ON RECENT ADVANCE OF NONLINEAR FILTERING THEORY: EMPHASES ON GLOBAL APPROACHES

Xue Luo

Abstract

The surveys in the field of nonlinear filtering (NLF) are enumerous. Most of them are application-oriented and served as the tutorials for the practioners. The local approaches, including Kalman filter and its invariants, have already been studied from various point of views, due to its off-the-shelf implementation and wide applications. However, it cannot give good estimation of the states in highly nonlinear system or with non-Gaussian initial conditional density functions. Moreover, while the local methods only approximate the mean and variance, the global ones seek the way to directly obtain the conditional density function of the states. Consequently, all the statistical information is acquired. In this survey, we shall briefly go through the local approaches and put emphases on the existing three major global approaches: finite-dimensional NLF, sequential Monte Carlo methods (particle filter) and the Yau-Yau's on- and off-line solver of Duncan-Mortensen-Zakai's equation [**75**]. The discussions are mainly from the mathematical point of view.

1. Introduction

The field of nonlinear filtering (NLF) has its origin from tracking and signal processing problems. Yet, the underlying formulation is so general and ubiquitous that it can be widely applied to various complex dynamical phenomenon modelled by stochastic processes. The aim of filtering is to obtain good estimates of the states in the stochastic dynamical system recursively in time, based on the noisy observations of the states. The states are also called signals. The states or signals represent all kinds of quantities in various applications. For example, the states in the tracking problem [**51**] are the moving target's position and velocity, and the observations are some nonlinear functions of the states corrupted by noise. The states in volatility calibration problem is the underlying volatility process while the observations are the security and derivative prices [**28**]. The signal process in the ion channel kinetics problem for nerve cells is the underlying molecular dynamics, while the observations are the channel conductances [**53**]. In the atmospheric data assimilation problems [**27**], the state refers to the location of a hurricane and the observations may be the measurements of the wind speed at various locations.

The study of stochastic filtering has a long story dated back to 1940s. It was first investigated in the pioneering work by Wiener [**67**] and Kolmogorov [**48**]. The most influential work in filtering theory is the classic Kalman filter (KF) published in 1960 [**45**] and subsequent Kalman-Bucy filter published one year later [**46**]. After the discovery of KF, its variants and itself have been dominated the field of filtering theory in signal processing and control area for more than half century. Till now, KF and its derivatives are still widely applied in various engineering and scientific problems, including tracking, communications, machine learning, economics, finance and etc. However, the KF performs poorly once either the linear or the Gaussian assumption is violated [**2**]. Consequently, the mathematicians and engineers are urged to pursue a computationally efficient, recursive optimal solution applicable to the more general NLF problems. Unfortunately, such algorithm only exists for the limited class of dynamic systems, say Beneš filter [**8**], Yau filter [**14**] and etc. It motivates the researches on the suboptimal solutions of NLF, which can be classified into two categories: the local and global approaches. The local ones approximate the posterior density function by some

Mathematics Subject Classification. 60G35,60F05,82C31,93E11.

Key words and phrases. Nonlinear filtering, global approach, Bayesian framework, Duncan-Mortensen-Zakai's equation.

particular form, say Gaussian or mixture of Gaussian; while the posterior density function in global approaches are directly computed without any assumptions on its type. More extensive discussions on the local and global approaches can be found in section 3.1 and 3.2, respectively.

From the formulation of NLF, Bayesian theory is no doubt one of the main tools, which is the most commonly used method for the study of the dynamic systems. Bayesian theory was originally discovered by [**7**] in 1763. It reveals the fundamental probability law governing the process of logical inference. However, it didn't receive much attention at that time until the re-discovery of its modern form by Laplace in "Théorie analytique des probailités". One of the important branches in statistics is the Bayesian statistics to statistical inference. Not surprisingly, Bayesian theory was also investigated in the field of filtering theory. Starting from the KF, although it is first developped by the orthogonal projection method [**45**, **40**], it has very nice interpretation of Bayesian framework. It is Ho and Lee who first explored the iterative Bayesian estimation. They specified in [**35**] the principle and procedure of Bayesian framework in the context of filtering. In general, the Bayesian filtering requires a dynamic state-space model (2.2), which consists of two processes: one describes the evolution of a hidden state of the system, while the other one is the observation process related to the states and corrupted with noises. In the Bayesian approach, the posterior density of the states, obtained from Bayes' theorem, provides a complete statistical description of the state variable at that time [**1**]. The procedure of Bayesian filtering consists of prediction-correction recursions. All sorts of variants of KF and the sequential Monte Carlo methods (particle filters) belong to this framework.

Besides the Bayesian framework, the conditional density function of the states can also be obtained by numerically solving the so-called Kusher's or Duncan-Mortensen-Zakai's equation. It is shown in [**50**] that the conditional density $p(\mathbf{x}_t|\mathcal{Y}_t)$ of the states $\mathbf{x}_t$ based on the observation history $\mathcal{Y}_t$ satisfies an Itô stochastic differential equation (SDE), which is called Kusher's equation. After the change of measure, the unnormalized conditional density $\pi(\mathbf{x}_t|\mathcal{Y}_t)$ satisfies a linear Itô SDE, so-called DMZ equation [**24**, **59**, **77**]. Apparently, the DMZ equation is the more preferable one. And the solution to the Kusher's equation $p(\mathbf{x}_t|\mathcal{Y}_t)$ and that to the DMZ equation $\pi(\mathbf{x}_t|\mathcal{Y}_t)$ is one-to-one correspondence. Detailed discussions can be found in section 2.2. Numerous efforts have been devoted in the past to solve DMZ equation for a general dynamic systems. We refer the interested readers to the survey [**32**] and references therein. By then, the algorithms are neither computational efficient nor recursive. In 2008, Yau and Yau [**75**] made a major breakthrough, due to a key observation (see Proposition 3.1) so that the heavy computation of solving the partial differential equation (PDE) can be pre-computed, stored and updated by synchronizing with the observations on-line. Thus, the real-time manner of the algorithm is foreseeable. The convergence of their algorithm has been rigorously shown when the drift function, the diffusion term and the observation function are time-invariant. We refer this method as Yau-Yau's on- and off-line algorithm, and Yau-Yau's method for short, in this survey. Recently, Yau and the author validated it also for the time-varying system, and numerically verified the real-time performance when the state is of one dimension [**56**, **57**]. More recently, Yueh et al. [**76**] present an efficient algorithm of Yau-Yau's method, and numerical simulations with two-dimensional states are performed well. In private communication, they claimed the feasibility and efficiency of Yau-Yau's method in even higher dimensions, say for the state with 6 dimension, which is a major breakthrough in this direction. We briefly summerize the idea of Yau-Yau's method in this survey.

This survey is aim to present various approaches studied in the literature with the emphases on the recent advance of the global approaches. We note that it is by no means exhaustive, in particular the local approaches, i.e. the discussion of some variants of KF, say ensemble Kalman filter (cf. [**3**, **36**]) is not mentioned in section 3.1, which may be effecient in certain extremely high-dimensional problems, for example the atmospheric data assimilation.

The paper is organized as follows. We present the general formulation of NLF problems in both discrete and continous-time versions. In section 2 we describe the stochastic filtering problem abstractly in two aspects: one is the Bayesian framework; the other one is the Kusher's and DMZ equations. Section 3 devotes to summarize local and global approaches of NLF. The KF, as the most influencial one, is re-derived from the viewpoint of DMZ equation, which provides a natural relation between the approaches based on Bayesian theory and DMZ equation. Following the KF, its

variants including extended Kalman filter (EKF), Gaussian sum filter (GSF) and unscented Kalman filter (UKF), etc are briefly presented. We emphasize the results of global approaches in section 3.2, where we display three of the kind: finite-dimensional filters, sequential Monte Carlo methods (particle filter) and the Yau-Yau's method. At last, we arrive the conclution and mention some possible future developments in section 4.

