INTERNATIONAL JOURNAL OF CONTROL, 2018
VOL. 91, NO. 3, 678–687
http://dx.doi.org/10.1080/00207179.2017.1288300

Taylor & Francis
Taylor & Francis Group

Direct method for Yau filtering system with nonlinear observations

Ji Shi, Zhiyu Yang and Stephen S. -T. Yau

Department of Mathematical Sciences, Tsinghua University, Beijing, China

ABSTRACT
For all known finite dimensional filters, one always assumes that the observation terms be degree one polynomials. However, in practice, the observation terms may be nonlinear, e.g. tracking problems. In this paper, we consider the Yau filtering system ($\frac{\partial f_j}{\partial x_i} - \frac{\partial f_i}{\partial x_j} = c_{ij}$ is constant for all i, j) with nonlinear observation terms and arbitrary initial condition. The novelty of the paper lies in (i) the real time computation of the solution of the Duncan-Mortensen-Zakai (DMZ) equation is reduced to the computation of Kolmogorov equation. Based on Gaussian approximation of the initial condition, the Kolmogorov equation can be solved in terms of ordinary differential equations; (ii) For a given probability density function, we give a new and original approach to do Gaussian approximation which is very effective and simple. The direct method developed here can be easily implemented in a real time and memoryless way. Besides, we do not need the controllability and observability assumption. Compared to the extended Kalman filter, our method is much stable and has theoretical proof. The numerical experiments show that the proposed Gaussian approximation method is very effective and our method can track the states very well.

ARTICLE HISTORY
Received 4 June 2016
Accepted 24 January 2017

KEYWORDS
Duncan-Mortensen-Zakai (DMZ) equation; Kolmogorov equation; nonlinear drift; nonlinear observation; Gaussian approximation

1. Introduction

In the early 1960s, Kalman and Bucy (1961) proposed the continuous version of Kalman filter which has been widely used in various fields of industry. However, Kalman filter is restricted to linear systems with Gaussian initial distribution. Actually, most of the filtering systems in practice is nonlinear with non-Gaussian initial distribution. The study of nonlinear filtering (NLF) is aimed at determining the conditional density $\rho(t, x)$ of the state $x(t)$ given the observation history $\{y(s): 0 \leq s \leq t\}$. In the late 1960s, Duncan (1967), Mortensen (1966), Zakai (1969) independently derived the so-called Duncan-Mortensen-Zakai (DMZ) equation for the NLF problem. Then $\rho(t, x)$ can be obtained by normalising the solution $\sigma(t, x)$ of the DMZ equation. Since the DMZ equation is a stochastic partial differential equation (PDE), there is no easy way to solve it.

Motivated by the Wei–Norman approach Wei and Norman (1964) of using Lie algebraic method to solve the linear time varying differential equation, Brockett and Clark (1980), Brockett (1981), and Mitter (1979) proposed the idea of using estimation algebra to construct finite dimensional nonlinear filter. However, in the Wei–Norman approach, one has to know explicitly the basis of the estimation algebra in order to reduce the DMZ equation to a finite system of ordinary differential equations (ODE)s, a Kolmogorov equation and first-order linear PDEs. In Chiou and Yau (1994), Tam, Wong, and Yau (1990), Yau (1994) and Yau and Hu (2005), all finite dimensional estimation algebra of maximal rank had been completely classified. Particularly, for a NLF system satisfying $\frac{\partial f_j}{\partial x_i} - \frac{\partial f_i}{\partial x_j} = c_{ij}$ is constant for all i, j is called the Yau filtering system in Chen (1994), which contains the Kalman filter and Benés (1981) filter as its special cases. However, one knows explicitly the basis of finite dimensional estimation algebra only for a few cases and one assumes that the linear system is controllable and observable.

In fact, for all known finite dimensional filters, one always needs the condition that the observation terms $h_i(x), 1 \leq i \leq m$ are degree one polynomials. However, the observation terms may be nonlinear in many situation. Yau and Hu (2001, 2005) first purposed the direct method to solve the DMZ equation. Later Yau and Lai (2003) gave the solution of Kolmogorov equation under certain conditions with Gaussian initial distribution in terms of ODEs. In Yau and Yau (2004a, 2004b), the Yau filtering system with linear observation and linear filtering system with nonlinear observation terms are considered, respectively. The direct method has several advantages. First, the method is easy to implement, and the derivation no longer needs controllability and

CONTACT Stephen S. -T. Yau yau@uic.edu
© 2017 Informa UK Limited, trading as Taylor & Francis Group

observability. Second, compared to the wildly used extended Kalman filter (EKF), the direct method is much stable and has theoretic convergence proof. Moreover, the necessity of integrating n first-order linear PDEs in the estimation algebra method is eliminated. Recently, in Luo and Yau (2013a, 2013b), a real-time algorithm for a general class of NLF problems was developed. Their method directly computes the Kolmogorov equation in advance and uses hermite orthogonal polynomials to approximate the initial condition at every time step.

In this paper, compared to the NLF systems considered in Yau and Yau (2004a, 2004b), we consider a more general situation, i.e. both the drift and the observations can be nonlinear . It is well known that any non-Gaussian density function can be well approximated by finite linear combination of Gaussian distributions and the most wildly used technique is expectation maximisation (EM) algorithm Dempster (1997). However, typically the EM algorithm uses a set of sample points to determine the Gaussian mixture parameters, which is not the case in this paper. A new and original way to do Gaussian approximation is proposed in this paper which is very effective as verified by the numerical experiments. Besides, the method is very simple and fast. Based on our Gaussian approximation algorithm, the conditional density function $\sigma(t, x)$ is explicitly given via solutions of ODEs.

The paper is organised as follows. In Section 2, we shall recall the basic filtering problem and some preliminary results. In Section 3, we first reduce the solution of the robust DMZ equation to the solution of Kolmogorov equation, followed by the Gaussian approximation algorithm. Numerical experiments are given in Section 4. Finally, the conclusions are presented in Section 5.

2. Basic concepts and preliminary results

In this paper, we consider the following signal observation model

$$\begin{cases} dx(t) = f(x(t))dt + g(x(t))dv(t), & x(0) = x_0, \\ dy(t) = h(x(t))dt + dw(t) & y(0) = 0, \end{cases} \tag{1}$$

in which $x(t)$, $v(t)$, $y(t)$ and $w(t)$ are, respectively, R^n, R^p, R^m and R^m valued processes and $v(t)$ and $w(t)$ have components that are independent, standard Brownian processes. We further assume that $n = p$, f and h are C^∞ smooth vector-valued functions, and that g is an orthogonal matrix function.

