Correspondence

Novel Suboptimal Filter via Higher Order Central Moments

In this paper, we construct a new suboptimal filter by deriving the Ito's stochastic differential equations of the estimation of higher order central moments, satisfy, and impose some conditions to form a closed system. The essentially infinite-dimensional cubic sensor problem has been investigated in detail numerically to illustrate the reasonableness of the imposed conditions, and the numerical experiments support our discussion. A two-dimensional polynomial filtering problem has also been experimented.

I. INTRODUCTION

The nonlinear filtering (NLF) problem involves the estimation of a stochastic process (called the signal or state process) that cannot be observed directly. Information containing the state is obtained from observations of a related process, i.e., the observation process. The main goal of NLF is to determine the conditional expectations, or perhaps even to compute the entire conditional density of the state, given the observation history. For an excellent introduction to NLF theory, we refer the readers to the book by Jazwinski [1].

In 1960, Kalman [2] published a historically important paper on linear filtering that is highly influential in modern industry. It is the so-called Kalman filter (KF). One year later, the continuous version of KF was investigated by Kalman and Bucy [3]. Since then, the Kalman-Bucy filter has been widely used in science and engineering, for example in navigation and guidance systems, radar tracking, sonar ranging, satellite and airplane orbit determination, and forecasting in weather, econometrics, and finance. However, the Kalman-Bucy filter has limited application due to the linearity assumptions of the drift term, the observation term, and the Gaussian assumption of the initial value.

The success of KF for the linear Gaussian estimation problems encouraged many researchers to generalize Kalman's results to nonlinear dynamical systems.

Manuscript received January 1, 2015; revised October 15, 2015, March 22, 2016; released for publication March 30, 2016.

DOI. No. 10.1109/TAES.2016.150003.

Refereeing of this contribution was handled by S. Mori.

X. Luo acknowledges the support of the Fundamental Research Funds for the Central Universities (YWF-15-SXXY-006, YWF-15-SXXY-002, YWF-16-SXXY-001) and National Natural Science Foundation of China (11501023). S. S.-T. Yau thanks the financial support of National Natural Science Foundation of China (11471184) and the start-up fund from Tsinghua University. He is grateful to National Center for Theoretical Sciences (NCTS) for providing excellent research environment while part of this research was done.

Dedicated to Peter Caine on the occasion of his 70th birthday.

However, the NLF problem is an essentially more difficult problem since the resulting optimal filter is, in general, infinite-dimensional, i.e., the conditional density depends on all its moments. Those methods which attempt to compute the density function directly or numerically are called the global approaches, see the survey paper [4] for details.

Although the global ones can completely solve the NLF problems, the heavy computation is one of the major obstacles in their real-time applications. Another way out is to use the approximate method to construct a suboptimal filter. The existing approximate filters for the NLF problems include the extended Kalman filter (EKF), the unscented Kalman filter (UKF), the ensemble Kalman filter (EnKF), particle filters (PF), and the splitting up method; see [5–8]. All of these methods have their own weakness. UKF and EnKF assume that the probability density of the state vector is Gaussian. PF could be inefficient and is sensitive to outliers. Resampling step is applied at every iteration, which results in a rapid loss of diversity in particles. Furthermore, PF are more applicable at low- and moderate high-dimensional systems; see [9] for the obstacles to high dimensional cases. The splitting up method requires g and h in the model (1) to be bounded, which even excludes the linear case. Recently, Germani et al. [10–11] developed a suboptimal method, so-called Carleman approach, based on the algorithm for the bilinear system [12]. However, recently the first and the last author found that the Carleman approach can fail completely in some one-dimensional NLF problem and developed a suboptimal method via Hermite polynomials [13]. The use of higher central moments to improve the performance of NLF has been studied by many researchers; see [14] and references therein. In fact, the cumulants can be a better choice than the central moments, and the study on the cumulants for NLF can be found in [15].

In this paper, we shall propose a new suboptimal filter by investigating the Ito's stochastic differential equation (SDE) which the higher central moments satisfy. Although the use of the higher central moments for NLF problems has been attempted for a long time and the second order EKF has been standard in the literature, see [1], the detailed derivation has never been clearly written down for NLF, especially the polynomial filtering problems, which can be viewed as the truncation of Taylor expansion of any nonlinear smooth functions. When arrived at an infinite dimensional system, the higher central moments are *conventionally* truncated to form a closed system, as in [16]. No one doubts the reasonableness of the truncation. It is in this paper that for the first time we investigate other options to form a closed system, say condition (12). The numerical experiments support the condition. Also we compare our methods with some existing ones. Our method works in nearly perfect agreement with theory.

An outline of this paper is as follows. In Section II, we introduce the continuous-time model in this paper. Our method is derived and described in Section III. Section IV is devoted to two numerical experiments, which validate our method. Our method is more flexible by choosing different truncation mode $\vec{N}$. The conclusion is in Section V.

II. PRELIMINARIES

The model we consider in this paper is the continuous-state-continuous-observation one:

$$\begin{cases} dx_t = f(x_t, t)dt + g(x_t, t)dv_t \\ dy_t = h(x_t, t)dt + dw_t, \end{cases} \tag{1}$$

where x_t, v_t, y_t, and w_t are $\mathbb{R}^n-$, $\mathbb{R}^p-$, $\mathbb{R}^m-$, and $\mathbb{R}^m-$valued processes, respectively, and $f:\mathbb{R}^n\times\mathbb{R}_+\to\mathbb{R}^n$, $g:\mathbb{R}^n\times\mathbb{R}_+\to\mathbb{R}^{n\times m}$, $h:\mathbb{R}^n\times\mathbb{R}_+\to\mathbb{R}^m$ are possibly nonlinear function of x. Assume that $\{v_t, t\geq 0\}$ and $\{w_t, t\geq 0\}$ are Brownian motion processes with $Var[dv_t]$ $= Q(t)dt$ and $Var[dw_t] = R(t)dt$, respectively. Moreover, $\{v_t, t\geq 0\}$, $\{w_t, t\geq 0\}$ and x_0 are independent. The initial observation is assumed to be $y_0 = 0$.

Without loss of generality, we assume $Q(t)$ is a diagonal matrix, $Q(t) = diag(q_1^2, \ldots, q_n^2)$. In fact, if it is not, we have spectral decomposition of $Q(t) = P\Lambda P'$, where $PP' = I$, Λ is diagonal matrix. By letting $g^* = gP$, $dv^* = P'dv$, then $Var[dv^*] = \Lambda dt$. We could further assume that $Q(t) = I$, due to the function g in front (replacing g by $gQ^{1/2}$).

Let us clarify the notations we shall use in this paper. Let $p \equiv p(x, t \mid Y_t)$ be the conditional probability density function of the state x_t, given the observation history $Y_t \equiv \{y_s, 0\leq s\leq t\}$, then the conditional expectation of x_t is defined as

$$\hat{x}_t \equiv E^t[x_t] \equiv E[x_t \mid Y_t]. \tag{2}$$

For conciseness, we may use the vector notations, denoted as $\vec{k} = (k_1, k_2, \cdots, k_n)$. We say $\vec{k}\leq\vec{\alpha}$, if $k_i\leq\alpha_i$, for all $1\leq i\leq n$. The strict inequality holds, if $k_i<\alpha_i$, for some $1\leq i\leq n$.

