s^{-1} h(s, x_s) ds \end{bmatrix}$$

are Wiener processes. Therefore, under the probability space $(\Omega, \mathcal{F}, P; \mathcal{F}_{0,t})$ the processes $\{x_t; t \in [0,T]\}, \{y_t; t \in [0,T]\}$ are solutions of

$$\begin{aligned} dx_t &= f(t, x_t) dt - \sigma(t, x_t) \alpha_t' C_t^{-1} h(t, x_t) dt \\ &\quad + \sigma(t, x_t) d\overline{w}_t, \qquad x(0) \in \mathbb{R}^n, \end{aligned} \tag{9}$$

$$dy_t = \alpha_t d\overline{w}_t + N_t^{1/2} d\overline{b}_t, \qquad y(0) = 0 \in \mathbb{R}^n, \tag{10}$$

Notation 2.3

1. $\{\mathcal{F}_{0,t}^y; t \in [0,T]\}$ denotes the complete filtration generated by σ-algebra $\sigma\{y_\tau; 0 \leq \tau \leq t\}$;

2. The measure-valued process

$$q_t(\Phi) = E[\Phi(x_t) \Lambda_{0,t} | \mathcal{F}_{0,t}^y]$$

is well defined. □

Lemma 2.4. [5, 6] Suppose $q_t(\cdot)$ has an $\mathcal{F}_{0,t}^y$-measurable density function $q : \mathbb{R}^n \times [0,T] \times \Omega \to \mathbb{R}$. Then

$$\widetilde{E}[\Phi(x_t) | \mathcal{F}_{0,t}^y] = \frac{q_t(\Phi)}{q_t(1)} = \frac{\int_{\mathbb{R}^n} \Phi(z) q(z,t) dz}{\int_{\mathbb{R}^n} q(z,t) dz}. \quad \square \tag{11}$$

Note that $\{\Lambda_{0,t}; t \in [0,T]\}$ is given by

$$\Lambda_{0,t} = 1 + \int_0^t \Lambda_{0,s} h'(s, x_s) C_s^{-1} dy_s. \tag{12}$$

Theorem 2.5. [5, 6] Suppose $q_t(\cdot)$ has a density function $q(x,t)$. The unnormalized density of the conditional distribution $\widetilde{P}(x_t \in A | \mathcal{F}_{0,t}^y), A \in \mathcal{B}(\mathbb{R}^n)$ is $q(\cdot)$ and satisfies the SPDE

$$dq(z,t) = A(t)^* q(z,t) dt + B(t)^* q(z,t) dy_t, \; q(z,0) = p_0(z), \tag{13}$$

where

$$\begin{aligned} A(t)^*\Phi(x) &= \tfrac{1}{2}\textstyle\sum_{i,j=1}^n \left(\frac{\partial^2}{\partial x_i \partial x_j}\left((\sigma(t,x)\sigma'(t,x))_{i,j} \Phi(x)\right)\right. \\ &\quad - \textstyle\sum_{i=1}^n \frac{\partial}{\partial x_i}(f_i(t,x)\Phi(x)), \\ B_k(t)^*\Phi(x) &= \textstyle\sum_{i=1}^d (C_t^{-1})_{i,k} h_i(t,x)\Phi(x) \\ &\quad - \textstyle\sum_{i=1}^n \frac{\partial}{\partial x_i}\left((\sigma(t,x)\alpha_t' C_t^{-1})_{i,k} \Phi(x)\right). \quad \square \end{aligned}$$

Definition 2.6. A fundamental solution of (13) is an $\mathcal{F}_{0,t}^y$-measurable function $q(z,t;x,s)$, with $(z,x) \in \mathbb{R}^n \times \mathbb{R}^n, 0 \leq s < t \leq T$ such that the following hold:

1. For fixed $(s,x) \in (0,t) \times \mathbb{R}^n$, $q(\cdot, t; x, s) \in C_z^2(\mathbb{R}^n)$ and $q(\cdot, \cdot; x, s)$ satisfies (13);

2. For $\varphi : \mathbb{R}^n \to \mathbb{R}$, which is continuous with compact support

$$\lim_{t \downarrow s} \int_{-\infty}^{\infty} q(z,t;x,s)\varphi(x) dx = \varphi(z). \quad \square \tag{14}$$

Theorem 2.7. Suppose for each $s \in [0,T]$ there exists 'a random process $\{q(z,t;x,s); 0 \leq s < t \leq T\}$, $(z,x) \in \mathbb{R}^n \times \mathbb{R}^n$ which is a solution of

$$\begin{aligned} dq(z,t;x,s) &= A(t)^* q(z,t;x,s) dt \\ &\quad + B(t)^* q(z,t;x,s) dy_s, \; \lim_{t \downarrow s} q(z,t;x,s) = \delta(z-x). \end{aligned} \tag{15}$$

Then

$$\widetilde{E}[\Phi(x_t) | \mathcal{F}_{0,t}^y] = \frac{q_t(\Phi)}{q_t(1)} = \frac{\int_{\mathbb{R}^{2n}} \Phi(z) q(z,t;x,0) p_0(x) dx dz}{\int_{\mathbb{R}^{2n}} q(z,t;x,0) p_0(x) dx dz}$$