Let $\sigma(t, x)$ denote the unnormalised conditional probability density function of the state given the observation $\{y(s): 0 \le s \le t\}$, which satisfies the following DMZ equation:

$$\begin{cases} d\sigma(t, x) = L_0\sigma(t, x)dt + \Sigma_{i=1}^n L_i\sigma(t, x)dy_i(t), \\ \sigma(0, x) = \sigma_0, \end{cases} \tag{2}$$

where

$$L_0 = \frac{1}{2}\sum_{i=1}^n \frac{\partial^2}{\partial x_i^2} - \sum_{i=1}^n f_i\frac{\partial}{\partial x_i} - \sum_{i=1}^n \frac{\partial f_i}{\partial x_i} - \frac{1}{2}\sum_{i=1}^m h_i^2 \tag{3}$$

and for $i = 1, \dots, m, L_i$ is the zero degree differential operator of multiplication by h_i. σ_0 is the probability density of the initial value x_0. In real application, we are interested in considering robust state estimator from observed sample paths with some properties of robustness. In Davis (1980), Davis considered this problem and proposed some robust algorithms. In our case, his basic idea reduced to define a new unnormalised density

$$u(t, x) = \exp\left(-\sum_{i=1}^m h_i(x)y_i(t)\right)\sigma(t, x), \tag{4}$$

then $u(t, x)$ satisfies the following equation:

$$\begin{cases} \frac{\partial u}{\partial t}(t, x) = L_0 u(t, x) + \sum_{i=1}^m y_i(t)[L_0, L_i]u(t, x) \\ \qquad + \frac{1}{2}\sum_{i,j=1}^m y_i(t)y_j(t)[[L_0, L_i], L_j]u(t, x) \\ u(0, x) = \sigma_0(x), \end{cases} \tag{5}$$

where $[\cdot, \cdot]$ is the Lie bracket. Equation (5) is called robust DMZ equation and we can rewrite it in the following equivalent form

$$\begin{cases} \frac{\partial u}{\partial t}(t, x) = \frac{1}{2}\sum_{i=1}^n \frac{\partial^2 u}{\partial x_i^2}(t, x) + \sum_{i=1}^n\{-f_i(x) \\ \qquad + \sum_{j=1}^m y_j(t)\frac{\partial h_j}{\partial x_i}(x)\}\frac{\partial u}{\partial x_i}(t, x) \\ \qquad -\{\sum_{i=1}^m\sum_{j=1}^n y_i(t)f_j(x)\frac{\partial h_i}{\partial x_j}(x) \\ \qquad -\frac{1}{2}\sum_{i=1}^m\sum_{j=1}^n y_i(t)\frac{\partial^2 h_i}{\partial x_j^2}(x) \\ \qquad -\frac{1}{2}\sum_{k=1}^n\sum_{i,j=1}^m y_i(t)y_j(t)\frac{\partial h_i}{\partial x_k}(x)\frac{\partial h_j}{\partial x_k}(x) \\ \qquad + \sum_{i=1}^n \frac{\partial f_i}{\partial x_i}(x) + \frac{1}{2}\sum_{i=1}^m h_i^2(x)\}u(t, x), \\ u(0, x) = \sigma_0(x), \end{cases} \tag{6}$$

In 1990, Yau (1990) first studied the filtering system (1) with the following conditions:

$$C_1') \quad \frac{\partial f_j}{\partial x_i} - \frac{\partial f_i}{\partial x_j} = c_{ij},\ 1 \le i, j \le n,$$

where c_{ij} is constant. The filtering system with condition C_1') is called the Yau filtering system in Chen (1994). Yau filtering systems include the Kalman–Bucy filtering systems and Benés filtering systems as special cases.

Theorem 2.1 (*Theorem* 1, *Yau* (1994)): *The condition* (C_1') *holds if and only if*

$$(f_1, \dots, f_n) = (l_1, \dots, l_n) + \left(\frac{\partial F}{\partial x_1}, \dots, \frac{\partial F}{\partial x_n}\right),$$

where $l_1, \dots, l_n$ *are polynomials of degree one and* F *is a* C^∞ *function.*

Define

$$\eta(x) = \sum_{i=1}^{n} f_i^2(x) + \sum_{i=1}^{n} \frac{\partial f_i}{\partial x_i}(x) + \sum_{i=1}^{m} h_i^2(x). \quad (7)$$

In this paper, we consider system (1) with the following three conditions:

(C_1) $f_i(x) = l_i(x) + \frac{\partial F}{\partial x_i}(x)$, $1 \le i \le n$;
(C_2) $\sum_{i=1}^{m} h_i^2(x) = \sum_{i,j=1}^{n} q_{ij}x_i x_j + \sum_{i=1}^{n} q_i x_i + q_0$;
(C_3) $\eta(x) = \sum_{i,j=1}^{n} \eta_{ij}x_i x_j + \sum_{i=1}^{n} \eta_i x_i + \eta_0$;

where $l_i(x) = \sum_{j=1}^{n} d_{ij}x_j + d_i$ and $d_{ij}, d_i, q_{ij} = q_{ji}, q_i, q_0, \eta_{ij}, \eta_i, \eta_0$, $1 \le i, j \le n$ are constants. We remark that nonlinear observation h_i's are allowed in this paper and the Kalman–Bucy filtering systems satisfies (C_3) and so does in Benés (1981).

3. Solution of the robust DMZ equation

In this section, we first reduce the robust DMZ equation to a Kolmogorov equation. Then the Kolmogorov equation can be solved in terms of ODEs based on an original Gaussian approximation method.

Let $P_N = \{0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_N = T\}$ be a partition of $[0, T]$. For each time interval $\tau_{k-1} \le t \le \tau_k$, let $u_k(t, x)$ be the solution of the following PDE (8), which is obtained from (6) by freezing the observation term $y(t)$ to $y(\tau_{k-1})$

$$\begin{cases} \frac{\partial u_k}{\partial t}(t,x) = & \frac{1}{2}\sum_{i=1}^{n}\frac{\partial^2 u_k}{\partial x_i^2}(t,x) + \sum_{i=1}^{n}\{-f_i(x) \\ & + \sum_{j=1}^{m} y_j(\tau_{k-1})\frac{\partial h_j}{\partial x_i}(x)\}\frac{\partial u_k}{\partial x_i}(t,x) \\ & -\{\sum_{i=1}^{m}\sum_{j=1}^{n} y_i(\tau_{k-1}) f_j(x)\frac{\partial h_i}{\partial x_j}(x) \\ & -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} y_i(\tau_{k-1})\frac{\partial^2 h_i}{\partial x_j^2}(x) \\ & -\frac{1}{2}\sum_{k=1}^{n}\sum_{i,j=1}^{m} y_i(\tau_{k-1})y_j(\tau_{k-1}) \\ & \quad \frac{\partial h_i}{\partial x_k}(x)\frac{\partial h_j}{\partial x_k}(x) \\ & +\sum_{i=1}^{n}\frac{\partial f_i}{\partial x_i}(x) + \frac{1}{2}\sum_{i=1}^{m} h_i^2(x)\}u_k(t,x), \\ u_k(\tau_{k-1}, x) = u_{k-1}(\tau_{k-1}, x). \end{cases} \quad (8)$$

Define the norm of the partition P_N by $|P_N| = \sup_{1 \le i \le N}(\tau_i - \tau_{i-1})$, it has been proved in Yau and Yau (2000, 2005) that in both pointwise sense and L^2-sense

$$u(\tau, x) = \lim_{|P_N| \to 0} u_i(\tau, x). \quad (9)$$

3.1 Reduction of the robust DMZ equation to Kolmogorov equation

Our reduction of the robust DMZ equation to Kolmogorov equation is based on the following important proposition.