2. Stochastic filtering problem

The aim of the stochastic filtering is to obtain the "best" estimate of the state or the signal in some sense, where the state is modelled by a stochastic process or a random sequence, denoted as $\{X_t,\, t \geq 0\}$ or $\{X_k,\, k \in \mathbb{N}\}$. The state itself can't be measured directly, while certain measurements of the state can be obtained, denoted as $\{Y_t,\, t \geq 0\}$ or $\{Y_k,\, k \in \mathbb{N}\}$, which is another stochatic process or random sequence. The observation usually is a function of the state with some measurement noise $\{W_t,\, t \geq 0\}$ or $\{W_k,\, k \in \mathbb{N}\}$. If we are in the continuous-time case, we assume further that X_t, Y_t and all the other processes in the sequel are defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{Y}_t = \sigma(Y_s,\, s \in [0,t])$ be the filtration generated by the observation process Y_t up to time t, which contains all the information from the observation history up to time t. The filtering problem is to estimate X_t based on $\mathcal{Y}_t$, i.e. $E[X_t|\mathcal{Y}_t]$.

Let us describe the discrete-time stochastic filtering as the vector-valued SDE [**40**], which is commonly used in the point-based filter.

$$\begin{cases} \mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}) + \mathbf{w}_{k-1} \\ \mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k \end{cases}, \tag{2.1}$$

where the state $\mathbf{x}_k$ is the N_x-vector and the measurement $\mathbf{y}_k$ is the N_y-vector; $\mathbf{w}_{k-1}$ and $\mathbf{v}_k$ are independent white Gaussian process noise and measurement noise with the covariance $\mathbf{Q}_{k-1}$ and $\mathbf{R}_k$, respectively. The aim of discrete estimation problem is to estimate the state $\mathbf{x}_k$ based on $\mathbf{y}_{1:k}$, given certain realization of observations $\mathbf{y}_{1:k} := \{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_k\}$, .

Suppose $\{x_t\}$ is a Markov process with an infinitesimal generator, the state-space equations can be written in the form of Itô stochastic differential equation [**62**]:

$$\begin{cases} d\mathbf{x}_t = \mathbf{f}(t, \mathbf{x}_t)dt + \mathbf{G}(t, \mathbf{x}_t)d\mathbf{w}_t \\ d\mathbf{y}_t = \mathbf{h}(t, \mathbf{x}_t)dt + d\mathbf{v}_t \end{cases}, \tag{2.2}$$

where $\mathbf{f}(t, \mathbf{x}_t)$ is the drift term, $\mathbf{G}(t, \mathbf{x}_t)$ is the volatility or diffusion coefficient, and $\mathbf{h}(t, \mathbf{x}_t)$ is the observation function. The two noise processes $\{\mathbf{w}_t,\, t \geq 0\}$ and $\{\mathbf{v}_t,\, t \geq 0\}$ are Wiener processes, with $E[d\mathbf{w}_t d\mathbf{w}_t^T] = \mathbf{Q}_t dt$ and $E[d\mathbf{v}_t d\mathbf{v}_t^T] = \mathbf{R}_t dt$, $\mathbf{R}_t > 0$, respectively. $\mathbf{x}_t \in \mathbb{R}^{N_x}$ and $\mathbf{y}_t \in \mathbb{R}^{N_y}$, where N_x and N_y are the dimension of the states and observations, respectively.

2.1. Bayesian estimation framework. In this framework, we assume that

1) The state is a Markov process, i.e. $p(\mathbf{x}_k|\mathbf{x}_{1:k-1}) = p(\mathbf{x}_k|\mathbf{x}_{k-1})$;
2) The observations are independent of the given states, i.e. $\mathbf{y}_{1:k-1}$ are independent of $\mathbf{x}_k$.

Let $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ denote the probability density function (pdf) of $\mathbf{x}_k$ under the condition of the observations $\mathbf{y}_{1:k}$, then from the Bayes' rule, we have

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{y}_{1:k}) &\overset{Bayes'}{=} \frac{p(\mathbf{y}_{1:k}|\mathbf{x}_k)p(\mathbf{x}_k)}{p(\mathbf{y}_{1:k})} = \frac{p(\mathbf{y}_k, \mathbf{y}_{1:k-1}|\mathbf{x}_k)p(\mathbf{x}_k)}{p(\mathbf{y}_k, \mathbf{y}_{1:k-1})} = \frac{p(\mathbf{y}_k|\mathbf{y}_{1:k-1}, \mathbf{x}_k)p(\mathbf{y}_{1:k-1}|\mathbf{x}_k)p(\mathbf{x}_k)}{p(\mathbf{y}_k|\mathbf{y}_{1:k-1})p(\mathbf{y}_{1:k-1})} \\ &\overset{Bayes'}{=} \frac{p(\mathbf{y}_k|\mathbf{y}_{1:k-1}, \mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1})p(\mathbf{y}_{1:k-1})p(\mathbf{x}_k)}{p(\mathbf{y}_k|\mathbf{y}_{1:k-1})p(\mathbf{y}_{1:k-1})p(\mathbf{x}_k)} = \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1})}{p(\mathbf{y}_k|\mathbf{y}_{1:k-1})} \\ &= \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1})}{\int p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1})d\mathbf{x}_k} \end{aligned} \tag{2.3}$$

It is clear to see from (2.3) that the posterior pdf $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ is obtained by three terms: the prior pdf $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$, the likelihood function $p(\mathbf{y}_k|\mathbf{x}_k)$ and the denominator in (2.3).

The Bayesian filtering consists of recursive prediction and update procedures [**40**].

Prediction: Given the prior pdf $p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})$, the conditional pdf of $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ satisfies the Chapman-Kolmogorov equation:

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})d\mathbf{x}_{k-1}; \tag{2.4}$$

Update: When $\mathbf{y}_k$ is available, the posterior pdf $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ is given by (2.3), i.e.,

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1})}{\int p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1})d\mathbf{x}_k}. \tag{2.5}$$

In the general NLF problems, both prior and posterior conditional pdf can't be computed in the analytic form. Therefore, it is essential to approximate prior and likelihood functions in (2.3). As the consequence, a variety of local approaches have been developped by using different approximations.

2.2. Kusher's and Duncan-Mortensen-Zakai's equation. In the continuous-time case, we can reformulate the stochastic filtering problem, by considering the infinitesimal generator of the state process $\{\mathbf{x}_t,\, t \geq 0\}$:

$$\mathcal{L}(\circ) := \frac{1}{2}\sum_{i,j=1}^{N_x} (\mathbf{GQG}^T)_{ij}(t,\mathbf{x}_t)\frac{\partial^2(\circ)}{\partial x_i \partial x_j} + \sum_{i=1}^{N_x} f_i(t,\mathbf{x}_t)\frac{\partial(\circ)}{\partial x_i},$$

where f_i and x_i are the ith component of the vector-valued function $\mathbf{f}$ and the vector state $\mathbf{x}_t$, respectively. The question now can be interpreted as how to find a recursive or finite-dimensional method to compute the conditional pdf of $\mathbf{x}_t$ with the filtration $\mathcal{Y}_t$, i.e. $p(\mathbf{x}_t|\mathcal{Y}_t)$. It turns out that $p(\mathbf{x}_t|\mathcal{Y}_t)$ satisfies the following Kusher's equation (cf. [**50**]):

$$dp(\mathbf{x}_t|\mathcal{Y}_t) = \mathcal{L}^* p(\mathbf{x}_t|\mathcal{Y}_t)dt + p(\mathbf{x}_t|\mathcal{Y}_t)\mathbf{e}_t\Sigma_{v,t}^{-1}dt,$$

where $\mathcal{L}^*$ is the adjoint operator of $\mathcal{L}$, i.e.

$$\mathcal{L}^*(\circ) = \frac{1}{2}\sum_{i,j=1}^{N_x} \frac{\partial^2((\mathbf{GQG}^T)_{ij}\circ)}{\partial x_i \partial x_j} - \sum_{i=1}^{N_x}\frac{\partial(f_i\circ)}{\partial x_i}, \tag{2.6}$$

$\mathbf{e}_t$ is the innovation process

$$\mathbf{e}_t = \mathbf{y}_t - \int_0^t \mathbb{E}[\mathbf{h}(s,\mathbf{x}_s)|\mathcal{Y}_s]ds, \quad \mathbb{E}[\mathbf{h}(s,\mathbf{x}_s)|\mathcal{Y}_s] = \int \mathbf{h}(s,\mathbf{x}_s)p(\mathbf{x}_s|\mathcal{Y}_s)d\mathbf{x}_s \tag{2.7}$$

and $\Sigma_{v,t} = \mathbb{E}[\mathbf{v}_t]$.

Although the Kusher's equation leads a way to solve the NLF problem completely, it needs to solve an infinite-dimensional system to get even the conditional mean (cf. [**13**]). Generally speaking, the solution is neither in a closed form nor easy to be computed numerically, due to the nonlinearity with respect to $p(\mathbf{x}_t|\mathcal{Y}_t)$ in (2.7).