Proof. Apply the Ito differential rule to $\widetilde{q}(z,t) = \int_{\mathbb{R}^n} q(z,t;x,0) p_0(x) dx$, to show that $\widetilde{q}(z,t)$ satisfies (13) and then follow Lemma 2.4. □

3. Moment Generating Functions

Definition 3.1. Let $f_1 : [0,T] \times \mathbb{R}^n \to \mathbb{R}, f_2 : [0,T] \times \mathbb{R}^n \to (\mathbb{R}^m)', f_3 : [0,T] \times \mathbb{R}^n \to (\mathbb{R}^d)'$ be such that $\widetilde{E}[\int_0^T |f_j(t, x_t)|^2]^k < \infty, \kappa \geq 1, j = 1, 2, 3$.

1. The integrals $L^{\kappa,1}_{0,t}, L^{\kappa,2}_{0,t}, L^{\kappa,3}_{0,t}$ are well-defined.
2. The measure-valued processes

$$M^{\kappa,j}_t(\Phi) = \mathrm{E}[\Phi(x_t)\Lambda_{0,t}L^{\kappa,j}_{0,t}|\mathcal{F}^y_{0,t}], \quad \kappa \geq 0, \quad (16)$$

are well-defined for $j = 1, 2, 3$. □

For $j = 1, 2, 3$, we wish to derive expressions for

$$\widetilde{\mathrm{E}}[L^{\kappa,j}_{0,t}|\mathcal{F}^y_{0,t}] = \frac{\mathrm{E}[\Lambda_{0,t}L^{\kappa,j}_{0,t}|\mathcal{F}^y_{0,t}]}{\mathrm{E}[\Lambda_{0,t}|\mathcal{F}^y_{0,t}]}, \quad \kappa \geq 1. \quad (17)$$

3.1 Recursive Equations

Theorem 3.2. Suppose $M^{\kappa,j}_t(\cdot)$ have $\mathcal{F}^y_{0,t}$-measurable density functions $M^{\kappa,j} : \mathbb{R}^n \times [0,T] \times \Omega \to \mathbb{R}, j = 1,2,3$. Then

$$M^{\kappa,j}(x,t)dx = \mathrm{E}[I_{x_t \epsilon dx}\Lambda_{0,t}L^{\kappa,j}_{0,t}|\mathcal{F}^y_{0,t}], \quad \kappa \geq 1, \quad (18)$$

$j = 1, 2, 3$, satisfy the following recursive system of SPDE's:

$$dM^{\kappa,1}(x,t) = A(t)^*M^{\kappa,1}(x,t)dt + B(t)^*M^{\kappa,1}(x,t)dy_t + \kappa f_1(t,x)M^{k-1,1}(x,t)dt, \ \kappa \geq 1, \quad (19)$$

$$\begin{aligned} dM^{\kappa,2}(x,t) &= A(t)^*M^{\kappa,2}(x,t)dt + B(t)^*M^{\kappa,2}(x,t)dy_t \\ &+ \frac{1}{2}\kappa(\kappa-1)|f_2'(t,x)|^2M^{\kappa-2,2}(x,t)dt \\ &- \kappa\sum_{i=1}^n \frac{\partial}{\partial x_i}\left(M^{\kappa-1,2}(x,t)\left(\sigma(t,x)f_2'(x,t)\right)_i\right)dt \\ &+ \kappa f_2(t,x)M^{\kappa-1,2}(x,t)\alpha_t'C_t^{-1}dy_t, \ \kappa \geq 1, \end{aligned} \quad (20)$$

$$\begin{aligned} dM^{\kappa,3}(x,t) &= A(t)^*M^{\kappa,3}(x,t)dt + B(t)^*M^{\kappa,3}(x,t)dy_t \\ &+ \frac{1}{2}\kappa(k-1)|C^{1/2}N^{-1/2}f_3'(t,x)|^2M^{\kappa-2,3}(x,t)dt \\ &+ \kappa f_3(t,x)M^{\kappa-1,3}(x,t)N^{1/2}C^{-1}dy_t, \ \kappa \geq 1, \end{aligned} \quad (21)$$

Here we use the convention $M^{p,j}(x,t) = 0$ for $p < 0$. Also,

$$M^{\kappa,j}(x,0) = 0, \ \kappa \geq 1, \ M^{0,j}(x,t) = q(x,t), \ j = 1,2,3. \quad (22)$$

Proof. Consider $\Phi(x_t)\Lambda_{0,t}L^{\kappa,1}_{0,t}$, where $\{x_t; t \in [0,T]\}$ and $\{\Lambda_{0,t}; t \in [0,T]\}$ are solutions of (9), (12), respectively. By the Itô product rule

$$L^{\kappa,1}_{0,t} = \kappa\int_0^t L^{\kappa-1,1}_{0,s}f_1(s,x_s)ds, \quad \kappa \geq 1. \quad (23)$$