We denote $P_{\vec{k}}$ as

$$\begin{aligned} P_{\vec{k}} &\equiv E^t\left[(x_1-\hat{x}_1)^{k_1}\ldots(x_n-\hat{x}_n)^{k_n}\right] \\ &\equiv E\left[(x_1-\hat{x}_1)^{k_1}\ldots(x_n-\hat{x}_n)^{k_n} \mid Y_t\right]. \end{aligned}$$

Say $P_{\vec{k}}$ is the lower order of $P_{\vec{\alpha}}$ if $\vec{k}<\vec{\alpha}$. By convention, $\vec{0} = (0, 0, \cdots, 0)$ and $\vec{e}_i$ denotes 1 for the i-th component, 0 otherwise. $P_{\vec{0}} = 1$ and $P_{\vec{e}_i} = 0$, for $1\leq i\leq n$.

Furthermore, $\min\{\vec{k}, \vec{l}\} = (\min\{k_1, l_1\}, \min\{k_2, l_2\}, \cdots, \min\{k_n, l_n\})$, $\vec{k}+\vec{l} = (k_1+l_1, \cdots, k_n+l_n)$, $|\vec{k}|_1 = \sum_{i=1}^n k_i$, and $|\vec{k}|_\infty = \max_{1\leq i\leq n} k_i$.

III. NEW SUBOPTIMAL FILTER

Let $f_i(x, t)$, $g_{ij}(x, t)$, and $h_i(x, t)$, $1\leq i\leq n$, $1\leq j\leq m$, be some smooth nonlinear functions in x. They can be approximated by their truncated Taylor expansions:

$$f_i(x,t) \approx \sum_{|\vec{m}|_1 \le M_f} f_{i;\vec{m}}(t) \prod_{a=1}^{n} x_a^{m_a} \tag{3}$$

$$g_{ij}(x,t) \approx \sum_{|\vec{m}|_1 \le M_g} g_{ij;\vec{m}}(t) \prod_{a=1}^{n} x_a^{m_a} \tag{4}$$

$$h_i(x,t) \approx \sum_{|\vec{m}|_1 \le M_h} h_{i;\vec{m}}(t) \prod_{a=1}^{n} x_a^{m_a} \tag{5}$$

where M_f, M_g, and M_h are the highest degrees kept in the expansions of $\{f_i\}_{1\le i\le n}$, $\{g_{ij}\}_{1\le i\le n, 1\le j\le p}$ and $\{h_i\}_{1\le i\le m}$, respectively.

In the sequel, we shall focus on the derivation of the method for the polynomial filtering problems.

PROPOSITION 3.1 For continuous filtering problem given by the system (1) with $f_i(x,t)$, $g_{ij}(x,t)$, $h_i(x,t)$ approximated by (3)–(5), the conditional mean $\hat{x}_i$ satisfies the following Ito's SDE

$$\begin{aligned} d\hat{x}_i = &\sum_{|\vec{m}|_1 \le M_f} \sum_{\vec{0}\le\vec{k}\le\vec{m}} f_{i;\vec{m}} \left(\prod_{a=1}^{n} \binom{m_a}{k_a} (\hat{x}_a)^{m_a-k_a} \right) P_{\vec{k}} dt \\ &+ \sum_{1\le j,s\le m} r^{js} \Bigg(dy_j - \sum_{|\vec{m}|_1 \le M_h} \sum_{\vec{0}\le\vec{k}\le\vec{m}} h_{j;\vec{m}} \Bigg(\prod_{a=1}^{n} \binom{m_a}{k_a} \\ &\times (\hat{x}_a)^{m_a-k_a} \Bigg) P_{\vec{k}} dt \Bigg) \cdot \Bigg(\sum_{|\vec{m}|_1 \le M_h} \sum_{\vec{0}\le\vec{k}\le\vec{m}} h_{s;\vec{m}} \Bigg(\prod_{a=1}^{n} \binom{m_a}{k_a} \\ &\times (\hat{x}_a)^{m_a-k_a} \Bigg) P_{\vec{k}+\vec{e}_i} \Bigg), \end{aligned} \tag{6}$$

where $(r^{js})_{m\times m}$ is the matrix R^{-1}.

PROOF According to [1], the conditional mean $\hat{x}_i$ satisfies

$$d\hat{x}_i = \hat{f}_i dt + \left(dy - \hat{h}dt\right)^T R^{-1} \left(\widehat{hx_i} - \hat{h}\hat{x}_i\right). \tag{7}$$

Using binomial expansion, we have

$$\begin{aligned} \prod_{a=1}^{n} x_a^{m_a} &= \prod_{a=1}^{n} (x_a - \hat{x}_a + \hat{x}_a)^{m_a} \\ &= \prod_{a=1}^{n} \sum_{\vec{0}\le\vec{k}\le\vec{m}} \binom{m_a}{k_a} (x_a-\hat{x}_a)^{k_a} (\hat{x}_a)^{m_a-k_a} \\ &= \sum_{\vec{0}\le\vec{k}\le\vec{m}} \prod_{a=1}^{n} \binom{m_a}{k_a} (x_a-\hat{x}_a)^{k_a} (\hat{x}_a)^{m_a-k_a} \end{aligned} \tag{8}$$

then

$$\begin{aligned} \prod_{a=1}^{n} (x_a-\hat{x}_a)^{\alpha_a} h_s = &\sum_{|\vec{m}|_1 \le M_h} \sum_{\vec{0}\le\vec{k}\le\vec{m}} h_{s;\vec{m}} \prod_{a=1}^{n} \binom{m_a}{k_a} \\ &\times (x_a-\hat{x}_a)^{\alpha_a+k_a} (\hat{x}_a)^{m_a-k_a}, \end{aligned}$$

and hence

$$\begin{aligned} &E^t\left[\prod_{a=1}^{n} (x_a-\hat{x}_a)^{\alpha_a} h_s\right] \\ &= \sum_{|\vec{m}|_1 \le M_h} \sum_{\vec{0}\le\vec{k}\le\vec{m}} h_{s;\vec{m}} \left(\prod_{a=1}^{n} \binom{m_a}{k_a} (\hat{x}_a)^{m_a-k_a} \right) P_{\vec{\alpha}+\vec{k}}. \end{aligned} \tag{9}$$

Similarly, we have

$$\begin{aligned} &E^t\left[\prod_{a=1}^{n} (x_a-\hat{x}_a)^{\alpha_a} f_i\right] \\ &= \sum_{|\vec{m}|_1 \le M_f} \sum_{\vec{0}\le\vec{k}\le\vec{m}} f_{i;\vec{m}} \left(\prod_{a=1}^{n} \binom{m_a}{k_a} (\hat{x}_a)^{m_a-k_a} \right) P_{\vec{\alpha}+\vec{k}}. \end{aligned} \tag{10}$$