Through the Kallianpur-Striebel formula [**11**], one can define the unnormalized conditional pdf $\pi(\mathbf{x}_t|\mathcal{Y}_t)$ through the following procedure. In particular, for any $\varphi \in \mathcal{B}(\Omega)$, the Borel σ-field on the state space Ω, which is a complete seperable metric space,

$$\mathrm{P}[\varphi] := \int \varphi(\mathbf{x}_t)p(\mathbf{x}_t|\mathcal{Y}_t)d\mathbf{x}_t = \frac{\tilde{\mathbb{E}}[\tilde{\mathbf{z}}_t\varphi(\mathbf{x}_t)|\mathcal{Y}_t]}{\tilde{\mathbb{E}}(\tilde{\mathbf{z}}_t|\mathcal{Y}_t)}, \quad \tilde{\mathbb{P}} - a.s.,$$

where the process $\tilde{\mathbf{z}}_t$ satisfying

$$d\tilde{\mathbf{z}}_t = \sum_{i=1}^{N_y} \tilde{\mathbf{z}}_t h_i(t,\mathbf{x}_t)dy_i,$$

with h_i and y_i the ith component of $\mathbf{h}$ and $\mathbf{y}$, respectively. $\tilde{\mathbb{P}}$ is the probablity measure introduced by the process $\tilde{\mathbf{z}}_t$, such that

$$\left.\frac{d\mathbb{P}}{d\tilde{\mathbb{P}}}\right|_{\mathcal{F}_t} = \tilde{\mathbf{z}}_t,$$

for all $t \geq 0$, where $\mathcal{F}_t$ is the filtration of $\mathbf{x}_t$. And $\tilde{\mathbb{E}}$ is the expectation with respect to $\tilde{\mathbb{P}}$. Hence, the unnormalized conditional pdf of $\mathbf{x}_t$ is defined as following. For any $\varphi \in \mathcal{B}(\mathbb{S})$, we define

$$\Pi[\varphi] = \mathrm{P}[\varphi]\Pi[\mathbf{1}], \quad \forall\, t \geq 0, \quad \tilde{\mathbb{P}}(\mathbb{P}) - a.s. \tag{2.8}$$

where $\Pi[\varphi] := \int \varphi(\mathbf{x}_t)\pi(\mathbf{x}_t|\mathcal{Y}_t)d\mathbf{x}_t$. Equation (2.8) explains the usage of the term "unnormalized" of $\pi(\mathbf{x}_t|\mathcal{Y}_t)$, since the denominator $\Pi[\mathbf{1}]$ can be viewed as the normalizing factor. Under certain mild condition, the unnormalized conditional pdf $\pi(\mathbf{x}_t|\mathcal{Y}_t)$ satisfies the Duncan-Mortensen-Zakai's (DMZ) equation (cf. [**24**, **59**, **77**])

$$d\pi(\mathbf{x}_t|\mathcal{Y}_t) = \mathcal{L}^*\pi(\mathbf{x}_t|\mathcal{Y}_t)dt + \mathbf{h}(t,\mathbf{x}_t)\pi(\mathbf{x}_t|\mathcal{Y}_t)d\mathbf{y}_t, \tag{2.9}$$

where $\mathcal{L}^*$ is defined in (2.6). There is an one-to-one correspondence between the solution of Kusher's equation and that of DMZ equation. And it is clear to see that DMZ equation is linear with respect to the unnormalized conditional pdf $\pi(\mathbf{x}_t|\mathcal{Y}_t)$. Therefore, studies on how to numerically solve the DMZ equation efficiently is the key to solve NLF problems completely.

3. Two categories: local and global approaches

3.1. Local approach. Around 1960s, the Kalman filtering (KF) has been developed in the seminal papers [**45**, **46**] by using the orthogonal projection method, under the linear and Gaussian assumptions. It has been shown to be optimal in the sense that it is unbiased, i.e. $\mathbb{E}[\hat{\mathbf{x}}_k] = \mathbb{E}[\mathbf{x}_k]$ and is a minimum variance estimate. In the late 1960s, Kailath [**44**] reformulated the KF with the innovation approach [**1**] and the tool of martingales theory [**23**]. The KF is also optimal from the viewpoint of innovation that it is whitening filter. The celebrated KF can also be derived within the Bayesian framework, which is reduced to the maximum a posteriori (MAP) solution [**13**] and the maximum likelihood (ML) solution [**64**]. The nice Bayesian interpretation of KF can be found in [**35**]. Recently, the derivation from DMZ equation is investigated in [**25**, **26**]. We refer the interested readers for a detailed history of KF and its variants to [**44**, **40**, **29**] and reference therein. To be somewhat self-contained, we briefly sketch the re-derivation of the KF from the discrete DMZ equation [**26**] under the linear and Gaussian assumptions. Equation (2.1) reduces to the following special case:

$$\begin{cases} \mathbf{x}_k = F_{k,k-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1} \\ \mathbf{y}_k = H_k\mathbf{x}_k + \mathbf{v}_k \end{cases}, \tag{3.1}$$

where $F_{k,k-1}$ and H_k are called transition matrix and measurement matrix, respectively. Let us further assume that the state process $\mathbf{x}_k$, the observation process $\mathbf{y}_k$ and the noise processes $\mathbf{w}_k$, $\mathbf{v}_k$ are mutually independent. To simplify notation, we suppose that $\{\mathbf{w}_k\}_{k=1}^{\infty}$ and $\{\mathbf{v}_k\}_{k=0}^{\infty}$ are sequences of independent $\mathcal{N}(\mathbf{0}, \mathbf{I}_{N_x})$ and $\mathcal{N}(\mathbf{0}, \mathbf{I}_{N_y})$ random variables, respectively.

The following theorem provides a recursive formula for unnormalized conditional pdf of $\mathbf{x}_k$ given $\mathbf{y}_{1:k}$. It is the discrete time version of DMZ equation.

Theorem 3.1 ([**25**, **26**]). *$\pi(\mathbf{x}_k|\mathbf{y}_{1:k})$ satisfies the recursion:*

$$\pi(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{\phi(\mathbf{y}_k - H_k\mathbf{x}_k)}{\phi(\mathbf{y}_k)} \int_{\mathbb{R}^{N_x}} \pi(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})\psi(\mathbf{x}_k - F_{k,k-1}\mathbf{x}_{k-1})d\mathbf{x}_{k-1}, \tag{3.2}$$

where $\psi(\mathbf{x}) = (2\pi)^{-\frac{N_x}{2}} \exp\left(-\frac{\mathbf{x}'\mathbf{x}}{2}\right)$ and $\phi(\mathbf{y}) = (2\pi)^{-\frac{N_y}{2}} \exp\left(-\frac{\mathbf{y}'\mathbf{y}}{2}\right)$, for $\mathbf{x} \in \mathbb{R}^{N_x}$ and $\mathbf{y} \in \mathbb{R}^{N_y}$, respectively.

The DMZ equation (3.2) is exact under the linear and Gaussian assumptions, and it has the form of a convolution equation. It is readily to verify that (3.2) yields the KF.