Proposition 3.1 (*Proposition* 3.1, *Yau and Yau* (2004*b*)): *For each* $\tau_{k-1} \le t \le \tau_k$, $1 \le k \le n$, $\tilde{u}_k(t, x)$ *satisfies the following parabolic equation:*

$$\frac{\partial \tilde{u}_k}{\partial t}(t,x) = \frac{1}{2}\Delta\tilde{u}_k(t,x) - \sum_{i=1}^{n} f_i(x)\frac{\partial \tilde{u}_k}{\partial x_i}(t,x) - \left(\sum_{i=1}^{n}\frac{\partial f_i}{\partial x_i}(x) + \frac{1}{2}\sum_{i=1}^{m} h_i^2\right)\tilde{u}_k(t,x) \quad (10)$$

for $\tau_{k-1} \le t \le \tau_k$, *if and only if*

$$u_k(t,x) = \exp\left(-\sum_{i=1}^{m} y_i(\tau_{k-1})h_i(x)\right)\tilde{u}_k(t,x) \quad (11)$$

satisfies (8).

The initial condition for (10) on $\tau_{k-1} \le t \le \tau_k$ is

$$\tilde{u}_k(\tau_{k-1}, x) = \begin{cases} \sigma_0(x)\exp(-\sum_{j=1}^{m} y_j(0)h_j(x)) = \sigma_0(x), & k = 1 \\ \exp\{\sum_{j=1}^{m}(y_j(\tau_{k-1}) - y_j(\tau_{k-2}))h_j(x)\} & \\ \tilde{u}_{k-1}(\tau_{k-1}, x), & k \ge 2 \end{cases} \quad (12)$$

From (4), (9) and (11), we have

$$\sigma(t,x) = \lim_{|P_k| \to 0} \tilde{u}_k(\tau_k, x). \quad (13)$$

Hence, to compute the unnormalised density $\sigma(t, x)$, we only need to find the solution $\tilde{u}_k(t, x)$ of the Kolmogorov Equation (10).

Lemma 3.1: *For each* k, $\tau_{k-1} < t < \tau_k$, (10) *is equivalent to the following equation*

$$\frac{\partial \tilde{u}_k}{\partial t}(t,x) = \frac{1}{2}\Delta\tilde{u}_k(t,x) + \sum_{i=1}^{n}\theta_i(x)\frac{\partial \tilde{u}_k}{\partial x_i}(t,x) + \theta(x)\tilde{u}_k(t,x), \quad (14)$$

where $\theta_i(x) = -f_i(x)$ *and* $\theta(x) = \frac{1}{2}(\sum_{i=1}^{n}\theta_i^2(x) + \sum_{i=1}^{n}\frac{\partial \theta_i}{\partial x_i}(x) - \eta(x))$.

Proof: Recall from (7) that

$$\eta(x)=\sum_{i=1}^{n} f_i^2(x)+\sum_{i=1}^{n}\frac{\partial f_i}{\partial x_i}(x)+\sum_{i=1}^{m} h_i^2(x).$$

Define $\theta_i(x)=-f_i(x)$, the coefficient of $\tilde{u}_k(t,x)$ in (10) is given by

$$\begin{aligned}
&-\left(\sum_{i=1}^{n}\frac{\partial f_i}{\partial x_i}(x)+\frac{1}{2}\sum_{i=1}^{m} h_i^2\right)\\
&=\sum_{i=1}^{n}\frac{\partial \theta_i}{\partial x_i}(x)-\frac{1}{2}(\eta(x)-\sum_{i=1}^{n} f_i^2(x)-\sum_{i=1}^{n}\frac{\partial f_i}{\partial x_i}(x))\\
&=\frac{1}{2}\sum_{i=1}^{n}\theta_i^2(x)-\frac{1}{2}\eta(x)+\sum_{i=1}^{n}\frac{\partial \theta_i}{\partial x_i}(x)+\frac{1}{2}\sum_{i=1}^{n}\frac{\partial f_i}{\partial x_i}(x)\\
&=\frac{1}{2}\left(\sum_{i=1}^{n}\theta_i^2(x)+\sum_{i=1}^{n}\frac{\partial \theta_i}{\partial x_i}(x)-\eta(x)\right)=:\theta(x).
\end{aligned}$$

■

In order to solve the Kolmogorov Equation (14) in terms of ODEs, we first introduce a new transformation in the following theorem.

Theorem 3.1: *For each k, $\tau_{k-1}\le t\le\tau_k$, suppose $\tilde{u}_k(t,x)$ is a solution of (14) and let*

$$\hat{u}_k(t,x)=e^{\Lambda(x)}\tilde{u}_k(t,x), \tag{15}$$

then $\hat{u}_k(t,x)$ is the solution of the following Kolmogorov equation

$$\begin{cases}\frac{\partial \hat{u}_k}{\partial t}(t,x) & =\frac{1}{2}\Delta\hat{u}_k(t,x)-\sum_{i=1}^{n} H_i(x)\frac{\partial \hat{u}_k}{\partial x_i}(t,x)\\ & \quad -P(x)\hat{u}_k(t,x)\\ \hat{u}_k(\tau_{k-1},x)=e^{\Lambda(x)}\cdot\tilde{u}_k(\tau_{k-1},x)\end{cases} \tag{16}$$

if we can choose $H_i(x)$, $P(x)$ and $\Lambda(x)$ such that the following equations:

$$-\frac{\partial \Lambda}{\partial x_i}(x)+H_i(x)+\theta_i(x)\equiv 0,\quad 1\le i\le n, \tag{17}$$

$$-\frac{1}{2}\eta(x)+\frac{1}{2}\sum_{i=1}^{n} H_i^2(x)-\frac{1}{2}\sum_{i=1}^{n}\frac{\partial H_i}{\partial x_i}(x)+P(x)\equiv 0 \tag{18}$$

hold.