Especially,

$$\begin{aligned} \hat{f}_i &= E^t\left[\prod_{a=1}^{n} (x_a-\hat{x}_a)^0 f_i\right] \\ &\overset{10}{=} \sum_{|\vec{m}|_1 \le M_f} \sum_{\vec{0}\le\vec{k}\le\vec{m}} f_{i;\vec{m}} \left(\prod_{a=1}^{n} \binom{m_a}{k_a} (\hat{x}_a)^{m_a-k_a} \right) P_{\vec{k}}, \\ \hat{h}_j &= E^t\left[\prod_{a=1}^{n} (x_a-\hat{x}_a)^0 h_j\right] \\ &\overset{(9)}{=} \sum_{|\vec{m}|_1 \le M_h} \sum_{\vec{0}\le\vec{k}\le\vec{m}} h_{j;\vec{m}} \left(\prod_{a=1}^{n} \binom{m_a}{k_a} (\hat{x}_a)^{m_a-k_a} \right) P_{\vec{k}}, \end{aligned}$$

and

$$\begin{aligned} &\widehat{h_s x_i} - \widehat{h_s}\hat{x}_i \\ &= E^t[(x_i-\hat{x}_i)h_s] \\ &\overset{(9)}{=} \sum_{|\vec{m}|_1 \le M_h} \sum_{\vec{0}\le\vec{k}\le\vec{m}} h_{s;\vec{m}} \left(\prod_{a=1}^{n} \binom{m_a}{k_a} (\hat{x}_a)^{m_a-k_a} \right) P_{\vec{k}+\vec{e}_i}. \end{aligned}$$

Equation (6) is followed immediately by plugging the above three equalities into (7) with the fact that $(dy - \hat{h}dt)' R^{-1} (\widehat{hx_i} - \hat{h}\hat{x}_i) = \sum_{j=1}^{m} (dy_j - \widehat{h}_j dt) \left[\sum_{s=1}^{m} r^{js} (\widehat{h_s x_i} - \widehat{h}_s \hat{x}_i)\right]$. ■

It is clear to see that in (6), the central moments $P_{\vec{k}+\vec{e}_i}$ for $\vec{k} \le \vec{m}$, with $|\vec{m}|_1 \le M_h$ and $P_{\vec{k}}$ for $\vec{k} \le \vec{m}$, with $|\vec{m}|_1 \le M_f$ are needed to compute $\hat{x}_i$. Thus, let us give the Ito's SDE for $P_{\vec{\alpha}}$ with $|\vec{\alpha}|_1 \ge 2$ in the following proposition.

PROPOSITION 3.2 For continuous filtering problem given by the system (1) with $f_i(x,t)$, $g_{ij}(x,t)$, and $h_i(x,t)$

approximated by (3)–(5), the SDE for $P_{\vec{\alpha}}$ is

$$
\begin{aligned}
dP_{\vec{\alpha}} &= \Bigg(-\sum_{a=1}^{n}\alpha_a \sum_{|\vec{m}|_1\le M_f}\sum_{\vec{0}\le\vec{k}\le\vec{m}} f_{a;\vec{m}} \\
&\times\left(\prod_{b=1}^{n}\binom{m_b}{k_b}(\hat{x}_b)^{m_b-k_b}\right)P_{\vec{k}}P_{\vec{\alpha}-\vec{e}_a} \\
&+\frac{1}{2}\sum_{a=1}^{n}\alpha_a\Bigg(\alpha_a-1)\Bigg(\sum_{1\le i,j\le n} r^{ij}\Bigg(\sum_{|\vec{m}|_1\le M_h}\sum_{\vec{0}\le\vec{k}\le\vec{m}} h_{i;\vec{m}} \\
&\times\left(\prod_{b=1}^{n}\binom{m_b}{k_b}(\hat{x}_b)^{m_b-k_b}\right)P_{\vec{k}+\vec{e}_a}\Bigg) \\
&\cdot\Bigg(\sum_{|\vec{m}|_1\le M_h}\sum_{\vec{0}\le\vec{k}\le\vec{m}} h_{j;\vec{m}}\left(\prod_{b=1}^{n}\binom{m_b}{k_b}(\hat{x}_b)^{m_b-k_b}\right)P_{\vec{k}+\vec{e}_a}\Bigg)P_{\vec{\alpha}-2\vec{e}_a}\Bigg) \\
&+\sum_{i=1}^{n}\sum_{|\vec{m}|_1\le M_f}\sum_{\vec{0}\le\vec{k}\le\vec{m}}\alpha_i f_{i;\vec{m}}\left(\prod_{a=1}^{n}\binom{m_a}{k_a}(\hat{x}_a)^{m_a-k_a}\right)P_{\vec{\alpha}+\vec{k}-\vec{e}_i} \\
&+\sum_{\substack{1\le i<j\le n,\\ 1\le l\le n}}\ \sum_{\substack{|\vec{m}^1|_1\le M_g,\\ |\vec{m}^2|_1\le M_g}}\ \sum_{\vec{0}\le\vec{k}\le\vec{m}^1+\vec{m}^2}\alpha_i\alpha_j g_{li;\vec{m}^1}g_{lj;\vec{m}^2} \\
&\times\left(\prod_{a=1}^{n}\binom{m_a^1+m_a^2}{k_a}(\hat{x}_a)^{m_a^1+m_a^2-k_a}\right)P_{\vec{\alpha}+\vec{k}-\vec{e}_i-\vec{e}_j} \\
&+\frac{1}{2}\sum_{i,l=1}^{n}\ \sum_{\substack{|\vec{m}^1|\le M_g,\\ |\vec{m}^2|_1\le M_g}}\ \sum_{\vec{0}\le\vec{k}\le\vec{m}^1+\vec{m}^2}\alpha_i(\alpha_i-1)g_{li;\vec{m}^1}g_{li;\vec{m}^2} \\
&\times\left(\prod_{a=1}^{n}\binom{m_a^1+m_a^2}{k_a}(\hat{x}_a)^{m_a^1+m_a^2-k_a}\right)P_{\vec{\alpha}+\vec{k}-2\vec{e}_i} \\
&+\sum_{a<b}\Bigg(\alpha_a\alpha_b P_{\vec{\alpha}-\vec{e}_a-\vec{e}_b}\Bigg(\sum_{1\le i,j\le n} r^{ij}\Bigg(\sum_{|\vec{m}|_1\le M_h}\sum_{\vec{0}\le\vec{k}\le\vec{m}} h_{i;\vec{m}} \\
&\times\left(\prod_{c=1}^{n}\binom{m_c}{k_c}(\hat{x}_c)^{m_c-k_c}\right)P_{\vec{k}+\vec{e}_a}\Bigg) \\
&\cdot\Bigg(\sum_{|\vec{m}|_1\le M_h}\sum_{\vec{0}\le\vec{k}\le\vec{m}} h_{j;\vec{m}}\left(\prod_{c=1}^{n}\binom{m_c}{k_c}(\hat{x}_c)^{m_c-k_c}\right)P_{\vec{k}+\vec{e}_b}\Bigg)\Bigg)\Bigg) \\
&-\sum_{a=1}\Bigg(\alpha_a\sum_{i,j} r^{ij}\Bigg(\sum_{|\vec{m}|_1\le M_h}\sum_{\vec{0}\le\vec{k}\le\vec{m}} h_{i;\vec{m}} \\
&\times\left(\prod_{b=1}^{n}\binom{m_b}{k_b}(\hat{x}_b)^{m_b-k_b}\right)P_{\vec{k}+\vec{e}_a}\Bigg) \\
&\cdot\Bigg(\sum_{|\vec{m}|_1\le M_h}\sum_{\vec{0}\le\vec{k}\le\vec{m}} h_{j;\vec{m}}\left(\prod_{b=1}^{n}\binom{m_b}{k_b}(\hat{x}_b)^{m_b-k_b}\right) \\
&\times(P_{\vec{\alpha}+\vec{k}-\vec{e}_a}-P_{\vec{\alpha}-\vec{e}_a}P_k)\Bigg)\Bigg)\Bigg)dt \\
&-(dy-\hat{h}dt)'R^{-1}\Bigg(\sum_{|\vec{m}|_1\le M_h}\sum_{\vec{0}\le\vec{k}\le\vec{m}}[h_{i;\vec{m}}]_{n\times 1} \\
&\times\left(\prod_{b=1}^{n}\binom{m_b}{k_b}(\hat{x}_b)^{m_b-k_b}\right) \\
&\cdot\Bigg(\sum_{a=1}^{n}\alpha_a P_{\vec{k}+\vec{e}_a}P_{\vec{\alpha}-\vec{e}_a}-P_{\vec{\alpha}+\vec{k}}+P_{\vec{\alpha}}P_{\vec{k}}\Bigg)\Bigg).
\end{aligned}
\tag{11}
$$