The KF consists of an iterative prediction-correction procedure. Let us denote $\mathbf{x}_{k|k-1} = \mathbb{E}[\mathbf{x}_k|\mathbf{y}_{1:k-1}]$ the conditional expectation of $\mathbf{x}_k$ given $\mathbf{y}_{1:k-1}$, and the conditional variance $\mathbf{\Sigma}_{k|k-1} = Var[\mathbf{x}_k|\mathbf{y}_{1:k-1}]$. Assume it is $\mathbf{x}_{k-1} \sim \mathcal{N}(\mu_{k-1}, \mathbf{\Sigma}_{k-1})$, that is, the normalized conditional pdf is

$$p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) = |\mathbf{\Sigma}_{k-1}|^{-\frac{1}{2}} \psi\left(\mathbf{\Sigma}_{k-1}^{-1}(\mathbf{x}_{k-1} - \mu_{k-1})\right).$$

Prediction: Starting from $\pi(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) \sim \psi\left(\mathbf{\Sigma}_{k-1}^{-1}(\mathbf{x}_{k-1} - \mu_{k-1})\right)$, we have

$$\mathbf{x}_{k|k-1} = \mathbb{E}[\mathbf{x}_k|\mathbf{y}_{1:k-1}] \overset{(3.1)}{=} \mathbb{E}[F_{k,k-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1}|\mathbf{y}_{1:k-1}] = \mathbb{E}[F_{k,k-1}\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}] = F_{k,k-1}\mu_{k-1},$$

with $\mu_{k-1} = \mathbb{E}[\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}]$, and

$$\begin{aligned}\mathbf{\Sigma}_{k|k-1} =&\mathbb{E}[(\mathbf{x}_k - \mathbf{x}_{k|k-1})(\mathbf{x}_k - \mathbf{x}_{k|k-1})'|\mathbf{y}_{1:k-1}]\\ =&\mathbb{E}[(F_{k,k-1}(\mathbf{x}_{k-1} - \mu_{k-1}) + \mathbf{w}_{k-1})\,(F_{k,k-1}(\mathbf{x}_{k-1} - \mu_{k-1}) + \mathbf{w}_{k-1})'\,|\mathbf{y}_{1:k-1}]\\ =&F_{k,k-1}\mathbf{\Sigma}_{k-1}F'_{k,k-1} + \mathbf{I}^2_{N_x\times N_x}\end{aligned}$$

Correction: The posterior conditional pdf is shown to be $\mathcal{N}(\mu_k, \mathbf{\Sigma}_k)$. That is, $\pi(\mathbf{x}_k|\mathbf{y}_{1:k}) \sim \psi\left(\mathbf{\Sigma}_k^{-1}(\mathbf{x}_k - \mu_k)\right)$, where μ_k and $\mathbf{\Sigma}_k$ are given in (3.4) and (3.3) below.

Theorem 3.2 ([**26**]). *Suppose* $\mathbf{x}_{k-1} \sim \mathcal{N}(\mu_{k-1}, \mathbf{\Sigma}_{k-1})$. *Then* $\mathbf{x}_k \sim \mathcal{N}(\mu_k, \mathbf{\Sigma}_k)$, *where*

$$\mathbf{\Sigma}_k = \mathbf{\Sigma}_{k|k-1} - \mathbf{\Sigma}_{k|k-1}H'_k\left(\mathbf{I}^2_{N_y\times N_y} + H_k\mathbf{\Sigma}_{k|k-1}H'_k\right)^{-1}H_k\mathbf{\Sigma}_{k|k-1}, \tag{3.3}$$

and

$$\mu_k = F_{k,k-1}\mu_{k-1} + \mathbf{\Sigma}_{k|k-1}H'_k\left(\mathbf{I}^2_{N_y\times N_y} + H_k\mathbf{\Sigma}_{k|k-1}H'_k\right)^{-1}(\mathbf{y}_k - H_kF_{k,k-1}\mu_{k-1}). \tag{3.4}$$

The quantity $K_k = \mathbf{\Sigma}_{k|k-1}H'_k\left(\mathbf{I}^2_{N_y\times N_y} + H_k\mathbf{\Sigma}_{k|k-1}H'_k\right)^{-1}$ *is the so-called Kalman gain.*

The KF is well-known to be optimal under linear Gaussian assumptions. However, real applications generally can't be set up with the model satisfying these assumptions. Therefore, many variants have been developed, following the idea of the KF in the hope of solving the general NLF problems.

3.1.1. Linearization methods: extended Kalman filter (EKF) [29]. The basic idea of EKF is to linearize (2.1) at the previous step's estimation, i.e.

$$\hat{F}_{k,k-1} = \left.\frac{d\mathbf{f}(\mathbf{x})}{d\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_{k-1}}, \quad \hat{H}_k = \left.\frac{d\mathbf{h}(\mathbf{x})}{d\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_{k-1}}.$$

Then the KF is applied to this linearized equation. The EKF is biased in general and it only works well when the true posterior conditional pdf is almost Gaussian. It could perform extremely poor especially when the true posterior is heavily skew or multimodal or the dynamics are highly nonlinear. Another drawback of EKF is the heavy computation to evaluate the Jacobian matrix at each time step. A detailed discussion on EKF and its applications can be found in many books, say [**29**, **40**] etc.

3.1.2. Finite sum approximation: Gaussian sum filter (GSF) [2, 49]. Unlike the EKF, which approximate the nonlinear term near the vicinity of the previous estimation, the GSF proposed to approximate the posterior pdf by a mixture of weighted Gaussians. That is,

$$p(\mathbf{x}) = \sum_{i=1}^{M} w_i\mathcal{N}(\mu^i, \mathbf{\Sigma}^i),$$

where the weighted coefficients $w_i > 0$ and $\sum_{i=1} w_i = 1$. Then the GSF runs a bank of EKF in parallel to obtain a suboptimal estimation.

3.1.3. Deterministic points approximation. Ito, et. al [**38**] improves the GSF further to avoid the intensive computational part of EKF, i.e. the evaluation of the Jacobian matrix. Indeed, with the Gaussian assumption, the Bayesian nonlinear filtering framework is given as follows:

Prediction:

$$\mathbf{x}_{k|k-1} = \int_{\mathbb{R}^{N_x}} \mathbf{f}(\mathbf{x}_{k-1})\mathcal{N}\left(\mathbf{x}_{k-1}; \mu_{k-1}, \mathbf{\Sigma}_{k-1}\right) d\mathbf{x}_{k-1} \tag{3.5}$$

$$\mathbf{\Sigma}_{k|k-1} = \int_{\mathbb{R}^{N_x}} \mathbf{f}(\mathbf{x}_{k-1})(\mathbf{f}(\mathbf{x}_{k-1}))'\mathcal{N}\left(\mathbf{x}_{k-1}; \mu_{k-1}, \mathbf{\Sigma}_{k-1}\right) d\mathbf{x}_{k-1} - \mathbf{x}_{k|k-1}\mathbf{x}'_{k|k-1} + \mathbf{I}_{N_x\times N_x}, \tag{3.6}$$

where $\mathcal{N}\left(\mathbf{x}_{k-1}; \mu_{k-1}, \mathbf{\Sigma}_{k-1}\right)$ represents the multivariate normal distribution with the mean μ_{k-1} and the covariance $\mathbf{\Sigma}_{k-1}$.

Correction:

$$\begin{aligned}\mu_k =&\mathbf{x}_{k|k-1} + \mathbf{L}_k(\mathbf{y}_k - \mathbf{z}_k)\\ \mathbf{\Sigma}_k =&\mathbf{\Sigma}_{k|k-1} - \mathbf{L}_k\mathbf{\Sigma}'_{\mathbf{xz}},\end{aligned}$$

where

$$\mathbf{L}_k =\mathbf{\Sigma}_{\mathbf{xz}}(\mathbf{R}_k + \mathbf{\Sigma}_{\mathbf{zz}})^{-1} \tag{3.7}$$

$$\mathbf{z}_k = \int_{\mathbb{R}^{N_x}} \mathbf{h}(\mathbf{x}_k)\mathcal{N}\left(\mathbf{x}_k; \mathbf{x}_{k|k-1}, \mathbf{\Sigma}_{k|k-1}\right) d\mathbf{x}_k \tag{3.8}$$

$$\mathbf{\Sigma}_{\mathbf{xz}} = \int_{\mathbb{R}^{N_x}} (\mathbf{x}_k - \mathbf{x}_{k|k-1})(\mathbf{h}(\mathbf{x}_k) - \mathbf{z}_k)'\mathcal{N}\left(\mathbf{x}_k; \mathbf{x}_{k|k-1}, \mathbf{\Sigma}_{k|k-1}\right) d\mathbf{x}_k \tag{3.9}$$

$$\mathbf{\Sigma}_{\mathbf{zz}} = \int_{\mathbb{R}^{N_x}} (\mathbf{h}(\mathbf{x}_k) - \mathbf{z}_k)(\mathbf{h}(\mathbf{x}_k) - \mathbf{z}_k)'\mathcal{N}\left(\mathbf{x}_k; \mathbf{x}_{k|k-1}, \mathbf{\Sigma}_{k|k-1}\right) d\mathbf{x}_k. \tag{3.10}$$

The integrals in (3.5)-(3.6) and (3.8)-(3.10) can be approximated by various numerical rules, such as Gauss quadrature rule, unscented transformation and cubature rule, etc. Consequently, they lead to different filtering methods, such as unscented Kalman filter (UKF) [**42**, **43**], Gaussian quadrature Kalman filter (GKF) [**38**, **5**] and cubature Kalman filter (CKF) [**4**, **6**]. Very recently, Jia, et al. [**41**] investigated the high-dimensional NLF problems by GKF with the sparse-grid algorithm [**66**].