Proof: Differentiating $\hat{u}_k(t,x)$ with respect to t and x, we have the following equations:

$$\frac{\partial \hat{u}_k(t,x)}{\partial t}=e^{\Lambda(x)}\frac{\partial \tilde{u}_k(t,x)}{\partial t}, \tag{19}$$

$$\frac{\partial \hat{u}_k(t,x)}{\partial x_i}=e^{\Lambda(x)}\left(\frac{\partial \Lambda(x)}{\partial x_i}\cdot\tilde{u}_k+\frac{\partial \tilde{u}_k(t,x)}{\partial x_i}\right), \tag{20}$$

$$\begin{aligned}\frac{\partial^2 \hat{u}_k(t,x)}{\partial x_i^2}&=e^{\Lambda(x)}\left\{\left(\frac{\partial^2 \Lambda(x)}{\partial x_i^2}+\left(\frac{\partial \Lambda(x)}{\partial x_i}\right)^2\right)\tilde{u}_k(t,x)\right.\\ &\left.\quad+2\frac{\partial \Lambda(x)}{\partial x_i}\cdot\frac{\partial \tilde{u}_k(t,x)}{\partial x_i}+\frac{\partial^2 \tilde{u}_k(t,x)}{\partial x_i^2}\right\}.\end{aligned} \tag{21}$$

Putting (19)–(21) into (16), we have

$$\begin{aligned}&\frac{\partial \hat{u}_k(t,x)}{\partial t}\\
&=e^{\Lambda(x)}\cdot\left\{\frac{1}{2}\Delta\tilde{u}_k(t,x)+\sum_{i=1}^{n}\left(\frac{\partial \Lambda(x)}{\partial x_i}-H_i\right)\cdot\frac{\partial \tilde{u}_k(t,x)}{\partial x_i}\right.\\
&\quad+\left(\frac{1}{2}\Delta\Lambda(x)+\frac{1}{2}\sum_{i=1}^{n}\left(\frac{\partial \Lambda(x)}{\partial x_i}\right)^2\right.\\
&\quad\left.\left.-\sum_{i=1}^{n} H_i\cdot\frac{\partial \Lambda(x)}{\partial x_i}-P(x)\right)\cdot\tilde{u}_k(t,x)\right\}\end{aligned} \tag{22}$$

Using (14), (19) and (22), we have

$$\begin{aligned}&\sum_{i=1}^{n}\theta_i(x)\frac{\partial \tilde{u}_k}{\partial x_i}(t,x)+\theta(x)\tilde{u}_k(t,x)\\
&=\sum_{i=1}^{n}\left(\frac{\partial \Lambda(x)}{\partial x_i}-H_i\right)\cdot\frac{\partial \tilde{u}_k(t,x)}{\partial x_i}\\
&\quad+\left(\frac{1}{2}\Delta\Lambda(x)+\frac{1}{2}\sum_{i=1}^{n}\left(\frac{\partial \Lambda(x)}{\partial x_i}\right)^2\right.\\
&\quad\left.-\sum_{i=1}^{n} H_i\cdot\frac{\partial \Lambda(x)}{\partial x_i}-P(x)\right)\cdot\tilde{u}_k(t,x).\end{aligned} \tag{23}$$

Observing the coefficients of $\frac{\partial \tilde{u}_k(t,x)}{\partial x_i}$ and $\tilde{u}_k(t,x)$, we have

$$\begin{aligned}\theta_i(x)&=\frac{\partial \Lambda(x)}{\partial x_i}-H_i,\\
\theta(x)&=\frac{1}{2}\Delta\Lambda(x)+\frac{1}{2}\sum_{i=1}^{n}\left(\frac{\partial \Lambda(x)}{\partial x_i}\right)^2\\
&\quad-\sum_{i=1}^{n} H_i\cdot\frac{\partial \Lambda(x)}{\partial x_i}-P(x).\end{aligned}$$

Recall that $\theta(x) = \frac{1}{2}(\sum_{i=1}^n \theta_i^2(x) + \sum_{i=1}^n \frac{\partial \theta_i}{\partial x_i}(x) - \eta(x))$, then (17) and (18) follows. ■

Noting the special structure of the drift f, i.e. condition (C_1), we have a special choice of $\Lambda(x)$ in (15).

Theorem 3.2: *Consider the NLF system (1) with conditions $(C_1)-(C_3)$. Then for each k, $\tau_{k-1} \le t \le \tau_k$, the solution $\tilde{u}_k(t,x)$ for (10) is reduced to the solution $\hat{u}_k(t,x)$ for the following Kolmogorov equation*

$$\begin{cases} \frac{\partial \hat{u}_k}{\partial t}(t,x) &= \frac{1}{2}\Delta \hat{u}_k(t,x) - \sum_{i=1}^n H_i(x)\frac{\partial \hat{u}_k}{\partial x_i}(t,x) \\ & \quad - P(x)\hat{u}_k(t,x) \\ \hat{u}_k(\tau_{k-1},x) = e^{G(x)-F(x)}\tilde{u}_k(\tau_{k-1},x) \end{cases} \tag{24}$$

where

$$\hat{u}_k(t,x) = e^{G(x)-F(x)}\tilde{u}_k(t,x), \tag{25}$$

if we can choose $H(x)$, $G(x)$ and $P(x)$ such that

$$\frac{1}{2}\sum_{i=1}^n H_i^2(x) - \frac{1}{2}\sum_{i=1}^n \frac{\partial H_i}{\partial x_i}(x) - \frac{1}{2}\eta(x) + P(x) \equiv 0 \tag{26}$$

where

$$H_i(x) - \frac{\partial G}{\partial x_i}(x) = l_i(x). \tag{27}$$

Proof: Here we use Theorem 3.1 with $\Lambda(x) = G(x) - F(x)$, where $F(x)$ is the one in Theorem 2.1. Then

$$\frac{\partial \Lambda}{\partial x_i}(x) = \frac{\partial G}{\partial x_i}(x) - \frac{\partial F}{\partial x_i}(x). \tag{28}$$

Putting (28) into (17) and note that $\theta_i(x) = -f_i(x)$, we have $H_i(x) - \frac{\partial G}{\partial x_i}(x) = l_i(x)$. ■

In the following corollary, we choose $H(x), G(x), P(x)$ which satisfy the conditions in Theorem 3.2 such that the coefficients of $\hat{u}_k(t,x)$ and $\frac{\partial \hat{u}_k}{\partial x_i}(t,x)$ are degree one polynomial and degree two polynomial, respectively. By this way, we can solve the transformed Kolmogorov Equation (24) in terms of ODEs if the initial condition is Gaussian.