With (6) and (11) in hand, we are ready to propose our new suboptimal method. Our idea is to cleverly impose some conditions to eliminate the terms $P_{\vec{\alpha}}$ in (6) and (11), for $|\vec{\alpha}|_\infty > |\vec{N}|_\infty$, for some given truncation $\vec{N}$, such that the equations of $\hat{x}_t$ and $P_{\vec{\alpha}}$, $\vec{\alpha}\le\vec{N}$, form a closed system. Thus, it is solvable and provides, generally speaking, more accurate approximation than its first order approximation—EKF.

We motivate by observing the last term of (11) for $\vec{\alpha} > \vec{e}_i$, for some $1\le i\le n$. That is, we exclude two trivial cases: (a) $P_{\vec{e}_i}=0$, for some $1\le i\le n$; (b) $P_{\vec{0}}=1$. It turns out that the last term vanishes if we impose the condition that

$$
P_{\vec{\alpha}+\vec{k}}=\sum_{a=1}^{n}\alpha_a P_{\vec{k}+\vec{e}_a}P_{\vec{\alpha}-\vec{e}_a}+P_{\vec{\alpha}}P_{\vec{k}}. \tag{12}
$$

Notice that $P_{\vec{\alpha}-\vec{e}_i}$, $P_{\vec{k}+\vec{e}_i}$, $P_{\vec{\alpha}}$, and $P_{\vec{k}}$ on the right-hand side of (12) are of lower order of $P_{\vec{\alpha}+\vec{k}}$.

Let us state our conditions more precisely. Given the truncation mode $\vec{N} > \vec{e}_i$, for some $1\le i\le n$, we shall form a closed system of equations for $\hat{x}_{it}$, $1\le i\le n$, and $P_{\vec{\alpha}}$, $\vec{\alpha}\le\vec{N}$. For arbitrary $\vec{\alpha} > \vec{e}_i$, for some $1\le i\le n$, there are three cases:

Case 1: $\vec{\alpha}\le\vec{N}$. Keep as it is, i.e. $P_{\vec{\alpha}}$;

Case 2: There exist $1\le i\ne j\le n$ such that $\alpha_i\le N_i$ and $\alpha_j > N_j$. We impose the condition (12) to $P_{\vec{\alpha}}=P_{\vec{\beta}+\vec{k}}$, where $\vec{\beta}=\min\{\vec{\alpha},\vec{N}\}$ and $\vec{k}=\vec{\alpha}-\vec{\beta}$;

Case 3: $\vec{\alpha} > \vec{N}$. Condition (12) is imposed to $P_{\vec{\alpha}}=P_{\vec{N}+\vec{k}}$, where$\vec{k}=\vec{\alpha}-\vec{N}$.

REMARK 3.3 Given any $\vec{\alpha}$ in Case 2 or 3, we shall impose the condition accordingly until it reduces to the combination of $P_{\vec{l}}$s, where all $\vec{l}$s belong to Case 1. Hence, the condition (12) may be imposed more than once to reduce certain $P_{\vec{\alpha}}$ in Case 2 or 3 to Case 1.

ALGORITHM OF OUR METHOD For continuous filtering problem given by system (1) with $f_i(x,t)$, $g_{ij}(x,t)$, and $h_i(x,t)$ approximated by (3)–(5), then a closed system of equations of $\hat{x}_i$, $1\le i\le n$, and $P_{\vec{\alpha}}$, $\vec{\alpha}\le\vec{N}$ is derived, if the condition (12) is imposed accordingly. Specifically, the closed system of the equations is given by: (6) for conditional mean $\hat{x}_i$, $1\le i\le n$; SDE (11) for $P_{\vec{\alpha}}$, for $\vec{\alpha} < \vec{N}$; ordinary differential (11) for $P_{\vec{N}}$ (the last term of

(11) vanishes here) and all the $P_{\vec{\alpha}}$ with $\vec{\alpha}$ in Case 2 or 3 are properly reduced to $P_{\vec{\alpha}}$, α in Case 1 by condition (12).

REMARK 3.4 By examining term-by-term in (6) and (11) with $|\vec{\alpha}|_1 = 2$, we see that when M_f, M_g, and $M_h \leq 1$, they form a closed system under the condition (12), which yields exactly the Kalman-Bucy filter. Indeed, if $f(x, t) = F(t)x$, $g(x, t) = G(t)$, and $h(x, t) = H(t)x$ in (1) for arbitrary $n \geq 1$, and the condition (12) is imposed, then our method gives

$$\begin{cases} d\hat{x} = F\hat{x}dt + PH^T R^{-1}(dy - H\hat{x}dt) \\ \frac{dP}{dt} = FP + PF^T + gQg^T - PH^T R^{-1} HP, \end{cases}$$

where $\hat{x} = [\hat{x}_1, \cdots, \hat{x}_n]$, $P = [P_{\vec{k}}]_{|\vec{k}|_\infty = 1}$.

REMARK 3.5 When $n = 1$, the lower bounds for some P_ks, $k \geq 2$, can be obtained by Jensen's inequality and Hölder's inequality, see details in Lemma 3.6 below. These lower bounds will be used to check the reasonableness of the conditions (12) imposed in cubic sensor problem in the next section.