Employing the Itô product rule once again, we have

$$\begin{aligned} \Phi(x_t)\Lambda_{0,t}L^{\kappa,1}_{0,t} &= \textstyle\int_0^t \Phi(x_s)d(\Lambda_{0,s}L^{\kappa,1}_s) \\ &+ \textstyle\int_0^t d\Phi(x_s)\Lambda_{0,s}L^{\kappa,1}_{0,s} + \int_0^t \langle\Phi(x), \Lambda L^{\kappa,1}\rangle_t, \end{aligned} \quad (24)$$

$$\begin{aligned} \Lambda_{0,t}L^{\kappa,1}_{0,t} &= \textstyle\int_0^t \Lambda_{0,s}dL^{\kappa,1}_{0,s} + \int_0^t L^{\kappa,1}_{0,s}d\Lambda_{0,s} + \int_0^t d\langle\Lambda, L^{\kappa,1}\rangle_t \\ &= \kappa\textstyle\int_0^t f_1(s,x_s)\Lambda_{0,s}L^{\kappa-1,1}_{0,s}ds \\ &+ \textstyle\int_0^t \Lambda_{0,s}L^{\kappa,1}_{0,s}h'(s,x_s)C_s^{-1/2}d\widetilde{y}_s. \end{aligned}$$

Substituting into (24) we obtain

$$\begin{aligned} \Phi(x_s)\Lambda_{0,t}L^{\kappa,1}_{0,t} &= \tfrac{1}{2}\textstyle\int_0^t \Lambda_{0,s}L^{\kappa,1}_{0,s}\mathrm{Tr}\,(\sigma(s,x_s) \\ &\sigma'(s,x_s)D_x^2\Phi(x_s))\,ds + \textstyle\int_0^t \Lambda_{0,s}L^{\kappa,1}_{0,s}D_x'\Phi(x_s)\sigma(s,x_s) \\ &.D_s^{1/2}d\widetilde{w}_s + \textstyle\int_0^t \Lambda_{0,s}L^{\kappa,1}_{0,s}\Phi(x_s)h'(s,x_s)C_s^{-1/2}d\widetilde{y}_s \\ &+ \textstyle\int_0^t \Lambda_{0,s}L^{\kappa,1}_{0,s}D_x'\Phi(x_s)\sigma(s,x_s)\alpha_s'C_s^{1/2}d\widetilde{y}_s \\ &+\kappa\textstyle\int_0^t \Lambda_{0,s}L^{\kappa-1,1}_{0,s}f_1(s,x_s)ds. \end{aligned} \quad (25)$$

Conditioning each side of () on $\mathcal{F}^y_{0,t}$ using (18), and then integrating by parts, we deduce (19). When $\kappa = 0, j = 1$, we have $M^{0,1}(x,t)dx = \mathrm{E}[I_{x_t \epsilon dx}\Lambda_{0,t}|\mathcal{F}^y_{0,t}]$, and thus $M^{0,1}(x,t)$ satisfies the DMZ equation. The remaining equations are obtained using the same procedure. □

Remark 3.3. Notice that the filtered estimates for $L^{\kappa,j}_t, \kappa \geq 1, j = 1,2,3$ can be computed from

$$\widetilde{\mathrm{E}}[L^{\kappa,j}_{0,t}|\mathcal{F}^y_{0,t}] = \frac{\int_{\mathbf{R}^n} M^{\kappa,j}(z,t)dz}{\int_{\mathbf{R}^n} q(z,t)dz}, \quad \kappa \geq 1. \ \square \quad (26)$$

Lemma 3.4. Suppose $M^{\kappa,j}_t(\cdot)$ have $\mathcal{F}^y_{0,t}$-measurable density functions. Then

$$M^{\kappa,1}(z,t) = \kappa\int_0^t\int_{\mathbf{R}^n} f_1(s,x)M^{\kappa-1,1}(x,s)q(z,t;x,s)dxds,$$

$$\begin{aligned} M^{\kappa,2}(z,t) &= \frac{1}{2}\kappa(\kappa-1)\int_0^t\int_{\mathbf{R}^n}|f_2(s,x)|^2M^{k-2,2}(x,s) \\ &.q(z,t;x,s)dxds - \kappa\int_0^t\int_{\mathbf{R}^n}\sum_{i=1}^n\frac{\partial}{\partial x_i}\Big(M^{k-1,2}(x,s) \\ &.\left(\sigma(s,x)f_2'(s,x)\right)_i\Big)q(z,t;x,s)dxds \\ &+ \kappa\int_0^t\int_{\mathbf{R}^n} f_2(s,x)M^{k-1,2}(x,s)\alpha_s'C_s^{-1}q(z,t;x,s)dxdy_s, \end{aligned}$$

$$\begin{aligned} M^{\kappa,3}(z,t) &= \frac{1}{2}\kappa(k-1)\int_0^t\int_{\mathbf{R}^n}|C_s^{1/2}N_s^{-1/2}f_3(s,x)|^2 \\ &.M^{k-2,3}(x,s)q(z,t;x,s)dxds + \kappa\int_0^t\int_{\mathbf{R}^n} f_3(s,x) \\ &.M^{k-1,3}(x,s)N^{1/2}C_s^{-1}q(z,t;x,s)C_s^{-1}dxdy_s, \end{aligned}$$

where $\kappa > 1$ and $M^{p,j}(x,t) = 0$ for $p < 0$, $j = 1,2,3$.