Corollary 3.1: *Choose $G(x) \equiv 0$, $P(x) = \frac{1}{2}\eta(x) - \frac{1}{2}\sum_{i=1}^n l_i^2(x) + \frac{1}{2}\sum_{i=1}^n \frac{\partial l_i}{\partial x_i}(x)$ and $H_i(x) = l_i(x)$, $1 \le i \le n$, then (26) and (27) hold in this case. For each k, $\tau_{k-1} \le t \le \tau_k$, the corresponding Kolmogorov Equation (24) is given by*

$$\begin{aligned} \frac{\partial \hat{u}_k}{\partial t}(t,x) &= \frac{1}{2}\Delta \hat{u}_k(t,x) - \sum_{i=1}^n l_i(x)\frac{\partial \hat{u}_k}{\partial x_i}(t,x) \\ &\quad + \frac{1}{2}(\sum_{i=1}^n l_i^2(x) - \sum_{i=1}^n \frac{\partial l_i}{\partial x_i}(x) \\ &\quad - \eta(x))\hat{u}_k(t,x) \end{aligned} \tag{29}$$

with

$$\hat{u}_k(\tau_{k-1},x) = \begin{cases} e^{-F(x)}\sigma_0(x), & k=1, \\ \exp\{\sum_{j=1}^m (y_j(\tau_{k-1}) - y_j(\tau_{k-2}))h_j(x)\}\cdot \\ \hat{u}_{k-1}(\tau_{k-1},x), & k \ge 2. \end{cases} \tag{30}$$

Then for $\tau_{k-1} \le t \le \tau_k$,

$$\tilde{u}_k(t,x) = e^{F(x)}\hat{u}_k(t,x). \tag{31}$$

Proof: Recall from (12) that

$$\tilde{u}_k(\tau_{k-1},x) = \begin{cases} \sigma_0(x), & k=1, \\ \exp\{\sum_{j=1}^m (y_j(\tau_{k-1}) - y_j(\tau_{k-2}))h_j(x)\} \\ \tilde{u}_{k-1}(\tau_{k-1},x), & k \ge 2. \end{cases}$$

Then, for $k=1$, $\hat{u}_1(0,x) = e^{-F(x)}\sigma_0(x)$. For $k \ge 2$,

$$\begin{aligned} \hat{u}_k(\tau_{k-1},x) &= e^{-F(x)}\tilde{u}_k(\tau_{k-1},x) \\ &= e^{-F(x)}\exp\left\{\sum_{j=1}^m (y_j(\tau_{k-1}) - y_j(\tau_{k-2}))h_j(x)\right\}\tilde{u}_{k-1}(\tau_{k-1},x) \\ &= \exp\left\{\sum_{j=1}^m (y_j(\tau_{k-1}) - y_j(\tau_{k-2}))h_j(x)\right\}\hat{u}_{k-1}(\tau_{k-1},x) \end{aligned}$$

■

Suppose $\hat{u}_k(\tau_{k-1},x)$ is well approximated by a sum of finite number of Gaussian distributions, it follows that a well approximated solution of (29) is obtained by linear combination of solutions of (29) with Gaussian initial condition since (29) is a linear PDE. The following theorem give the solution of (29) with Gaussian initial distribution in terms of ODEs .

Theorem 3.3 (Theorem 3.2, Yau & Lai, 2003)**:** *Consider the following Kolmogorov equation with Gaussian initial* *condition*

$$\begin{cases} \frac{\partial \hat{u}}{\partial t}(t,x) = \frac{1}{2}\Delta\hat{u} - \sum_{i=1}^{n} l_i(x)\frac{\partial \hat{u}}{\partial x_i}(t,x) \\ \quad +\frac{1}{2}\big(\sum_{i=1}^{n} l_i^2(x) - \sum_{i=1}^{n}\frac{\partial l_i(x)}{\partial x_i} \\ \quad -\eta(x)\big)\hat{u}(t,x) \\ \hat{u}(t_0,x) = e^{x^T A(t_0)x + B^T(t_0)x + C(t_0)} \end{cases} \tag{32}$$

where $A(t_0) = (A_{ij}(t_0))$ *is an* $n \times n$ *symmetric matrix,* $B^T(t_0) = (B_1(t_0), \dots, B_n(t_0))$, $x^T = (x_1, \dots, x_n)$ *are row vectors and* $C(t_0)$ *is a scalar,* $t_0 \geq 0$.

Let

$$\begin{aligned} q(x) &= \frac{1}{2}\left(\sum_{i=1}^{n} l_i^2(x) - \sum_{i=1}^{n}\frac{\partial l_i}{\partial x_i}(x) - \eta(x)\right) \\ &= x^T Q x + p^T x + r \end{aligned} \tag{33}$$

where $l_i(x) = \sum_{j=1}^{n} d_{ij}x_j + d_i$, $Q = (q_{ij})$ *is an* $n \times n$ *symmetric matrix,* $p^T = (p_1, \dots, p_n)$ *is a row vector and* r *is a scalar. Then the solution of* (32) *is of the following form*

$$\hat{u}(t,x) = e^{x^T A x + B^T x + C} \tag{34}$$

where $A(t) = (A_{ij}(t))$ *is an* $n \times n$ *symmetric matrix valued function of* t, $B^T(t) = (B_1(t), \dots, B_n(t))$ *is a row vector valued function of* t, *and* $C(t)$ *is a scalar function of* t. *Moreover,* $A(t)$, $B(t)$ *and* $C(t)$ *satisfy the following system of nonlinear ODEs:*

$$\frac{dA(t)}{dt} = 2A^2(t) - [A(t)D + D^T A(t)] + Q, \tag{35}$$

$$\frac{dB^T(t)}{dt} = 2B^T(t)A(t) - B^T(t)D - 2d^T A(t) + p^T, \tag{36}$$

$$\frac{dC(t)}{dt} = trA(t) + \frac{1}{2}B^T B(t) - d^T B(t) + r, \tag{37}$$

where $D = (d_{ij})$ *is a* $n \times n$ *matrix and* $d^T = (d_1, \dots, d_n)$ *is a* $1 \times n$ *matrix.*

3.2 Algorithms

In this section, we first give a new way to do Gaussian approximation in Algorithm 1. Then the computation of $\tilde{u}_k(t,x)$ is summarised as an application in Algorithm 2.

The idea of our Gaussian approximation is the following: given a probability density $\phi(x)$, we fit it with Gaussian distributions using the peaks of $\phi(x)$ as the mean. The procedure is repeated until the peaks of $\phi(x) - g(x)$ is no larger than some threshold E, where $g(x)$ is the sum of Gaussians from previous fitting steps. In Section 4.1, the numerical experiments show that our Gaussian approximation method works very well.

Algorithm 1 Gaussian approximation

1: Let $f(x) = \phi(x)$ and the threshold $E = \alpha \max \phi(x)$, where α is a given small number.
2: Fitting the peaks of $f(x)$ which are larger than E with Gaussian distributions. Specifically, for a peak $P_i(x_i, y_i)$ of $f(x)$ with $y_i \geq E$, we use the function $g_i(x) = y_i \exp(-\frac{(x-x_i)^2}{2\sigma_i^2})$ to fit $P_i(x_i, y_i)$ with points in a neighborhood of $P_i(x_i, y_i)$ where no other peaks exists, and the best fitting parameter σ_i is obtained by fitting. Suppose the sum of Gaussian distributions $g_i(x)$ in this step is $g(x)$.
3: Let $f_1(x) = f(x) - g(x)$. If $f_1(x)$ has no peaks whose values larger than E, then go to step 4. Otherwise, let $f(x) = f_1(x)$ and go to step 2.
4: Let $f_2(x) = -f_1(x)$. If $f_2(x)$ has no peaks which are larger than E, then done. Otherwise, let $f(x) = f_2(x)$ and go to step 2.