LEMMA 3.6 (LOWER BOUND OF $P_k s$) Let $P_k = E^t[(x - \hat{x})^k]$, with convention that $P_0 = 1$, we have

1) $P_k \geq P_l^{\frac{k}{l}}$, for all $k \geq l \geq 1$ and k, $\frac{k}{l}$ are even integers greater than 2;

2) If k, l and $\frac{(k-l)p}{1-p}$ are all even integers, then $P_k \leq P_{lp}^{\frac{1}{p}} P_{\frac{(k-l)p}{p-1}}^{1-\frac{1}{p}}$, where $p \geq 1$, for all $k \geq l \geq 0$.

PROOF 1) It is trivial to see that when $k = l$ the equality holds. So let us assume that $k > l$ and look at P_{2k}:

$$\begin{aligned} P_k &= \int_{\mathbb{R}} (x - \hat{x})^k p(x|Y_t)dx \geq \int_{\mathbb{R}} \left[(x - \hat{x})^l p(x|Y_t)\right]^{\frac{k}{l}} dx \\ &\geq \left[\int_{\mathbb{R}} (x - \hat{x})^l p(x|Y_t)\right]^{\frac{k}{l}} dx = P_l^{\frac{k}{l}}, \end{aligned}$$

as long as $k \geq l \geq 1$, where the first inequality is due to the fact that $0 \leq p(x|Y_t) \leq 1$ and the second one follows from Jensen's inequality. It is Jensen's inequality that requires that $\frac{k}{l}$ is an even integer greater than 2, so that $x^{\frac{k}{l}}$ is convex in $\mathbb{R}$.

2) Similar to before, we have

$$\begin{aligned} P_k &= \int_{\mathbb{R}} (x - \hat{x})^k p(x|Y_t)dx \\ &= \int_{\mathbb{R}} (x - \hat{x})^{l+(k-l)} p(x|Y_t)^{m+(1-m)} dx \\ &\leq \left(\int_{\mathbb{R}} \left[(x - \hat{x})^l p(x|Y_t)^m\right]^p dx\right)^{\frac{1}{p}} \\ &\quad \times \left(\int_{\mathbb{R}} \left[(x - \hat{x})^{k-l} p(x|Y_t)^{1-m}\right]^{\frac{p}{p-1}} dx\right)^{1-\frac{1}{p}}, \end{aligned}$$

for all $p \geq 1$ and $0 \leq l \leq k$. The conclusion follows by letting $mp = 1$. ■

REMARK 3.7 Lemma 3.6 indicates that, in general, the moment sequence P_ks satisfy the following lower bounds: $P_4 \geq P_2^2$ (by 1)); $P_6 \geq P_2^3$ (by 1)) or $P_6 \geq \frac{P_4^2}{P_2}$ (by 2)), and etc. The lower bounds for $P_{\vec{k}}s$ with $n \geq 2$ are not clear [17].

IV. NUMERICAL EXPERIMENTS

In this section, we shall illustrate our method applied to two different filtering problems: cubic sensor problem and a polynomial filtering problem with two-dimensional state. In the cubic sensor problem, we compare our method with $N = 2$, 3 with EKF, and PF with 50 particles. Further, we formulate and implement our method to a polynomial filtering problem with two-dimensional state. The numerical result has been also compared with EKF, UKF, and EnKF with 20 ensembles.

A. Cubic Sensor Problem

This problem is modeled by SDE (1) with $f(x, t) = 0$, $g(x, t) = 1$, and $h(x, t) = x^3$, which has been shown rigorously that it is essentially infinite-dimensional in [18] and has been studied by many authors, refer to [19–21]. In order to get a fair comparison with EKF in computational complexity, we first propose to pick $N = 2$. Intuitively, the larger N is, the more accurate approximation is obtained for the state. Hence, we also pick $N = 3$ in our method for comparison.

Notice that $M_h = 3$, $M_f = M_g = 0$. On the right-hand sides of (6) and (11) with $\alpha \leq 2$, P_3-P_5 show up and need to be reduced to some functions of P_2, $P_1 = 0$, and $P_0 = 1$. The conditions we imposed are:

$$\begin{aligned} P_3 &= P_{2+1} \overset{12}{=} 2P_2 P_1 + P_1 P_2 = 3P_1 P_2 = 0; \\ P_4 &= P_{2+2} \overset{12}{=} 2P_3 P_1 + P_2 P_2 = P_2^2; \\ P_5 &= P_{2+3} \overset{12}{=} 2P_4 P_1 + P_2 P_3 = P_2 P_3 \overset{13}{=} 0. \end{aligned} \tag{13}$$

The condition on P_4 satisfies the lower bound in Remark 3.7. Our method for $\hat{x}_t$ and P_2 gives

$$\begin{cases} d\hat{x}_t = \frac{1}{R}(P_2^2 + 3P_2(\hat{x}_t^2))(dy - (3P_2\hat{x}_t + (\hat{x}_t)^3)dt) \\ \frac{dP_2}{dt} = 1 - \frac{1}{R}(P_2^2 + 3P_2(\hat{x}_t)^2)^2 \end{cases} \tag{14}$$

When choosing $N = 3$ in our method, the conditions imposed are:

$$\begin{aligned} P_4 &= P_{3+1} \overset{12}{=} 3P_2^2 + P_1 P_3 = 3P_2^2; \\ P_5 &= P_{3+2} \overset{12}{=} 3P_2 P_3 + P_2 P_3 = 4P_2 P_3; \\ P_6 &= P_{3+3} \overset{13}{=} 3P_2 P_4 + P_3^2 = 9P_2^3 + P_3^2. \end{aligned} \tag{15}$$

Again from Remark 3.7, the condition on P_4, P_6 are also reasonable, in the sense that $P_4 \geq P_2^2$ and $P_6 \geq \frac{P_4^2}{P_2} = \frac{(3P_2^2)^2}{P_2} = 9P_2^3$. The SDE given by our method for $\hat{x}_t$, P_2, and P_3 is:

$$\begin{cases} d\hat{x}_t = \frac{1}{R}[dy - (\hat{x}_t^3 + 3\hat{x}_t P_2 + P_3)dt] \\ \qquad \cdot(3\hat{x}_t^2 P_2 + 3\hat{x}_t P_3 + 3P_2^2) \\ \frac{dP_2}{dt} = 1 - \frac{1}{R}(3\hat{x}_t^2 P_2 + 3\hat{x}_t P_3 + 3P_2^2)^2 \\ \frac{dP_3}{dt} = -\frac{3}{R}(3\hat{x}_t^2 P_2 + 3\hat{x}_t P_3 + 3P_2^2) \\ \qquad \cdot(3\hat{x}_t^2 P_3 + 6\hat{x}_t P_2^2 + 3P_2 P_3) \end{cases} \tag{16}$$