Proof. Follow the derivation of Theorem 2.7. □

3.2 Moment Generating Functions

Next we introduce moment generating functions for computing the conditional moments of integrals and stochastic integrals (17).

Definition 3.5. Let $\theta = i\omega, i = \sqrt{-1}$.

1. The measure-valued conditional moment generating functions of the stochastic processes $\{L^{1,j}_{0,t}; t \in [0,T]\}$, given by

$$\widetilde{\beta}^{\theta,j}_t(\Phi) = \widetilde{\mathrm{E}}[\Phi(x_t)\exp\left(\theta L^{1,j}_{0,t}\right)|\mathcal{F}^y_{0,t}], \ j = 1,2,3, \quad (27)$$

are well-defined.

2. The measure-valued unnormalized conditional moment generating functions of the stochastic processes $\{L_{0,t}^{1,j}; t \in [0,T]\}$ given by

$$\beta_t^{\theta,j}(\Phi) = \mathrm{E}[\Phi(x_t)\Lambda_{0,t} \exp\left(\theta L_{0,t}^{1,j}\right) |\mathcal{F}_{0,t}^y],\ j = 1,2,3, \tag{28}$$

are well-defined. □

Lemma 3.6. Suppose $\beta_t^{\theta,j}(\cdot)$ have $\mathcal{F}_{0,t}^y$-measurable density function $\beta^{\theta,j} : \mathbb{R}^n \times [0,T] \times \Omega \to \mathbb{R}$.

1. Then

$$\widetilde{\mathrm{E}}[\Phi(x_t) \exp\left(\theta L_{0,t}^{1,j}\right) |\mathcal{F}_{0,t}^y] = \frac{\beta_t^{\theta,j}(\Phi)}{q_t(1)}$$
$$= \frac{\int_{\mathbb{R}^n} \Phi(z)\beta^{\theta,j}(z,t)dz}{\int_{\mathbb{R}^n} q(z,t)dz}, \quad j = 1,2,3. \tag{29}$$

2. The conditional characteristic functions of of the stochastic processes $\{L_{0,t}^{1,j}; t \in [0,T]\}$, are given by

$$\widetilde{\mathrm{E}}[\exp\left(i\omega L_{0,t}^{1,j}\right) |\mathcal{F}_{0,t}^y] = \frac{\beta_t^{i\omega,j}(1)}{q_t(1)}$$
$$= \frac{\int_{\mathbb{R}^n} \beta^{i\omega,j}(z,t)dz}{\int_{\mathbb{R}^n} q(z,t)dz}, \quad j = 1,2,3. \tag{30}$$

Proof. Similar to Lemma 2.4. □

Theorem 3.7. Suppose $\beta_t^{\theta,j}(\cdot)$ have $\mathcal{F}_{0,t}^y$-measurable density functions $\beta^{\theta,j}(\cdot), j = 1,2,3$.
The densities of the measure-valued unnormalized conditional moment generating functions, namely,

$$\beta^{\theta,j}(x,t)dx = \mathrm{E}[I_{x_t \in dx}\Lambda_{0,t} \exp\left(\theta L_{0,t}^{1,j}\right) |\mathcal{F}_{0,t}^y], \tag{31}$$

where $j = 1,2,3$ satisfy the following system of SPDE's:

$$d\beta^{\theta,1}(x,t) = A(t)^*\beta^{\theta,1}(x,t)dt + B(t)^*\beta^{\theta,1}(x,t)dy_t$$
$$+ \theta f_1(t,x)\beta^{\theta,1}(x,t)dt, \tag{32}$$

$$d\beta^{\theta,2}(x,t) = A(t)^*\beta^{\theta,2}(x,t)dt + B(t)^*\beta^{\theta,2}(x,t)dy_t$$
$$+ \frac{\theta}{2}|f_2'(t,x)|^2\beta^{\theta,2}(x,t)dt$$
$$- \theta \sum_{i=1}^n \frac{\partial}{\partial x_i}\left(\left(\sigma(t,x)f_2'(x,t)\right)_i \beta^{\theta,2}(x,t)\right) dt$$
$$+ \theta f_2(t,x)\beta^{\theta,2}(x,t)\alpha_t' C_t^{-1} dy_t, \tag{33}$$

$$d\beta^{\theta,3}(x,t) = A(t)^*\beta^{\theta,3}(x,t)dt + B(t)^*\beta^{\theta,3}(x,t)dy_t$$
$$+ \frac{\theta^2}{2}|C^{1/2}N^{-1/2}f_3'(t,x)|^2\beta^{\theta,3}(x,t)dt$$
$$+ \theta f_3(t,x)\beta^{\theta,3}(x,t)N^{1/2}C^{-1}dy_t. \tag{34}$$