Using the above Gaussian approximation procedure, we can decompose (30) into a finite number of Gaussian distributions. By Theorem 3.3, the Kolmogorov equation (29) with Gaussian initial condition is solved in terms of ODEs. The algorithm to compute $\tilde{u}_k(t,x)$ is list in Algorithm 2.

Algorithm 2 Compute $\tilde{u}_k(t,x)$

1: Choose the total computing time T, Δt and the parameter α in Algorithm 1. Let $N = \frac{T}{\Delta t}$, and partition the time interval $[0, T]$ by $\{0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_N = T\}$.
2: **for** $k = 1 : N$ **do**
3: Using Algorithm 1, suppose $\hat{u}_k(\tau_{k-1}, x)$ is decomposed into $\sum_{i=1}^{N(k)} c_{k,i} G(\mu_{k,i}, \sigma_{k,i})$.
4: For each Gaussian distribution $G(\mu_{k,i}, \sigma_{k,i})$, suppose the solution of (29) with initial condition $G(\mu_{k,i}, \sigma_{k,i})$ is $\hat{u}_{k,i}(\tau_k, x)$. Solving (35)–(37), we obtain $\hat{u}_{k,i}(\tau_k, x)$. Then $\hat{u}_k(\tau_k, x) = \sum_{i=1}^{N(k)} c_{k,i}\hat{u}_{k,i}(\tau_k, x)$.
5: From (31), we have $\tilde{u}_k(\tau_k, x) = e^{-F(x)}\hat{u}_k(\tau_k, x)$.
6: By (30), we obtain $\hat{u}_{k+1}(\tau_k, x)$.
7: **end for**

4. Numerical experiments

In this section, we first use two examples to show the effectiveness of the above Gaussian approximation algorithm, then two concrete NLF models are considered to verify the effectiveness of the direct method. The computing platform we used have Intel(R) Xeon(R) CPU E5-2670 v3 @ 2.30GHz. We use the mean-squared error (MSE) as the performance metric. The average MSE over 100 times simulations is provided.

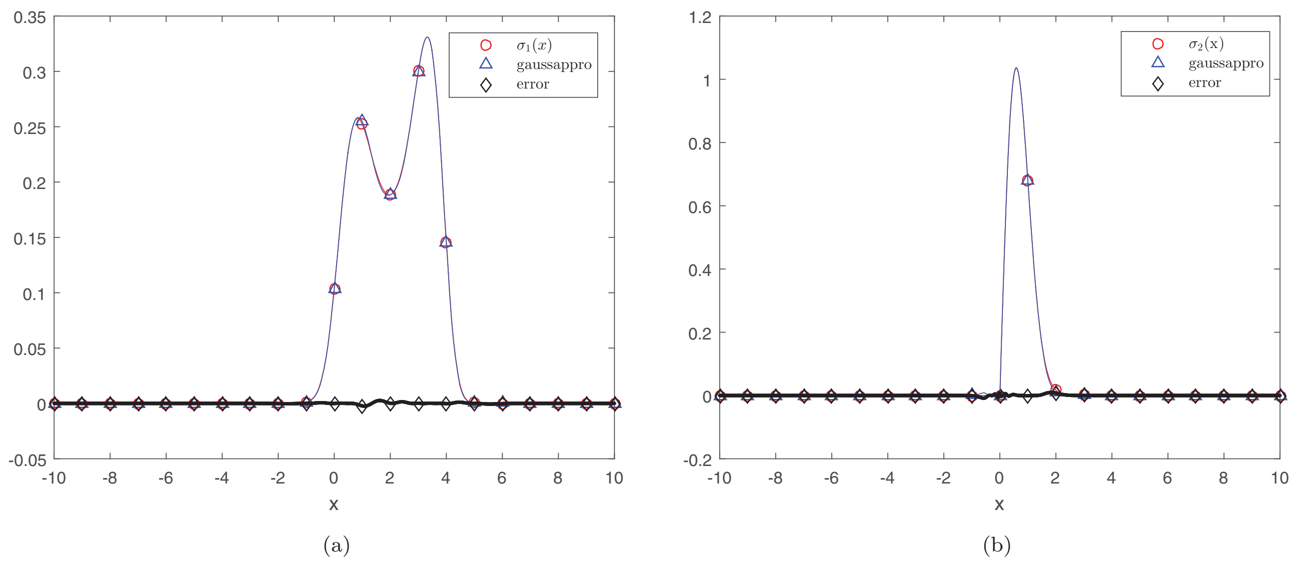


Figure 1. Gaussian approximation of a given probability density function, (a) Gaussian approximation of $\sigma_1(x)$, (b) Gaussian approximation of $\sigma_2(x)$.

4.1 Gaussian approximation

We approximate a given probability density function by $\sum_{i=1}^{N} c_i e^{\frac{-(x-\mu_i)^2}{2\sigma_i^2}}$ by using Algorithm 1. Let $\alpha = 0.01$ and we consider the following two examples.

Example 4.1: Let the probability density function be $\sigma_1(x) = \frac{1}{71.2186} e^{-x\sin x - \frac{1}{2}x\cos x - x^2 + 3x + 2}$. The performance of our Gaussian approximation algorithm is given in Figure 1(a). The MSE is 2.5680×10^{-7}. The average elapsed time is 0.2971 seconds. The Gaussian distributions we used is given in Table 1.

Example 4.2: Let the probability density function be Rayleigh distribution: $\sigma_2(x) = \frac{x}{b^2} e^{-\frac{x^2}{2b^2}}$ with $b = 0.585$. The performance of our Gaussian approximation algorithm is given in Figure 1(b). The MSE is 3.2128×10^{-6}. The average elapsed time is 0.3110 seconds. The Gaussian distributions we used is given in Table 2.

From the above examples, we can see that the proposed Gaussian approximation algorithm is very effective and fast.

4.2 Direct method

In this section, the performance of the direct method and EKF is compared. We consider the following NLF system which satisfies condition (C_1)–(C_3)

$$\begin{cases} dx_t = f(x_t)dw_t \\ dy_1(t) = x_t \sin(x_t)dt + dv_1(t) \\ dy_2(t) = x_t \cos(x_t)dt + dv_2(t) \end{cases} \tag{38}$$

Here $w(t)$, $v_1(t)$, $v_2(t)$ are scalar independent standard Brownian motions. The initial distribution is taken as $\sigma_0(x) = \frac{1}{71.2186} e^{-x\sin x - \frac{1}{2}x\cos x - x^2 + 3x + 2}$. The initial values for EKF are $\hat{x}_0$ and P_0. In Gaussian approximation algorithm, we choose the parameter $\alpha = 0.01$. The total simulation time is T and the time step is Δt.