3.2. Global approach. The local approaches performs more effective than the global ones. The real-time manner is very appealing in many real applications. However, the common drawbacks inherited from KF are the follows:

1) They perform well only when the posterior conditional pdf is close to the Gaussian and the dynamic system is almost linear.
2) Only mean and variance are obtained. No more statistical information is available.

Unlike the local approaches, the global ones are aim to obtain the approximation of the conditional pdf. No apriori assumptions need to be imposed on the system or the posterior conditional pdf. All statistical information is obtained automatically. In this sense, the NLF problems are solved completely. The only problems are the real-time manner and the heavy computation in high-dimensional states NLF problems.

In general, the NLF problem is intractable with finite statistics, say mean and moments. It is interesting to understand under what conditions certain NLF problems can be transformed into finite dimensional ones. And is there any NLF problem essentially infinite-dimensional?

3.2.1. Finite-dimensional filters. KF is a typical finite-dimensional filter in the sense that it can be implemented by integrating a finite number of (actually two) ordinary differential equations (ODE). Or say, it has the sufficient statistics with finite (two) variables, i.e. the conditional mean and variance. However, not all NLF problems are finite-dimensional. For instance, Hazewinkel et al. have shown in [**34**] the nonexistence of finite-dimensional filter for the cubic sensor problem. Hence, it is meanful to construct finite-dimensional filter for more general NLF problems and to study the necessary and sufficient conditions to guarantee the existence of such filters.

As far as the author knows, Beneš [**8**] is the first one to investigate the exact finite-dimensional filter in the NLF context. Later, Yau [**14**] gives a more general case including the KF and Beneš filter as special cases. Around 2000, the exact finite-dimensional filter from the differential geometric point of view is studied by Brigo et al. [**10**] and reference therein, which is the so-called projection filters.

At the International Congress of Mathematicians in 1983, Brockett [**11**] proposed to systematically study the finite-dimensional filters by using the estimation algebra to classify all the finite-dimensional ones. The estimation algebra E of the filtering model (2.2) is defined as the Lie algebra generated by $\{L_0, L_1, \ldots, L_m\}$, where L_0 is related to $\mathcal{L}^*$ and L_i, $i = 1, \ldots, m$, are the zero degree differential operators of multiplication by h_i. As an immediate application of the classification, it can be used to construct new exact finite-dimensional filters for NLF problems. The following theorem given by Ocone [**61**] is the first one characterized the functions in a finite-dimensional estimation algebra.

Theorem 3.3 ([**61**]). *Let E be a finite-dimensional estimation algebra. If a function ξ is in E, then ξ is a polynomial of degree at most two.*

In particular, if $\mathbf{G} = \mathbf{Q} = \mathbf{R} = \mathbf{I}$, $\mathbf{I}$ is the identity matrix, then

$$L_0(\circ) := \frac{1}{2}\sum_{i=1}^{N_x} \frac{\partial^2 \circ}{\partial x_i^2} - \sum_{i=1}^{N_x} f_i \frac{\partial \circ}{\partial x_i} - \sum_{i=1}^{N_x} \frac{\partial f_i}{\partial x_i} \circ - \frac{1}{2}\sum_{i=1}^{N_y} h_i^2 \circ, \tag{3.11}$$

and $L_i(\circ) := h_i \circ$, where f_i and h_i are the ith component of $\mathbf{f}$ and $\mathbf{h}$, respectively. In real applications, the actual observations consist of piecewise smooth sample paths $y(t)$. Davis [**19**] was interested in constructing robust estimators from these kind of observation paths. He considered a version of (2.9) dealing with path-wise observation $y(t)$. It follows immediately from an exponential transformation:

$$u(t, \mathbf{x}) = \exp\left(\sum_{i=1}^{N_y} h_i(\mathbf{x}) y_i(t)\right) \pi(t, \mathbf{x}).$$

Equation (2.9) is reduced to the following PDE, which is called robust DMZ equation in our context.

$$\frac{\partial u}{\partial t}(t, \mathbf{x}) = L_0 u(t, \mathbf{x}) + \sum_{i=1}^{N_y} y_i(t)[L_{N_y}, L_i]u(t, \mathbf{x}) + \frac{1}{2}\sum_{i,j=1}^{N_y} y_i(t) y_j(t)[[L_0, L_i], L_j]u(t, \mathbf{x}), \tag{3.12}$$

where $[L_i, L_j]$ is the Lie bracket of the differential operators L_i and L_j, defined as $[L_i, L_j](\phi) := L_i(L_j(\phi)) - L_j(L_i(\phi))$, for any $\phi \in C^\infty$.

Yau [**69**] constructs a class of finite-dimensional filter for NLF problem using estimation algebra techniques. It is referred as Yau filter in [**14**], which includes the Kalman-Bucy filter and Beneš filter as special cases. Yau also gave a necessary and sufficient condition to guarantee the estimation algebra to be finite-dimensional.

In particular, the following theorem from [**69**] shows how to construct finite-dimensional filters from finite-dimensional estimation algebras with maximal rank. The estimation algebra E is said to be of maximal rank if, for any $1 \le i \le N_x$, there exists a constant c_i such that $x_i + c_i$ is in E.

Theorem 3.4 (Yau [**69**]). *Let E be an estimation algebra of (2.2) satisfying $\frac{\partial f_j}{\partial x_i} - \frac{\partial f_i}{\partial x_j} = c_{ij}$, where the c_{ij}s are constants for all $1 \le i, j \le N_x$. Suppose that E is a finite dimensional estimation algebra of maximal rank. Then E has a basis of the form $1, x_1, \ldots, x_{N_x}, D_1, \ldots, D_{N_x}$ and L_0, and $\sum_{i=1}^{N_x} \frac{\partial f_i}{\partial x_i} + \sum_{i=1}^{N_x} f_i^2 + \sum_{i=1}^{N_y} h_i^2$ is a degree two polynomial $\sum_{i,j=1}^{N_x} a_{ij} x_i x_j + \sum_{i=1}^{N_x} b_i x_i + d$, where $D_i = \frac{\partial}{\partial x_i} - f_i$ and L_0 is defined in (3.11). The robust DMZ equation (3.12) has a solution for all $t \ge 0$ of the form*

$$u(t, \mathbf{x}) = e^{T(t)} e^{r_{N_x}(t) x_{N_x}} \ldots e^{r_1(t) x_1} e^{s_{N_x}(t) D_{N_x}} \ldots e^{s_1(t) D_1} e^{t L_0} \sigma_0$$

where $T(t), r_1(t), \ldots, r_{N_x}(t), s_1(t), \ldots, s_{N_x}(t)$ satisfies the following ODEs:

$$\begin{aligned}
\frac{ds_i}{dt}(t) =& r_i(t) + \sum_{j=1}^{N_x} s_j(t) c_{ji} + \sum_{k=1}^{N_x} h_{ki} y_k(t), \quad 1 \le i \le N_x;\\
\frac{dr_j}{dt}(t) =& \frac{1}{2}\sum_{i=1}^{N_x} s_i(t)(a_{ij} + a_{ji}), \quad 1 \le j \le N_x;\\
\frac{dT}{dt} =& -\frac{1}{2}\sum_{i=1}^{N_x} r_i^2(t) - \frac{1}{2}\sum_{i=1}^{N_x} s_i^2(t)\left(\sum_{j=1}^{N_x} c_{ij}^2 - a_{ij}\right) + \sum_{i=1}^{N_x} r_i(t) - \sum_{j=2}^{N_x}\sum_{i=1}^{j} s_j(t) c_{ij}\\
&+ \sum_{1 \le i < k \le N_x} s_i(t) s_k(t)\left[\sum_{j=1}^{N_x} c_{ij} c_{jk} + \frac{1}{2}(a_{ik} + a_{ki})\right] + \frac{1}{2}\sum_{i=1}^{N_x} s_i(t) b_i\\
&+ \frac{1}{2}\sum_{i,j=1}^{N_y} y_i(t) y_j(t) \sum_{k=1}^{N_x} h_{ik} h_{jk} - \sum_{i,j=1}^{N_x} s_i(t) r_j(t) c_{ij},
\end{aligned}$$

where $h_k(\mathbf{x}) = \sum_{j=1}^{N_x} h_{kj}x_j + e_k$, $1 \le k \le N_y$, h_{kj} *and* e_k *are constants. In particular, a universal finite-dimensional filter exists.*

The characterization of the condition $\frac{\partial f_j}{\partial x_i} - \frac{\partial f_i}{\partial x_j} = c_{ij}$, where c_{ij} are constants for all $1 \le i, j \le N_x$, is also given in [**69**].