The initial conditions are

$$\beta^{\theta,j}(x,0) = p_0(x), \qquad x \in \mathbb{R}^n, \qquad j = 1,2,3. \tag{35}$$

Proof. First, absorb $\exp\left(\theta L_{0,t}^{1,j}\right)$ in the exponential term $\Lambda_{0,t}$ by setting

$$\widehat{\Lambda}_{0,t}^j = \Lambda_{0,t} \exp\left(\theta L_{0,t}^{1,j}\right).$$

Second, apply the Itô product rule as in Theorem 3.2. This derivation is along the lines of information state equations in [6]. □

Proposition 3.8. Suppose $\widetilde{\mathrm{E}}[\exp\left(\int_0^T |f_j(t,x_t)|^2 dt\right)] < \infty, j = 1,2,3$.
Then

$$E\left[\Phi(x_t)\Lambda_{0,t} \exp\left(\theta L_{0,t}^{1,1}\right) |\mathcal{F}_{0,t}^y\right] = E\left[\Phi(x_t)\Lambda_{0,t}|\mathcal{F}_{0,t}^y\right]$$
$$+ \sum_{\kappa=1}^\infty \frac{\theta^\kappa}{\kappa!} E\left[\Phi(x_t)\left(\int_0^t f_i(s,x_s)\right)^\kappa |\mathcal{F}_{0,t}^y\right], \tag{36}$$

where the infinite series converges in $L^1(\Omega, \mathcal{F}_{0,t}^y, P)$.
Moreover,

$$\beta_t^{\theta,j}(\Phi) = q_t(\Phi) + \sum_{\kappa=1}^\infty \frac{\theta^\kappa}{\kappa!} M_t^{\kappa,j}(\Phi), \qquad j = 1,2,3. \tag{37}$$

Proof. Similar to [2]. □

Theorem 3.9. Suppose $\widetilde{\mathrm{E}}[|L_{0,t}^{1,j}|^\kappa] < \infty$ for some positive integer κ, $j = 1,2,3$.
Then for $j = 1,2,3$

1. $\beta^{i\omega,j}(1)$ have κ continuous derivatives with respect to ω *w.p.*1;

2.

$$\lim_{\theta \to 0} \frac{d^\kappa}{d\theta^\kappa} \frac{\beta_t^{\theta,j}(\Phi)}{q_t(1)} = \widetilde{\mathrm{E}}[\Phi(x_t)L_{0,t}^{\kappa,j}|\mathcal{F}_{0,t}^y] \quad \text{w.p.1.} \tag{38}$$

$$\lim_{\theta \to 0} \frac{d^\kappa}{d\theta^\kappa} \frac{\beta_t^{\theta,j}(1)}{q_t(1)} = \widetilde{\mathrm{E}}[L_{0,t}^{\kappa,j}|\mathcal{F}_{0,t}^y] \quad \text{w.p.1.} \tag{39}$$

Proof. The derivation is based on Kolmogorovs continuity theorem and its application to parameter dependent diffusion processes (see [7]). First, note that if the measure-valued processes $\beta_t^{\theta,j}(\cdot)$ have density function then (32)-(34) hold. If $\beta^{\theta,j}(x,t)$ are in the function space of continuous functions, then their derivatives with respect to θ will also be continuous; this is done as in [7]. Hence, by normalizing (37), as $\theta \to 0$, the left-hand-side of (38) and (39) converge in distribution provided the density functions $M^{\kappa,j}(x,t)$ exist. The a.s. convergence is established as follows. For each measure-valued process $\beta_t^{\theta,j}(\cdot), j = 1,2,3$ there is a stochastic ordinary differential equation analogous to (25). A direct application of the Blagovescenskii and Freidlin [7] result establishes the a.s. continuity and convergence. □

3.3 Expectation-Maximization

Consider the system

$$dx_t = Fx_t dt + G dw_t, \qquad x(0) \in \mathbb{R}^n,$$
$$dy_t = Hx_t dt + N^{\frac{1}{2}} db_t, \qquad y(0) = 0,$$
$$f_1(t,x) = \frac{1}{2}x'Qx, \quad f_2(t,x) = x'R, \quad f_3(t,x) = x'S.$$

Here $Q = Q'$. We assume $x(0)$ is a Gaussian random variable.

Suppose F, H are random matrices which we wish to identify or estimate. The expectation-maximization algorithm, (see [3]) enables computation of maximum-likelihood estimates of F, H, in terms of filtered estimates of the processes $\int_0^t f_1(s,x_s)ds$, $\int_0^t f_2(s,x_s)dw_s$, $\int_0^t f_3(s,x_s)db_s$. Here we apply Theorem 3.9 to obtain the filtered estimates of these integrals.