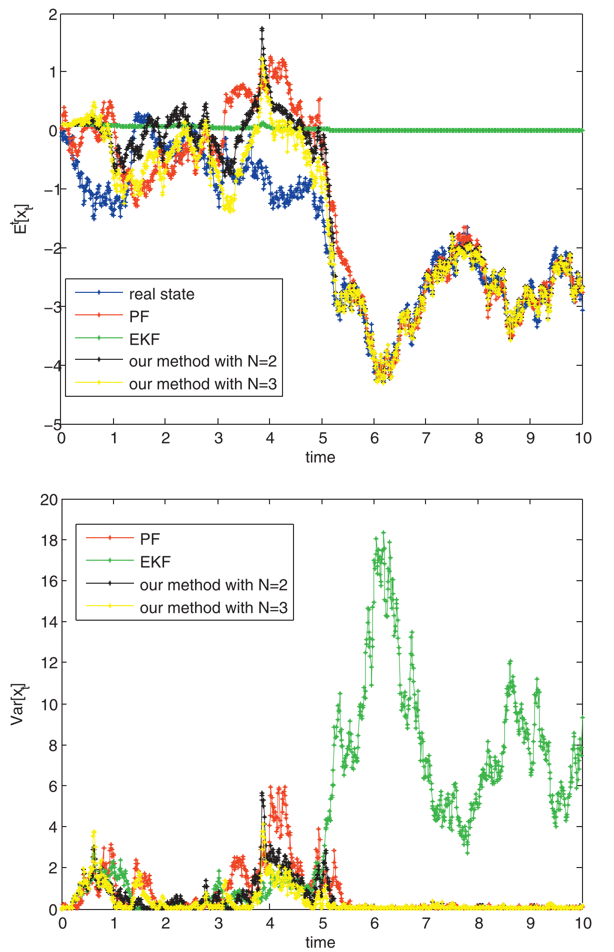


Fig. 1. Our method with $N = 2, 3$ for the cubic sensor problem are compared with the EKF and the PF with 50 particles. Left: the averaged mean vs. time; right: the averaged variance vs. time.

We randomly generate 100 sample paths (except those EKF explodes before T) with $Q = R = 1$ and $P_0 = 0.01$, and apply EKF, PF with 50 particles, our method with $N = 2$ (14) and $N = 3$ (16) to estimate the real state. The PF used in our experiment is the SIR algorithm; see Algorithm 4, [22]. It is worth noting that there has been much progress in PF since the SIR algorithm, including: regularised PFs [23], auxiliary PFs [24], particle flow filters [25], Gaussian PFs [26], transport PFs [27], various Markov Chain Monte Carlo methods (e.g., Metropolis adjusted Langevin or MALA, hybrid Monte Carlo, Girolami's geodesic flow on Riemannian manifolds, etc.). The SDEs of EKF and our methods are numerically solved by Euler-Maruyama scheme [28]. The total experimental time is $T = 10$ and the time step is $dt = 0.01$. The averaged mean and variance of the 100 experiments using EKF, PF, and our methods have been displayed in Fig. 1. The figure shows that our method with $N = 3$ is superior than the other three. The variance of the estimation errors and the average CPU time has been list in Table I.

To explain why in Table I the number of particles is chosen to be 500 in PF, we experiment the cubic sensor problem by generating the sample path using

TABLE I
Variance of the Estimation Errors and Average CPU Time of Different Filters Applied to the Cubic Sensor Problem

Filters	Variance of the Errors	Average CPU Time
PF with 500 particles	0.4566	4.146493 s
EKF	4.4487	0.002505 s
our method with $N = 2$	0.4562	0.002325 s
our method with $N = 3$	0.3425	0.003405 s

TABLE II
Number of Particles vs. Variance of Estimation Error

Number of Particles	Variance of Estimation Errors	CPU Time
50	0.5167	0.465100 s
100	0.4246	0.909719 s
200	0.4493	2.642290 s
500	0.3596	4.251382 s
1,000	0.4765	8.555768 s
5,000	0.4461	37.203790 s

randn('*state*',*100*), with $T = 10$ and $dt = 0.01$. The performance is measured by variance of estimation errors. In Table II, we display the errors and the CPU times with different number of particles from 50 to 5000. It shows that using 500 particles the PF accuracy is roughly the same as our method. Presumably, this is the optimal accuracy, which explains why the performance stops to be improved by using more particles.

REMARK 4.1 The condition (12) on $P_{\tilde{\alpha}}$ can't be shown rigorously. It is just like no one can show that the truncation (conventionally operation to form a close system) yields the theoretically best approximation of $P_{\tilde{\alpha}}$.

In the sequel, we shall use the global method proposed in [29–30] to numerically compute the P_ks of cubic sensor problem. This investigation will give us some indication on the reasonableness of our condition (12). [29–30] introduced a method to directly approximate the conditional density function $\rho(x, t)$, and then we can obtain the approximate higher central moment of the states by

$$P_l = E^t[(x - \hat{x})^l] = \int_{\mathbb{R}} \frac{(x - \hat{x})^l \rho(x, t)}{\int_{\mathbb{R}} \rho(x, t) dx} dx,$$

where $l \geq 2$, for the one-dimensional state. We apply the method in [29–30] with appropriately chosen parameters ($\alpha = 2.5$, truncation modes $N_f = 45$) to 10 randomly generated real states. All the real states are generated with $Q = R = 1$ and the initial density function is assumed to be $u_0(x) = e^{-\frac{x^2}{2}}$. The total experimental time is $T = 10$, and time step is $dt = 0.001$. The approximate higher central moments are computed numerically by Gaussian-Hermite quadrature rule. The averaged higher central moments P_2–P_6 obtained by method in [29–30] have been plotted in Fig. 2. It indicates that we probably should impose $P_{2k+1} \approx 0$ and $P_{2k} \neq 0$, which matches the condition (13) and (15).

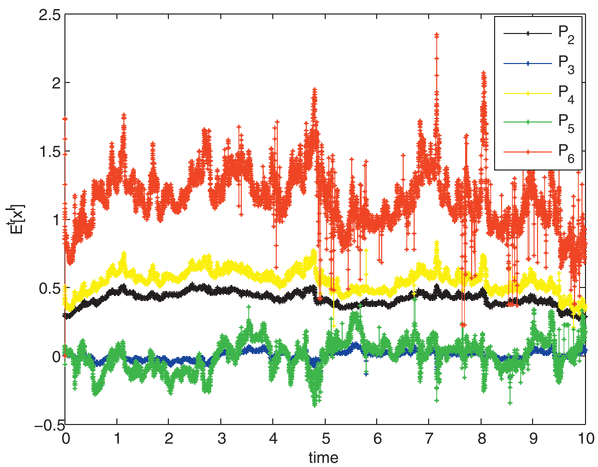


Fig. 2. The averaged higher central moments for cubic sensor problem are displayed.