Theorem 3.5 ([**69**]). $\frac{\partial f_j}{\partial x_i} - \frac{\partial f_i}{\partial x_j} = c_{ij}$, *where* c_{ij} *are constants for all* $1 \le i, j \le N_x$, *if and only if*

$$(f_1, \ldots, f_{N_x}) = (l_1, \ldots, l_{N_x}) + \left(\frac{\partial \psi}{\partial x_1}, \ldots, \frac{\partial \psi}{\partial x_{N_x}}\right),$$

where $l_1, \ldots, l_{N_x}$ *are polynomials of degree one and* ψ *is a* C^∞ *function.*

And the classification of the finite-dimensional estimation algebra with maximal rank has been completed in [**70**, **72**].

Theorem 3.6 ([**72**]). *Suppose that the state space of the filtering model* (2.2) *is of dimension* N_x. *If* E *is the finite-dimensional estimation algebra with maximal rank, then* $f = \nabla\phi + (\alpha_1, \ldots, \alpha_{N_x})$, *where* ϕ *is a smooth function and* α_i, $1 \le i \le N_x$ *are affine functions and* E *is a real vector space of dimension* $2N_x + 2$ *with basis given by* $1, x_1, \ldots, x_{N_x}, D_1, \ldots, D_{N_x}$ *and* L_0.

The finite-dimensional filter can also be constructed from the finite-dimensional estimation algebra with non-maximal rank, see [**63**]. However, the classification of the non-maximal rank ones is still wide open, except some partial results, including those for low-dimensional estimation algebra with arbitrary states' dimension [**73**, **15**]; the classification with state dimension 2 and arbitrary dimensional estimation algebra [**68**].

Besides the classification of the estimation algebra, Yau et al. [**74**] introduced the direct method to solve the NLF with finite-dimensional estimation algebra, which has been further generalized by [**37**, **72**]. Based on the Wei-Norman approach of the estimation algebra to solve the DMZ equation, one needs to know the basis of the estimation algebra explicitly, so that the DMZ equation can be reduced to a finite system of ODE and several first-order linear PDEs. Unfortunately, the basis can only be known when the estimation algebra has maximal rank. The direct method in [**74**, **37**, **71**] is easy to implement and don't rely on the explicit basis of the estimation algebra, which can be applied to all Yau filters [**14**]. Moreover, the number of sufficient statistics required to acquire the conditional pdf is N_x. More precisely, in [**37**] Yau et al. assume that the following conditions are satisfied:

1) $\frac{\partial f_j}{\partial x_i} - \frac{\partial f_i}{\partial x_j} = c_{ij}$, where c_{ij} are constants, $1 \le i, j \le N_x$. This is so-called Yau filter in [**14**]. This condition is equivalent to

$$f_i(\mathbf{x}) = l_i(\mathbf{x}) + \frac{\partial F}{\partial x_i}(\mathbf{x}), \tag{3.13}$$

   for $1 \le i \le N_x$, where $l_i(x) = \sum_{j=1}^{N_x} d_{ij}x_j + d_i$ for $1 \le i \le N_x$ and F is a C^∞ function.
2) Yau showed in [**69**] that the observation functions $h_1, \cdots, h_{N_y}$ are polynomials of degree at most one for all the Yau filters with finite-dimensional estimation algebra. Without loss of generality, we assume that

$$h_i(\mathbf{x}) = \sum_{j=1}^{N_x} c_{ij}x_j + c_i, \tag{3.14}$$

   for $1 \le i \le N_y$, where c_{ij} and c_i are constants.
3) It is also shown in [**69**] that

$$\eta(\mathbf{x}) := \sum_{i=1}^{N_x} \frac{\partial f_i}{\partial x_i} + \sum_{i=1}^{N_x} f_i^2 + \sum_{i=1}^{N_y} h_i^2$$

is a polynomial of degree at most two for all the Yau filter with finite-dimensional estimation algebra. Without loss of generality, let us assume that

$$\eta(\mathbf{x}) = \sum_{i,j=1}^{N_x} \eta_{ij} x_i x_j + \sum_{i=1}^{N_x} \eta_i x_i + \eta_0, \tag{3.15}$$

where η_{ij}, η_i and η_0 are constants.

Under the conditions above, the solution of the robust DMZ equation (3.12) can be solved directly as described in the following theorem:

Theorem 3.7 ([**71**]). *Consider the filtering model* (2.2) *with* $\mathbf{Q} = \mathbf{G} = \mathbf{R} = \mathbf{I}$ *with the conditions* (3.13)-(3.15). *Then the solution* $u(t, \mathbf{x})$ *for the robust DMZ equation* (3.12) *is reduced to the solution of* $\tilde{u}(t, \mathbf{x})$ *for the forward Kolmogorov equation*

$$\begin{cases} \dfrac{\partial \tilde{u}}{\partial t}(t, \mathbf{x}) = \dfrac{1}{2} \triangle \tilde{u}(t, \mathbf{x}) - \displaystyle\sum_{i=1}^{N_x} H_i(\mathbf{x}) \frac{\partial \tilde{u}}{\partial x_i}(t, \mathbf{x}) - P(\mathbf{x}) \tilde{u}(t, \mathbf{x}) \\ \tilde{u}(0, \mathbf{x}) = e^{G(\mathbf{x}) - F(\mathbf{x})} \sigma_0(\mathbf{x}) \end{cases} \tag{3.16}$$

where

$$\tilde{u}(t, \mathbf{x}) = exp\left[c(t) + G(\mathbf{x}) - \sum_{i=1}^{N_x} a_i(t) x_i - F(\mathbf{x} + b(t)) \right] u(t, \mathbf{x} + b(t))$$

and $a_i(t)$, $b_i(t)$ *and* $c(t)$ *satisfy the following system of ODEs:*

$$\begin{cases} a_i'(t) - \dfrac{1}{2} \displaystyle\sum_{j=1}^{N_x} (\eta_{ij} + \eta_{ji}) b_j(t) + \sum_{j=1}^{N_x} d_{ji} b_j'(t) = 0 \\ a_i(0) = 0 \end{cases}, \tag{3.17}$$

$$\begin{cases} b_i'(t) - a_i(t) - \displaystyle\sum_{j=1}^{N_x} d_{ij} b_j(t) + \sum_{j=1}^{N_x} c_{ji} y_j(t) = 0 \\ b_i(0) = 0 \end{cases}, \tag{3.18}$$

$$\begin{cases} c_i'(t) = -\dfrac{1}{2} \displaystyle\sum_{i=1}^{N_x} (b_i'(t))^2 + \sum_{i=1}^{N_x} a_i(t) b_i'(t) - \sum_{i=1}^{N_x} d_i b_i'(t) + \frac{1}{2} \sum_{i,j=1}^{N_x} \eta_{ij} b_i(t) b_j(t) + \frac{1}{2} \sum_{i=1}^{N_x} \eta_i b_i(t) \\ c(0) = 0 \end{cases}, \tag{3.19}$$

for $1 \leq i \leq N_x$, *if we can choose* $H(\mathbf{x})$, $G(\mathbf{x})$ *and* $P(\mathbf{x})$ *such that*

$$\frac{1}{2} \sum_{i=1}^{N_x} H_i^2(\mathbf{x}) - \frac{1}{2} \sum_{i=1}^{N_x} \frac{\partial H_i}{\partial x_i}(\mathbf{x}) - \frac{1}{2} \eta(\mathbf{x}) + P(\mathbf{x}) = 0,$$

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

The possible choices of $H(\mathbf{x})$, $G(\mathbf{x})$ and $P(\mathbf{x})$ in [**71**] include the follows:

1) Choose a C^∞ function $G(\mathbf{x})$ such that

$$\begin{aligned} \triangle G(\mathbf{x}) + |\nabla G|^2(\mathbf{x}) + 2 \sum_{i=1}^{N_x} l_i(\mathbf{x}) \frac{\partial G}{\partial x_i}(\mathbf{x}) &= \eta(\mathbf{x}) - \sum_{i=1}^{N_x} l_i^2(\mathbf{x}) - \sum_{i=1}^{N_x} \frac{\partial l_i}{\partial x_i}(\mathbf{x}), \\ H_i(\mathbf{x}) &= \frac{\partial G}{\partial x_i}(\mathbf{x}) + l_i(\mathbf{x}), \end{aligned}$$

and

$$P(\mathbf{x}) = \sum_{i=1}^{N_x} \frac{\partial H_i}{\partial x_i}(\mathbf{x}) = \sum_{i=1}^{N_x} \left(\frac{\partial^2 G}{\partial x_i^2}(\mathbf{x}) + \frac{\partial l_i}{\partial x_i}(\mathbf{x}) \right).$$

2) Choose

$$\begin{aligned} G(\mathbf{x}) \equiv & 0; \\ P(\mathbf{x}) = & \frac{1}{2}\eta(\mathbf{x}) - \frac{1}{2}\sum_{i=1}^{N_x} l_i^2(\mathbf{x}) - \frac{1}{2}\sum_{i=1}^{N_x}\frac{\partial l_i}{\partial x_i}(\mathbf{x}); \\ H_i(\mathbf{x}) = & l_i(\mathbf{x}), \end{aligned}$$

for $1 \leq i \leq N_x$.

3) Choose a function $G(\mathbf{x})$ such that $\frac{\partial G}{\partial x_i}(\mathbf{x}) = -l_i(\mathbf{x})$ if $d_{ij} = d_{ji}$ for $1 \leq i, j \leq N_x$. Let $P(\mathbf{x}) = \frac{1}{2}\eta(\mathbf{x})$ and $H_i(\mathbf{x}) \equiv 0$, $1 \leq i \leq N_x$.

4) Choose

$$\begin{aligned} G(\mathbf{x}) = & F(\mathbf{x}); \\ P(\mathbf{x}) = & \frac{1}{2}\eta(\mathbf{x}) - \frac{1}{2}\sum_{i=1}^{N_x} f_i^2(\mathbf{x}) + \frac{1}{2}\sum_{i=1}^{N_x}\frac{\partial f_i}{\partial x_i}(\mathbf{x}); \\ H_i(\mathbf{x}) = & f_i(\mathbf{x}), \end{aligned}$$

for $1 \leq i \leq N_x$.

3.2.2. Sequential Monte Carlo methods and particle filters (PF). The use of Monte Carlo methods for NLF can be traced back to [**33**]. The algorithm is so-called sequential importance sampling (SIS). Although it has been known since 1970s, it is not commonly used in the NLF problems, due to some major drawbacks until [**30**], the so-called bootstrap filter has been developed. In [**30**], Gordon et. al. identified the degeneracy of the importance weights as sample improverishment. In brief, it asserts that most of the samples are annihilated due to the very small normalized importance weights in the long run. The remedy is to rejuvenate by replicating the samples with high importance weights and removing those with low weights. This is similar as the algorithm in [**65**], so-called sampling and importance resampling (SIR). Starting from the bootstrap filter [**30**], various similar filtering have been studied, including Monte Carlo filter [**47**], particle filter [**21**] and etc. A good introduction to this field has been written by Künsch [**52**], while the interesting recent developments in theory and applications are covered in [**20**].

The sequential Monte Carlo method is within the Bayesian framework. In the NLF context, we are concerned to compute the expectations of the form:

$$\mathbb{E}(\varphi) = \int \varphi(x)p(x)dx,$$

where $\varphi(\circ)$ are some functions for estimation. For example, $\varphi(x) = x$ gives the mean. The approximation of integral by Monte Carlo method can be achieved by generating random samples from p, denoted as $\left\{x^{(i)}\right\}_{i=1}^{N}$, and approximate p by point masses, i.e. $p(x) = \sum_{i=1}^{N}\delta_{x^{(i)}}(x)$, where $\delta_a(x)$ is the Kronecker-delta function. Henceforth, the expectation $\mathbb{E}(\varphi)$ is given by

$$\mathbb{E}(\varphi) \approx \frac{1}{N}\sum_{i=1}^{N}\varphi\left(x^{(i)}\right). \tag{3.20}$$

Intuitively, as $N \to +\infty$, $\mathbb{E}(\varphi)$ is well approximated.

Sampling directly from the distribution p is no doubt a good choice. However, in the NLF context, neither the prior pdf $p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})$ nor the posterior one $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ are known. Generally speaking, we can't sample directly from p. Instead, we sample from another convenient distribution q, which is called importance distribution or instruction distribution. To guarantee the unbiased estimation of $\mathbb{E}(\varphi)$, we need to make a correction by

$$\mathbb{E}(\varphi) = \int \varphi(x)p(x)dx = \int \varphi(x)\frac{p(x)}{q(x)}q(x)dx \overset{(3.20)}{\approx} \frac{1}{\sum_{j=1}^{N} w^{(j)}}\sum_{i=1}^{N} w^{(i)}\varphi\left(x^{(i)}\right), \tag{3.21}$$

where $w^{(i)} := \frac{p\left(x^{(i)}\right)}{q\left(x^{(i)}\right)}$ is the unnormalized importance weight.

Back to the Bayesian framework, let us apply the Monte Carlo sampling technique as follows. We sample N particles $\left\{\mathbf{x}_k^{(i)}\right\}_{i=1}^N$ from an importance distribution $q_k(\mathbf{x}_k|\mathbf{y}_{1:k})$ and compute the unnormalized importance weights

$$w_k^{(i)} = \frac{p\left(\mathbf{x}_k^{(i)}|\mathbf{y}_{1:k}\right)}{q\left(\mathbf{x}_k^{(i)}|\mathbf{y}_{1:k}\right)}, \tag{3.22}$$

for $i = 1, 2, \ldots, N$. Then the conditional expectation of any function φ can be approximated by the weighted sample $\left\{\mathbf{x}^{(i)}, w_k^{(i)}\right\}_{i=1}^N$:

$$\mathbb{E}(\varphi) = \int \varphi(\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{y}_{1:k}) d\mathbf{x}_k \approx \sum_{i=1}^N \frac{w_k^{(i)}}{\sum_{j=1}^N w_k^{(j)}} \varphi\left(\mathbf{x}_k^{(i)}\right). \tag{3.23}$$

How does the pair $\left(\mathbf{x}_k^{(i)}, w_k^{(i)}\right)$ propagate through the dynamic system? The samples $\left\{\mathbf{x}_k^{(i)}\right\}_{i=1}^N$ are propagated as

$$\mathbf{x}_{k+1}^{(i)} \sim \tilde{q}\left(\mathbf{x}_{k+1}^{(i)}|\mathbf{x}_k^{(i)}, \mathbf{y}_{1:k+1}\right) = \frac{q\left(\mathbf{x}_{k+1}^{(i)}|\mathbf{y}_{1:k+1}\right)}{q\left(\mathbf{x}_k^{(i)}|\mathbf{y}_{1:k}\right)}; \tag{3.24}$$

and the unnormalized weights $\left\{w_k^{(i)}\right\}_{i=1}^N$ are updated as

$$\begin{aligned} w_{k+1}^{(i)} =& \frac{p\left(\mathbf{x}_{k+1}^{(i)}|\mathbf{y}_{1:k+1}\right)}{q\left(\mathbf{x}_{k+1}^{(i)}|\mathbf{y}_{1:k+1}\right)} \overset{(3.24)}{=} \frac{f\left(\mathbf{x}_{k+1}^{(i)}|\mathbf{x}_k^{(i)}\right) p\left(\mathbf{x}_k^{(i)}|\mathbf{y}_{1:k+1}\right)}{\tilde{q}\left(\mathbf{x}_{k+1}^{(i)}|\mathbf{x}_k, \mathbf{y}_{1:k+1}\right) q\left(\mathbf{x}_k^{(i)}|\mathbf{y}_{1:k}\right)} \\ \overset{(3.22)}{=}& w_k^{(i)} \frac{f\left(\mathbf{x}_{k+1}^{(i)}|\mathbf{x}_k^{(i)}\right) p\left(\mathbf{x}_k^{(i)}|\mathbf{y}_{1:k+1}\right)}{\tilde{q}\left(\mathbf{x}_{k+1}^{(i)}|\mathbf{x}_k, \mathbf{y}_{1:k+1}\right) p\left(\mathbf{x}_k^{(i)}|\mathbf{y}_{1:k}\right)} = w_k^{(i)} \frac{f\left(\mathbf{x}_{k+1}^{(i)}|\mathbf{x}_k^{(i)}\right) h\left(\mathbf{y}_{k+1}|\mathbf{x}_k^{(i)}\right)}{\tilde{q}\left(\mathbf{x}_{k+1}^{(i)}|\mathbf{x}_k, \mathbf{y}_{1:k+1}\right) l(\mathbf{y}_{k+1}|\mathbf{y}_{1:k})}, \end{aligned}$$

where $f(\mathbf{x}_k|\mathbf{x}_{k-1})$ and $h(\mathbf{y}_k|\mathbf{x}_k)$ are the transition density and the observation density, respectively, and $l(\mathbf{y}_k|\mathbf{y}_{1:k-1})$ is the predictive distribution of $\mathbf{y}_k$ given $\mathbf{y}_{1:k-1}$. $l(\mathbf{y}_k|\mathbf{y}_{1:k-1})$ is usually difficult to evaluate. But it does not depend on the state, and hence it is not necessary to be computed, since the weights will be renormalized as in (3.21). The algorithm described above is so-called SIS.