A solution of (13) is

$$q(x,t) = \frac{1}{(2\pi)^{\frac{n}{2}}|P_t^0|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}|P_t^{0,-\frac{1}{2}}\left(x - \widehat{x}_t^0\right)|^2\right) \times \widehat{\Lambda}_{0,t}^0,$$

where $\widehat{x}^0(\cdot), P^0(\cdot), \widehat{\Lambda}^0(\cdot)$ are given by

$$\begin{aligned} d\widehat{x}_t^0 &= F\widehat{x}_t^0 dt + P_t^0 H' N^{-1}\left(dy_t - H\widehat{x}_t^0 dt\right), \quad \widehat{x}^0(0) = \xi, \\ \dot{P}_t^0 &= FP_t^0 + P_t^\theta F' - P_t^0 H' N^{-1} H P_t^0 \\ &\quad + GG', \quad P^0(0) = P_0, \\ \widehat{\Lambda}_{0,t}^0 &= \exp\left(\int_0^t \left(H\widehat{x}_s^0\right)' N^{-1} dy_s \right. \\ &\quad \left. - \frac{1}{2}\int_0^t \left(H\widehat{x}_s^0\right)' N^{-1} H\widehat{x}_s^0 ds\right). \end{aligned}$$

1. Computation of $\widehat{L}_{0,t}^{1,1} = \widetilde{\mathrm{E}}[\frac{1}{2}\int_0^t x_s' Q x_s ds | \mathcal{F}_{0,t}^y]$:
A solution of (32), (35) is, (see for example [5, 6])

$$\begin{aligned} \beta^{\theta,1}(x,t) &= \frac{1}{(2\pi)^{\frac{n}{2}}|P_t^\theta|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}|P_t^{\theta,-\frac{1}{2}}\left(x - \widehat{x}_t^\theta\right)|^2\right) \\ &\quad \times \widehat{\Lambda}_{0,t}^\theta \times \exp\left(\frac{\theta}{2}\int_0^t Tr(P_s^\theta Q) ds\right), \end{aligned} \tag{40}$$

where

$$d\widehat{x}_t^\theta = \left(F + \theta P_t^\theta Q\right)\widehat{x}_t^\theta dt + P_t^\theta H' N^{-1}\left(dy_t - H\widehat{x}_t^\theta dt\right), \quad \widehat{x}(0) = \xi, \tag{41}$$

$$\dot{P}_t^\theta = FP_t^\theta + P_t^\theta F' - P_t^\theta\left(H' N^{-1} H - \theta Q\right) P_t^\theta + GG', \quad P^\theta(0) = P_0, \tag{42}$$

$$\widehat{\Lambda}_{0,t}^\theta = \exp\left(\int_0^t \left(H\widehat{x}_s^\theta\right)' N^{-1} dy_s - \frac{1}{2}\int_0^t \left(H\widehat{x}_s^\theta\right)' N^{-1} H\widehat{x}_s^\theta ds\right). \tag{43}$$

In fact, we can show that $\lim_{\theta\to 0} P_t^\theta = P_t^0$, uniformly on compact subsets of $[0,T]$, and $\lim_{\theta\to 0}\widehat{x}_t^\theta = \widehat{x}_t^0$ a.s.

According to Theorem 3.9 we need

$$\frac{d}{d\theta}\frac{\beta_t^{\theta,1}(1)}{\widehat{\Lambda}_{0,t}^0} = \frac{d}{d\theta}\left[\widehat{\Lambda}_{0,t}^\theta\left(\widehat{\Lambda}_{0,t}^0\right)^{-1}\exp\left(\frac{\theta}{2}\int_0^t Tr(P_s^\theta Q)\right)\right]. \tag{44}$$

Let

$$r_t^\theta = \frac{d}{d\theta}\widehat{x}_t^\theta, \qquad \Sigma_t^\theta = \frac{d}{d\theta}P_t^\theta.$$

Then from the differentiability of parameter dependent solutions of stochastic differential equations we know that

$$\begin{aligned} r_t^\theta &= \int_0^t \left(F + \theta P_s^\theta Q\right) r_s^\theta ds + \int_0^t P_s^\theta H' N^{-1}\left(dy_s - H r_s^\theta ds\right) \\ &\quad + \int_0^t \theta \Sigma_s^\theta Q \widehat{x}_s^\theta ds + \int_0^t \Sigma_s^\theta H' N^{-1}\left(dy_s - H\widehat{x}_s^\theta ds\right) \\ &\quad + \int_0^t P_s^\theta Q \widehat{x}_s^\theta ds, \end{aligned} \tag{45}$$

$$\begin{aligned} \Sigma_t^\theta &= \int_0^t F\Sigma_s^\theta ds + \int_0^t \Sigma_s^\theta F' ds - \int_0^t \Sigma_s^\theta \left(H' N^{-1} H\right. \\ &\quad \left. - \theta Q\right) P_s^\theta ds - \int_0^t P_s^\theta \left(H' N^{-1} H - \theta Q\right)\Sigma_s^\theta ds \\ &\quad + \int_0^t P_s^\theta Q P_s^\theta ds, \end{aligned} \tag{46}$$