B. Polynomial Filtering Problem With Two-Dimensional State

In this subsection, we shall illustrate our method formulated for polynomial filtering problems of higher dimensional states. Let us take the following example:

$$\begin{cases} f_1 = 0 \\ f_2 = x_1^2 \end{cases}, \quad \begin{cases} h_1 = x_1 x_2 \\ h_2 = x_2^2 \end{cases}, \quad g = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}, \quad R = I_2, \tag{17}$$

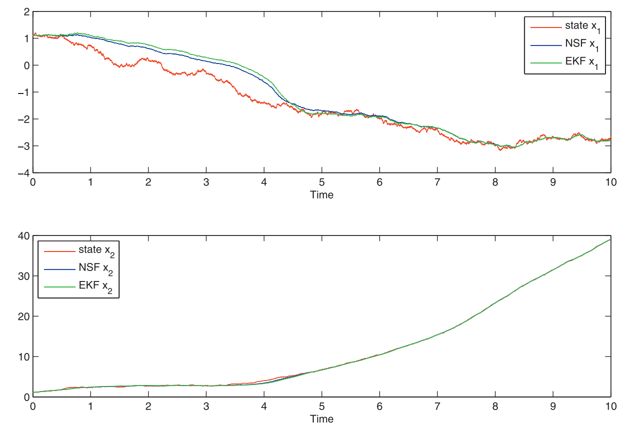
and the initial state

$$\begin{pmatrix} x_{10} \\ x_{20} \end{pmatrix} \sim N\left(\begin{pmatrix} 1.1 \\ 1.1 \end{pmatrix}, \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix} \right). \tag{18}$$

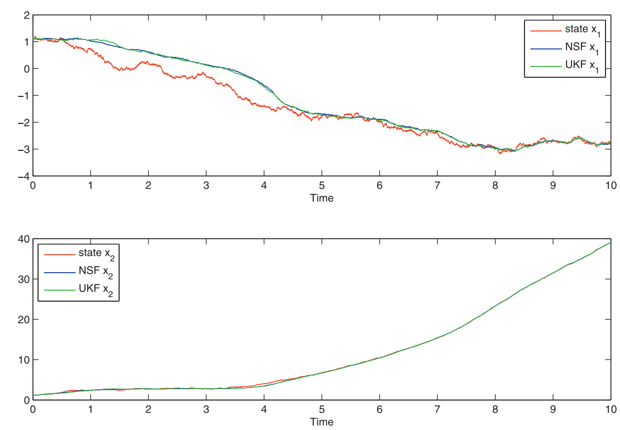
Let us choose $\vec{N} = (2, 2)$ in our method. Notice that $M_f = 2$ and $M_h = 2$. Observing the right-hand side of (6) and (11) for $P_{\vec{\alpha}}$ with $\vec{\alpha} \leq \vec{N}$, it contains all $P_{\vec{\alpha}}$, $\vec{\alpha} \leq \vec{N} + \vec{k}$, for $|\vec{k}|_1 \leq M_h$. We need to reduce all $P_{\vec{\alpha}}$, $\vec{\alpha}$ in case 2 or 3 by condition (12).

$$\begin{aligned}
P_{30} &= P_{(2,0)+(1,0)} \overset{12}{=} 2P_{20}P_{10} + P_{20}P_{10} \\
&= 3P_{20}P_{10} = 0; \\
P_{31} &= P_{(2,1)+(1,0)} \overset{12}{=} 2P_{20}P_{11} + P_{11}P_{20} + P_{21}P_{10} \\
&= 3P_{20}P_{11}; \\
P_{32} &= P_{(2,2)+(1,0)} \overset{12}{=} 2P_{20}P_{12} + 2P_{11}P_{21} + P_{22}P_{10} \\
&= 2P_{20}P_{12} + 2P_{11}P_{21}; \\
P_{33} &= P_{(2,2)+(1,1)} \overset{12}{=} 2P_{12}P_{21} + 2P_{21}P_{12} + P_{22}P_{11} \\
&= P_{12}P_{21} + P_{22}P_{11}; \\
P_{40} &= P_{(2,0)+(2,0)} \overset{12}{=} 2P_{30}P_{10} + P_{20}^2 = P_{20}^2; \\
P_{41} &= P_{(2,1)+(2,0)} \overset{12}{=} 2P_{30}P_{11} + P_{21}P_{20} + P_{21}P_{20} \overset{19}{=} 2P_{20}P_{21}; \\
P_{42} &= P_{(2,2)+(2,0)} \overset{12}{=} 2P_{12}P_{30} + 2P_{21}P_{21} \\
&\quad + P_{22}P_{20} \overset{19}{=} 2P_{21}^2 + P_{20}P_{22}.
\end{aligned} \tag{19}$$

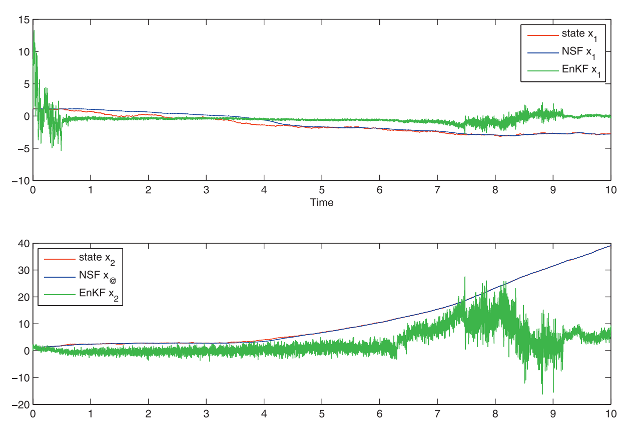
Similar arguments could be used to obtain $P_{03} = 0$, $P_{13} = 3P_{02}P_{11}$, $P_{23} = 2P_{02}P_{21} + 2P_{11}P_{21}$, $P_{04} = P_{02}^2$, $P_{14} = 2P_{02}P_{12}$, and $P_{24} = 2P_{12}^2 + P_{02}P_{22}$. According to (6) and (11), our method yields a SDE of $\hat{x}_1$, $\hat{x}_2$, P_{02}, P_{11},



IV.3.a: *NSF and EKF.*



IV.3.b: *NSF and UKF.*



IV.3.c: *NSF and EnKF with* 20 *ensembles.*

Fig. 3. NSF compared with EKF, UKF, and EnKF with 20 ensembles are displayed for the two-dimensional polynomial filtering problem (17), (18). The upper one in each subfigure is the trajectory of $\hat{x}_1$, while the lower one is that of $\hat{x}_2$.

P_{20}, P_{12}, P_{21} and P_{22}. We don't write down the lengthy expression here due to the page limitation.

Numerical results for this example are displayed in Fig. 3. In this example, we generate 20 sample paths randomly. The total experimental time is $T = 10$, and the time step is $dt = 0.001$. The figures are the average of 20 runs. One can see that our method tracks as well as EKF and UKF. But EnKF with 20 ensembles does not perform very well. As to the efficiency, our method takes 15.4 s while it costs 163.4 s for UKF to obtain the similar result.