Example 4.3: In this example, we take the drift $f(x) = \tanh(x)$ in (38).

(1) With $T = 30$ seconds, $\Delta t = 0.1$ seconds and $\hat{x}_0 = 1$. For $P_0 = 1$, the performance of our method and EKF is given in Figure 2(a). One can see that the EKF completely fails at about $t = 13$ seconds. The average MSE for 100 experiments by direct method and EKF are given in Table 3.

Table 1. Gaussian distributions used to approximate $\sigma_1(x)$.

i	1	2	3	4	5	6	7	8	9	10
c_i	0.2574	0.3311	0.0539	− 0.0409	− 0.0063	− 0.0791	− 0.0034	− 0.0071	0.0047	0.0085
μ_i	0.8400	3.3200	2.0600	− 0.2500	2.9700	4.2500	− 1.0200	4.9700	0.4800	3.5900
σ_i	0.7511	0.7039	0.4013	0.4204	0.2033	0.3849	0.3407	0.3171	0.2197	0.2249

Table 2. Gaussian distributions used to approximate $\sigma_2(x)$.

i	1	2	3	4	5	6	7	8
c_i	1.0367	0.0984	− 0.3590	− 0.0165	− 0.0596	0.0520	0.0662	− 0.0456
μ_i	0.5900	1.2900	0	0.7700	− 0.3100	− 0.0800	0.0900	0
σ_i	0.4051	0.3066	0.1375	0.1546	0.1316	0.0614	0.0673	0.0189

Table 3. Average MSE of direct method and EKF.

P_0	MSE-direct method[a]	Time-direct metehod[b](s)	MSE-EKF[c]
1	0.1786	19.4947	–
5	0.2206	19.7331	–
10	0.1821	18.8867	–

[a] The average MSE of direct method for 100 experiments.
[b] The average elapsed time of direct method for 100 experiments.
[c] The average MSE of EKF for 100 experiments.

Table 4. Performance of direct method with different Δt.

Δt	0.1	0.01	0.001
MSE	0.3451	0.2236	0.1714

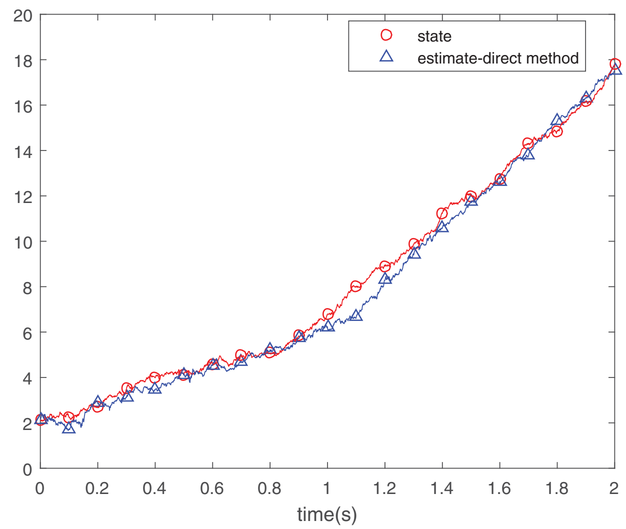


Figure 3. State estimation by direct method with $T = 2$, $\Delta t = 0.001$ seconds.

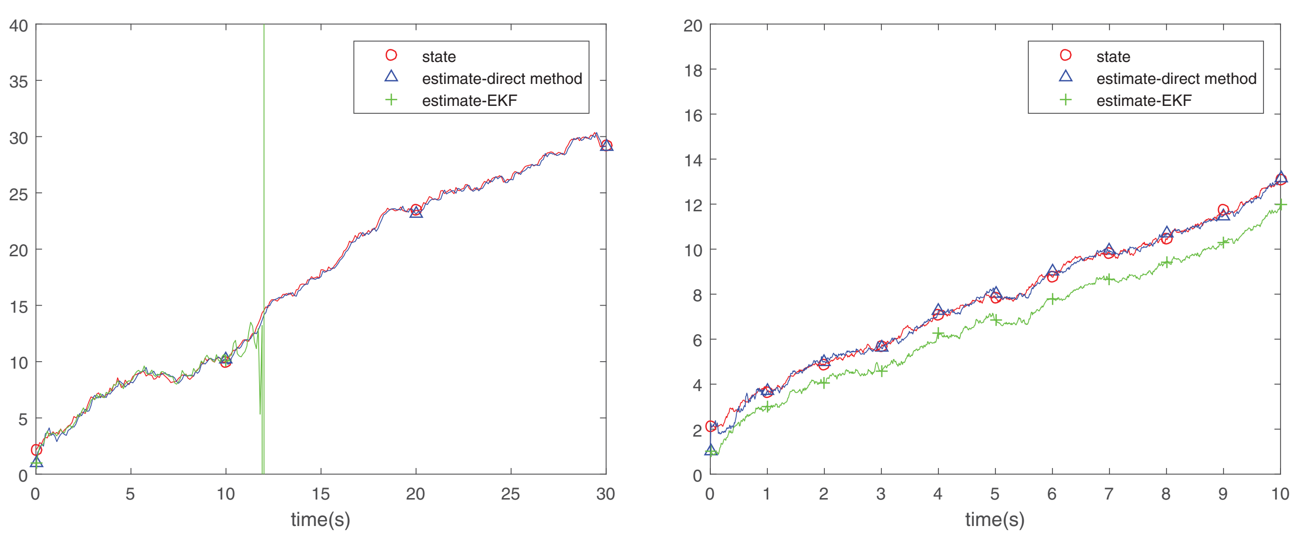
(2) In this case we take $T = 10$ seconds and $\Delta t = 0.01$ seconds. For $\hat{x}_0 = 1$ and $P_0 = 5$, the performance of the direct method and EKF is given in Figure 2(b). The average MSE for 100 experiments of the direct method and EKF is 0.1826 and 6.1364, respectively.

(3) With $T = 10$ seconds and different time step, the corresponding MSE of our method is given in Table 4, from which we can see that with smaller time step, the higher the estimation precision is.

Example 4.4: In this example, we take $T = 2$ seconds, $\Delta t = 0.001$ seconds. The drift $f(x) = x + 1 + \frac{dF}{dx}$, where

$$F(x) = \int_{-\infty}^{x} \{e^{-(z-\frac{1}{2})^2} \Big/ \int_{-\infty}^{z} e^{-(y-\frac{1}{2})^2} dy - 3/2\} dz.$$

For $\hat{x}_0 = 1$, $P_0 = 3$, the EKF just blow up, the performance of our method is given in Figure 3. The MSE of our method is 0.2016.