are continuous in (t,θ) $w.p.1$. Similarly as before we have $\lim_{\theta\to 0} r_t^\theta = r_t^0$ (a.s.), $\lim_{\theta\to 0}\Sigma_t^\theta = \Sigma_t^0$, where

$$\begin{aligned} r_t^0 &= \int_0^t P_s^0 Q \widehat{x}_s^0 ds + \int_0^t \Sigma_s^0 H' N^{-1}\left(dy_s - H\widehat{x}_s^0 ds\right) \\ &\quad + \int_0^t F r_s^0 ds + \int_0^t P_s^0 H' N^{-1}\left(dy_s - H r_s^0 ds\right), \end{aligned} \tag{47}$$

$$\begin{aligned} \Sigma_t^0 &= \int_0^t \left(F - P^0 H' N^{-1} H\right)\Sigma_s^0 ds \\ &\quad + \int_0^t \Sigma_s^0 \left(F - P^0 H' N^{-1} H\right)' ds + \int_0^t P_s^0 Q P_s^0 ds. \end{aligned} \tag{48}$$

Consequently,

$$\begin{aligned} &\lim_{\theta\to 0}\frac{\beta_t^{\theta,1}(1)}{\widehat{\Lambda}_{0,t}^0} \\ &= \lim_{\theta\to 0}\left\{\left(\int_0^t (H r_s^\theta)' N^{-1} dy_s - \int_0^t (H r_s^\theta)' N^{-1} H\widehat{x}_s^\theta ds\right.\right. \\ &\quad \left. + \frac{1}{2}\int_0^t Tr\left(P_s^\theta Q + \theta\Sigma_s^\theta Q\right) ds\right) \\ &\quad \left.\times \widehat{\Lambda}_{0,t}^\theta\left(\widehat{\Lambda}_{0,t}^0\right)^{-1}\exp\left(\frac{\theta}{2}\int_0^t Tr(P_s^\theta Q) ds\right)\right\}. \end{aligned}$$

Finally, $\widehat{L}_{0,t}^{1,1} = \widetilde{\mathrm{E}}[\frac{1}{2}\int_0^t x_s' Q x_s ds | \mathcal{F}_{0,t}^y]$ is given by

$$\widehat{L}_{0,t}^{1,1} = \frac{1}{2}\int_0^t Tr\left(P_s^0 Q\right) ds + \int_0^t \left(H r_s^0\right)' N^{-1}\left(dy_s - H\widehat{x}_s^0 ds\right). \tag{49}$$

2. Computation of $\widehat{L}^{1,2}_{0,t} = \widetilde{\mathrm{E}}[\int_0^t x'_s R dw_s | \mathcal{F}^y_{0,t}]$: A solution of (33), (35) is, (see [5])

$$\beta^{\theta,2}(x,t) = \frac{1}{(2\pi)^{\frac{n}{2}} |P^\theta_t|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} |P^{\theta,-\frac{1}{2}}_t \left(x - \widehat{x}^\theta_t\right)|^2 \right)$$
$$\times \widehat{\Lambda}^\theta_{0,t} \times \exp\left(\frac{\theta}{2} \int_0^t Tr(P^\theta_s RR') ds \right), \quad (50)$$

where

$$d\widehat{x}^\theta_t = \left(F + \theta P^\theta_t RR' + \theta GR'\right) \widehat{x}^\theta_t dt$$
$$+ P^\theta_t H' N^{-1} \left(dy_t - H\widehat{x}^\theta_t dt\right), \ \widehat{x}(0) = \xi, \quad (51)$$
$$\dot{P}^\theta_t = \left(F + \theta GR'\right) P^\theta_t + P^\theta_t \left(F + \theta GR'\right)'$$
$$- P^\theta_t \left(H'N^{-1}H - \theta RR'\right) P^\theta_t + GG', \ P^\theta(0) = P_0, \quad (52)$$
$$\widehat{\Lambda}^\theta_{0,t} = \exp\left(\int_0^t \left(H\widehat{x}^\theta_s\right)' N^{-1} dy_s \right. \quad (53)$$
$$\left. - \frac{1}{2} \int_0^t \left(H\widehat{x}^\theta_s\right)' N^{-1} H\widehat{x}^\theta_s ds \right). \quad (54)$$

By Theorem 3.9 we need

$$\frac{d}{d\theta} \frac{\beta^{\theta,2}_t(1)}{\widehat{\Lambda}^0_{0,t}} = \frac{d}{d\theta} \left[\widehat{\Lambda}^\theta_{0,t} \left(\widehat{\Lambda}^0_{0,t}\right)^{-1} \exp\left(\frac{\theta}{2} \int_0^t Tr(P^\theta_s RR') \right) \right]. \quad (55)$$