V. CONCLUSIONS

In this paper, given a truncation $\vec{N}$, starting from (11) for $P_{\vec{\alpha}}$, we construct our method by imposing some conditions (12) to reduce all the higher order central moments to the combination of the lower order ones $P_{\vec{\alpha}}$, $\vec{\alpha} \leq \vec{N}$. After the reduction, our method arrives at a closed system of (6) for $\hat{x}_{it}$, $1 \leq i \leq n$ and (11) for $P_{\vec{\alpha}}$, $\vec{\alpha} \leq \vec{N}$. This is completely new and different from the conventional operation–truncation. Since no one can show the truncation yields the best approximation, our procedure provides another reasonable way to form a closed system. Our method is a natural generalization of EKF. It is also more flexible by choosing the truncation $\vec{N}$ according to the desired accuracy and the demand of computational complexity. The imposed condition (12) in our method satisfies the lower bounds of P_ks, and it is justified numerically for the cubic sensor problem by using the higher central moments obtained from Yau-Yau's method [29]. Our method has also been formulated and implemented for the filtering problems with a two-dimensional state. Numerical results verify that our method works in nearly perfect agreement with theory.

XUE LUO, Senior Member, IEEE
Beihang University
Beijing, P. R. China

YANG JIAO
University of Illinois at Chicago
Chicago, IL, USA

WEN-LIN CHIOU
Fu-Jen University
New Taipei City, Taiwan, ROC

STEPHEN S.-T. YAU, Fellow, IEEE
Tsinghua University
Beijing, P. R. China
yau@uic.edu

REFERENCES

[1] Jazwinski, A. H.
Stochastic processes and filtering theory. New York: Academic, 1970.

[2] Kalman, R. E.
A new approach to linear filtering and prediction problems.
Transactions of the American Society of Mechanical Engineers, Series D, Journal of Basic Engineering, **82**, (1960), 35–45.

[3] Kalman, R. E., and Bucy, R. S.
New results in linear filtering and prediction theory.
Transactions of the American Society of Mechanical Engineers, Series D, Journal of Basic Engineering, **83**, (1961), 95–108.

[4] Luo, X.
On recent advance of nonlinear filtering theory: Emphases on global approaches.
Pure Applied Mathematics Quarterly, **10**, 4 (2014), 685–721; reprinted in *International Congress of Chinese Mathematics Notices*, **3**, 1 (2015), 43–57.

[5] Wan, E. A., and Van der Merwe, R.
The unscented Kalman filter for nonlinear estimation.
In *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (AS-SPCC)*, 2000, 153–158.

[6] Gillijns, S., Barrero Mendoza, O., Chandrasekar, J., De Moor, B. L. R., Bernstein, D. S., and Ridley, A.
What is the ensemble Kalman filter and how well does it work?
In *Proceedings of the IEEE of American Control Conference (ACC)*, 2006, 4,448–4,453.

[7] Crisan, D., and Lyons, T.
A particle approximation of the solution of the Kusher-Stratonovich equation.
Probability Theory and Related Fields, **115**, (1999), 549–578.

[8] Kan, R.
From moments of sum to moments of product.
Journal of Multivariate Analysis, **99**, 3 (2008), 542–554.

[9] Budhiraja, A., Chen, L., and Lee, C.
A survey of numerical methods for nonlinear filtering problems.
Physics D, **230**, (2007), 27–36.

[10] Germani, A., Manes, C., and Palumbo, P.
Filtering of differential nonlinear systems via a Carleman approximation approach.
In *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference*, Seville, Spain, Dec. 12–15, 2005, 5,917–5,922.

[11] Germani, A., Manes, C., and Palumbo, P.
Filtering of stochastic nonlinear differential systems via a Carleman approximation approach.
IEEE Transactions on Automatic Control, **52**, 11 (2007), 2,166–2,172.

[12] Carravetta, F., Germani, A., and Shuakayev, M.
A new suboptimal approach to the filtering problem for bilinear stochastic differential systems.
SIAM Journal on Control and Optimization, **38**, 4 (2000), 1,171–1,203.

[13] Luo, X., and Yau, S. S.-T.
The suboptimal nonlinear filtering with augmented states via probabilists' Hermite polynomials.
To be published, 2015.

[14] Majji, M., Junkins, J. L., and Turner, J. D.
Jth moment extended Kalman filtering for estimation of nonlinear dynamic systems.
In *Proceedings of the AIAA Guidance, Navigation and Control Conference and Exhibit*, Aug. 2008.

[15] Willsky, A. S.
Dynamical systems defined on groups: Structural properties and estimation.
PhD thesis, MIT, MA, 1973.

[16] Luo, X., Jiao, Y., Chiou, W.-L., and Yau, S. S.-T.
A novel suboptimal method for solving polynomial filtering problems.
Automatica Journal of the International Federation of Automatic Control, **62**, (2015), 26–31.

[17] Kushner, H.
Approximations to optimal nonlinear filters.
IEEE Transactions on Automatic Control, **AC-12**, 5 (1967), 546–556.

[18] Hazewinkel, M., Marcus, S. I., and Sussmann, H. J.
Nonexistence of finite dimensional filters for conditional statistics of the cubic sensor problem.
Systems and Control Letters, **3**, (1983), 331–340.

[19] Bucy, R. S., and Pages, J.
A priori error bounds for the cubic sensor problem.
IEEE Transactions on Automatic Control, **23**, (1978), 88–91.

[20] Ocone, D.
Probability densities for conditional statistics in the cubic sensor problem.
Mathematics of Control, Signals, and Systems, **1**, (1988), 183–202.

[21] Steiberg, Y., Bobrowsky, B. Z., and Schuss, Z.
On the optimal filtering problem for the cubic sensor.
Circuits, Systems, and Signal Processing, **7**, (1988), 381–408.

[22] Arulampalam, M. S., Maskell, S., Gordon, N., and Clapp, T.
A tutorial on particle filters for online nonlinear non-Gaussian Bayesian tracking.
IEEE Transactions on Signal Processing, **50**, 2 (2002), 174–188.

[23] Musso, C., Oudjane, N., and Le Gland, F. Improving regularised particle filters. Sequential Monte Carlo methods in practice. *Statistics for Engineering and Information Science*, New York: Springer, 2001, pp. 247–271.

[24] Pitt, M. K., and Shephard, N.
Filtering via simulation: Auxiliary particle filters.
Journal of the American Statistical Association, **94**, 446 (1999), 590–591.

[25] Daum, F., and Huang, J.
Particle flow for nonlinear filters.
In *Proceedings of the IEEE International Conference on Acoustics*, **7698**, 2 (2011), 5,920–5,923.

[26] Bunch, P., and Godsill, S.
Particle filtering with progressive Gaussian approximations to the optimal importance density.
In *Proceedings of the IEEE 5th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Dec. 2013, 360–363.

[27] Reich, S. A.
Gaussian-mixture ensemble transform filter.
Quarterly Journal of the Royal Meteorological Society, **138**, 662 (2011), 222–233.

[28] Higham, D. J.
An algorithmic introduction to numerical simulation of stochastic differential equations.
SIAM Review, **43**, (2001), 525–546.

[29] Luo, X., and Yau, S. S.-T.
Complete real time solution of the general nonlinear filtering problem without memory.
IEEE Transactions on Automatic Control, **58**, 10 (2013), 2,563–2,578.

[30] Luo, X., and Yau, S. S.-T.
Hermite spectral method to 1D forward Kolmogorov equation and its application to nonlinear filtering problems.
IEEE Transactions on Automatic Control, **58**, 10 (2013), 2,495–2,507.