(a) (b)

Figure 2. State estimation by direct method and EKF, (a) $T = 30$, $\Delta t = 0.1$ seconds, (b) $T = 10$, $\Delta t = 0.01$ seconds.

From the above two examples, we can see that our method can track the state very well and it can be implemented in a real time manner. Besides, compared to EKF, our method is much more stable.

5. Conclusion

In this paper, we consider a general class of NLF sytems, called Yau filtering system with arbitrary initial condition. The observations can also be nonlinear which occurs in many practical examples. We show that the solution of the robust DMZ equation is reduced to the solution of Kolmogorov equation. Based on Gaussian approximation, the Kolmogorov equation can be solved in terms of ODEs. On the one hand, the direct method can be implemented in real time and memoryless way. On the other hand, it is easy to implement and do not need the controllability and observability assumption. Compared to EKF, it is much stable and has theoretical convergence proof. Besides, we give a new approach to decompose a given probability density function into a finite number of Gaussian distributions. The proposed Gaussian approximation method is very simple and fast. The numerical experiments show that our Gaussian approximation method is very effective and the direct method can track the states very well.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Natural Science Foundation of China [grant number 11471184] and the start-up fund from Tsinghua University.

References

Benés, V. (1981). Exact finite dimensional filters for certain diffusions with nonlinear drift. *Stochastics, 5*, 65–92. doi: 10.1080/17442508108833174.

Brockett, R.W. (1981). Nonlinear systems and nonlinear estimation theory. InM. Hazewinkel J.S. Willems (Eds.) *The mathematics of filering and identification and applications* (pp. 441–477). Berlin, Germany: Springer Netherlands.

Brockett, R.W., & Clark J.M.C,. (1980). The geometry of the conditional density equation. In R.W. Brokett J.M.C. Clark (Eds.) *International Conference on Analysis and optimisation of stochastic systems* (pp. 299–309). New York, NY: Academic Press.

Chen, J. (1994, June). On ubiquity of Yau filters. *Proceedings of the American Control Conference* (pp. 252–254). Baltimore, MD: IEEE. doi: 10.1109/ACC.1994.751736.

Chiou, W.L., & Yau, S.S.T. (1994). Finite-dimensional filters with nonlinear drift II: Brockett's problem on classification of finite-dimensional estimation algebras. *SIAM Journal of Control Optimization, 32*, 297–310.

Davis, M.H.A. (1980). On a multiplicative functional transformation arising in nonlinear filtering theory. *Z. Wahrsch. Verw Gebiete, 54*, 125–139.

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, Series B *39*(1), 1–38.

Duncan, T.E. (1967). *Probability densities for diffusion processes with applications to nonlinear filtering system* (Ph D dissertation). Stanford University.

Kalman, R.E., & Bucy, R.S. (1961). New results in linear filtering and prediction theory. *Transactions of the ASME-Journal of Basic Engineering, 83*, 95–108.

Luo, X., & Yau, S.S.T. (2013a). Hermite spectral method to 1-D forward Kolmogorov equation and its application to nonliner filtering problems. *IEEE Transaction on Automatic Control, 58*(10), 2495–2507.

Luo, X., & Yau, S.S.T. (2013b). Complete real time solution of the general nonlinear filtering problem without memory. *IEEE Transaction on Automatic Control, 58*(10), 2563–2578.

Mortensen, R.E. (1966). *Optimal control of continuous time stochastic systems* (Ph D dissertation). University of California, Berkeley.

Mitter, S.K. (1979). On the analogy between mathematical problems of nonlinear filtering and quantum physics. *Ricerche Automat, 10*, 163–216.

Tam, L.F., Wong, W.S., & Yau, S.S.T. (1990). On a necessary and sufficient condition for finite dimensionality of estimation algebras. *SIAM Journal on Control and Optimization, 28*(1), 173–185.

Wei, J., & Norman, E. (1964). On the global representation of the solutions of linear differential equations as a product of exponentials. *Proceedings of the American Mathematical Society, 15*, 327–334. doi: 10.1090/S0002-9939-1964-0160009-0.

Yau, S.S.T. (1990, December). Recent results on linear filtering: New class of finite dimensional filter. *Proceeding of the 29th Conference Decision and Control* (pp. 231–233). Honolulu, HI: IEEE. doi: 10.1109/CDC.1990.203587.

Yau, S.S.T. (1994). Finite-dimensional filters with nonlinear drift I: A class of filters including both Kalman-Bucy filters and Benes filters. *Journal of Mathematical Systems, Estimation, and Control, 4*, 181–203.

Yau, S.S.T., & Hu, G.Q. (2001). Finite-dimensional filters with nonlinear drift X: Explicit solution of DMZ equation. *IEEE Transactions on Automatic Control, 46*, 142–148.

Yau, S.S.T., & Hu, G.Q. (2005). Classification of finite-dimensional estimation algebras of maximal rank with arbitrary state-space dimension and mitter conjecture. *International Journal of Control, 78*, 689–705.

Yau, S.S.T., & Lai, Y.T. (2003). Explicit solution of DMZ equation in nonlinear filtering via solution of ODE. *IEEE Transactions on Automatic Control, 48*, 505–508.

Yau, S.S.T., & Yau, S.T. (1994, December). New direct method for Kalman-Bucy filtering system with arbitrary initial condition. *Proceeding of the 33rd Conference on Decision and Control* (pp. 1221–1225). Lake Buena Vista, FL. doi: 10.1109/CDC.1994.411164.

Yau, S.S.T., & Yau, S.T. (1997). Finite dimensional filters with nonlinear drift III: Duncan-Mortensen-Zakai equation with arbitrary initial condition for linear filtering system and the Benés filtering system. *IEEE Transactions on Aerospace and Electronic Systems, 33*, 1277–1294.

Yau, S.T., & Yau, S.S.T. (2000). Real time solution of nonlinear filtering problem withour memory. *Mathematics Research Letters, 7*, 671–693.

Yau, S.T., & Yau, S.S.T. (2004a). Nonlinear filtering and time varying Schrödinger eqution I. *IEEE Transactions on Aerospace and Electronic Systems, 40*, 284–292.

Yau, S.S.T., & Yau, S.T. (2004b, December). Linear filtering with nonlinear observations. In *43rd IEEE Conference on Decision and Control* (pp. 2112–2117). Atlantis, Paradise island. doi: 10.1109/CDC.2004.1430360.

Yau, S.S.T., & Yau, S.T. (2005). Solution of filtering problems with nonlinear observations. *SIAM Journal on Control Optimization, 44*, 1019–1039.

Zakai, M., & Wahrsch, Z. (1969). On the optimal filtering of diffusion processes. *Gujarat Electricity Board, 11*, 230–243.