Computing $\lim_{\theta\to 0} r^\theta_t = \lim_{\theta \to 0} \frac{d}{d\theta} \widehat{x}^\theta_t = r^0_t, \lim_{\theta\to 0} \Sigma^\theta_t = \lim_{\theta\to 0} \frac{d}{d\theta} P^\theta_t = P^0_t$, similarly as before, we have

$$r^0_t = \int_0^t \left(P^0_s RR' + GR'\right) \widehat{x}^0_s ds$$
$$+ \int_0^t \Sigma^0_s H' N^{-1} \left(dy_s - H\widehat{x}^0_s ds\right)$$
$$+ \int_0^t F r^0_s ds + \int_0^t P^0_s H' N^{-1} \left(dy_s - H r^0_s ds\right), \quad (56)$$
$$\Sigma^0_t = \int_0^t \left(F - P^0 H' N^{-1} H\right) \Sigma^0_s ds$$
$$+ \int_0^t \Sigma^0_s \left(F - P^0 H' N^{-1} H\right)' ds + \int_0^t P^0_s RR' P^0_s ds$$
$$+ \int_0^t GR' P^0_s ds + \int_0^t P^0_s RG' ds. \quad (57)$$

Hence

$$\lim_{\theta\to 0} \frac{d}{d\theta} \frac{\beta^{\theta,2}_t(1)}{\widehat{\Lambda}^0_{0,t}} = \frac{1}{2} \int_0^t Tr\left(P^0_s RR'\right) ds$$
$$+ \int_0^t \left(H r^0_s\right)' N^{-1} \left(dy_s - H\widehat{x}^0_s\right) ds, \quad (58)$$

Finally, $\widehat{L}^{1,2}_{0,t} = \widetilde{\mathrm{E}}[\int_0^t x'_s R dw_s | \mathcal{F}^y_{0,t}]$ is given by

$$\widehat{L}^{1,2}_{0,t} = \frac{1}{2} \int_0^t Tr\left(P^0_s RR'\right) ds$$
$$+ \int_0^t \left(H r^0_s\right)' N^{-1} \left(dy_s - H\widehat{x}^0_s ds\right). \quad (59)$$

3. Computation of $\widehat{L}^{1,3}_{0,t} = \widetilde{\mathrm{E}}[\int_0^t x'_s S db_s | \mathcal{F}^y_{0,t}]$: Following the same approach we deduce that $\widehat{L}^{1,3}_{0,t} = \widetilde{\mathrm{E}}[\int_0^t x'_s S db_s | \mathcal{F}^y_{0,t}]$ is given by

$$\widehat{L}^{1,3}_{0,t} = \int_0^t \left(H r^0_s\right)' N^{-1} \left(dy_s - H\widehat{x}^0_s ds\right)$$
$$+ \int_0^t \left(N^{-\frac{1}{2}} S' \widehat{x}^0_s\right)' N^{-1} \left(dy_s - N^{-\frac{1}{2},'} S' \widehat{x}^0_s ds\right), \quad (60)$$

where

$$r^0_t = \int_0^t \left(\Sigma^0_s H' N^{-1} + P^0_s (SN^{-\frac{1}{2}})' N^{-1}\right) \left(dy_s - H\widehat{x}^0_s ds\right)$$
$$+ \int_0^t P^0_s H' N^{-1} \left(dy_s - (SN^{-\frac{1}{2}})' \widehat{x}^0_s ds\right)$$
$$+ \int_0^t F r^0_s ds + \int_0^t P^0_s H' N^{-1} \left(dy_s - H r^0_s ds\right), \quad (61)$$
$$\Sigma^0_t = \int_0^t \left(F - P^0 H' N^{-1} H\right) \Sigma^0_s ds$$
$$+ \int_0^t \Sigma^0_s \left(F - P^0 H' N^{-1} H\right)' ds$$
$$- \int_0^t P^0_s \left((SN^{-\frac{1}{2}}) N^{-1} H + (N^{-1}H)'(SN^{-\frac{1}{2}})'\right) P^0_s ds. \quad (62)$$

References

[1] S. Marcus, "Algebraic structure and finite dimensional nonlinear estimation," *SIAM Journal on Mathematical Analysis*, vol. 9, no. 2, pp. 312–327, 1978.

[2] D. Ocone, *Topics in Nonlinear Filtering Theory.* PhD thesis, M.I.T., Massachusetts, 1980.

[3] R. Elliott and V. Krishnamurthy, "Exact finite dimensional filters for maximum likelihood parameter estimation of continuous-time linear Gaussian systems," *IEEE Transactions on Information Theory (submitted)*, 1996.

[4] R. S. Liptser and A. N. Shiryayev, *Statistics of Random Processes, Vol.1.* New York: Springer-Verlag, 1977.

[5] C. Charalambous and J. Hibey, "Minimum principle for partially observable nonlinear risk-sensitive control problems using measure-valued decompositions," *Stochastics and Stochastics Reports*, vol. 57, 1996.

[6] C. Charalambous and R. Elliott, "Certain nonlinear stochastic optimal control problems with explicit control laws equivalent to LEQG/LQG problems (to appear in)," *IEEE Transactions on Automatic Control*, vol. 42, no. 4, 1997.

[7] J. Blagovescenskii and M. Freidlin, "Some properties of diffusion processes depending on a parameter," *Soviet Mathematics*, vol. 2, pp. 633–636, 1961.