Although SIS achieves great success for short data records, it is doomed to fail in the long run, since the probability mass concentrated on a small portion of the samples after a few iteration steps, see [**22**]. The remedy is to resample the particles. The procedure surely will introduce some additional Monte Carlo variance, but in the long run it alleviates the accumulative error over time and help to eliminate the particle improverishment. The standard particle filtering algorithm is to resample the particles according to the normalized weights, and after that, the weights are reset to be $\frac{1}{N}$. In detail, the particles with small importance weights are eliminated; while those with large ones are replicated. This improved algorithm is referred as SIR.

The structure of the particle filter [**18**] can be summerized abstractly as follows:

1) *Mutation:* Draw for $i = 1, \ldots, N$,

$$\mathbf{x}_k^{(i)} \sim K_k\left(\hat{\mathbf{x}}_{k-1}^{(i)}, d\mathbf{x}_k\right),$$

where $\hat{\mathbf{x}}_k^{(i)}$ are the ith resampled particles at time step k, $K_k : X_{k-1} \to P(X_k)$ is a given probability kernel, and X_k is the sample space at time step k.

2) *Correction:* Assign weights to particles so that, for $i = 1, \ldots, N$,

$$w_k^{(i)} \propto \frac{p\left(\mathbf{x}_k^{(i)}|\mathbf{y}_{1:k}\right)}{\tilde{p}\left(\mathbf{x}_k^{(i)}|\mathbf{y}_{1:k-1}\right)},$$

where $\tilde{p}(\cdot|\mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})K_k(\mathbf{x}_{k-1},\cdot)d\mathbf{x}_{k-1}$.

3) *Selection:* Resample, according to a given selection scheme,

$$\left(\mathbf{x}_k^{(i)}, w_k^{(i)}\right)_{i=1}^N \rightarrow \left(\hat{\mathbf{x}}_k^{(i)}, 1\right)_{i=1}^N.$$

Various resampling strategies give different algorithms. Multinomial resampling [**30**] amounts to drawing N independent new particles from the multinomial distribution which produces $\left\{\mathbf{x}_k^{(i)}\right\}_{i=1}^N$ with the probability $\tilde{w}_k^{(i)}$, where $\tilde{w}_k^{(i)} := \frac{w_k^{(i)}}{\sum_{j=1}^N w_k^{(j)}}$ with $w_k^{(i)}$ defined in (3.22). Residual resampling [**54**] consists of reproducing $\left\lfloor N\tilde{w}_k^{(i)} \right\rfloor$ times each particle $\mathbf{x}_k^{(i)}$, where $\lfloor\cdot\rfloor$ stands for the integer part. The number of new particles need to draw from the multinomial distribution is $N_r = N - \sum_{i=1}^N \left\lfloor N\tilde{w}_k^{(i)} \right\rfloor$. This strategy yields N particles $\left\{\mathbf{x}_k^{(i)}\right\}_{i=1}^N$ with probability $\frac{N\tilde{w}_k^{(i)} - \left\lfloor N\tilde{w}_k^{(i)} \right\rfloor}{N_r}$. Systematic resampling [**12**, **17**] is the selection method such that the number of replicates of certain particle $\mathbf{x}_k^{(i)}$ with the probability in the range of $N\tilde{w}_k^{(i)} \pm 1$.

It has been discussed in [**16**, **18**] that to what extent (3.23) yields a good approximation of the expectation as the number of the particles N tends to infinity. The following theorem gives the central limit theorem of the PF with either multinomial resampling or residual resampling strategies.

Theorem 3.8 ([**18**])**.** *If the selection strategies are either multinomial resampling or residual resampling, and provided that the unit function $\mathbf{x}_k \mapsto 1$ belongs to $\Phi_k^{(1)}$ for every k, where $\Phi_k^{(d)}$ is the set of measurable functions $\varphi : X_k \rightarrow \mathbb{R}^d$ such that for some $\delta > 0$,*

$$\mathbb{E}_{p(\mathbf{x}_k|\mathbf{y}_{1:k-1})}||w_k \cdot \varphi||^{2+\delta} < +\infty,$$

where X_k is the sample space at time step k. Then for any $\varphi \in \Phi_k^{(d)}$, $\mathbb{E}_{p(\mathbf{x}_k|\mathbf{y}_{1:k-1})}(\varphi)$, $V_k(\varphi)$ and $\hat{V}_k(\varphi)$ are finite quantities, and the following convergences in distribution hold as $N \rightarrow +\infty$:

$$N^{\frac{1}{2}}\left\{\frac{\sum_{i=1}^N w_k^{(i)}\varphi\left(\mathbf{x}_k^{(i)}\right)}{\sum_{j=1}^N w_k^{(j)}} - \mathbb{E}_{p(\mathbf{x}_k|\mathbf{y}_{1:k})}(\varphi)\right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, V_k(\varphi));$$

$$N^{\frac{1}{2}}\left\{N^{-1}\sum_{i=1}^N \varphi\left(\hat{\mathbf{x}}_k^{(i)}\right) - \mathbb{E}_{p(\mathbf{x}_k|\mathbf{y}_{1:k})}(\varphi)\right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \hat{V}_k(\varphi)),$$

where

$$\begin{aligned}\tilde{V}_k(\varphi) =& \hat{V}_{k-1}\{\mathbb{E}_{K_k}(\varphi)\} + \mathbb{E}_{p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})}\{Var_{K_k}(\varphi)\};\\ V_k(\varphi) =& \tilde{V}_k\{w_k \cdot (\varphi - \mathbb{E}_{p(\mathbf{x}_k|\mathbf{y}_{1:k})}\varphi)\}.\end{aligned}$$

For multinomial resampling, we have

$$\hat{V}_k(\varphi) = V_k(\varphi) + Var_{p(\mathbf{x}_k|\mathbf{y}_{1:k})}(\varphi);$$

while for residual resampling, we have

$$\hat{V}_k(\varphi) = V_k(\varphi) + R_k(\varphi),$$

with

$$\begin{aligned}R_k(\varphi) =& \mathbb{E}_{\tilde{p}(\mathbf{x}_k|\mathbf{y}_{1:k-1})}\{r(w_k)\varphi\varphi'\}\\ &- \frac{1}{\mathbb{E}_{\tilde{p}(\mathbf{x}_k|\mathbf{y}_{1:k-1})}\{r(w_k)\}}\left[\mathbb{E}_{\tilde{p}(\mathbf{x}_k|\mathbf{y}_{1:k-1})}\{r(w_k)\varphi\}\right]\left[\mathbb{E}_{\tilde{p}(\mathbf{x}_k|\mathbf{y}_{1:k-1})}\{r(w_k)\varphi\}\right]',\end{aligned}$$

and $r(\cdot) = \cdot - \lfloor\cdot\rfloor$. The notation $\mathbb{E}_{K_k}(\varphi)$ and $Var_{K_k}(\varphi)$ are the short for $\mathbb{E}_{K_k(\mathbf{x}_{k-1},\cdot)}\{\varphi(\cdot)\}$ and $Var_{K_k(\mathbf{x}_{k-1},\cdot)}\{\varphi(\cdot)\}$, respectively.

3.2.3. Yau-Yau's method. Various numerical schemes to solve the PDEs can applied to (2.9) to obtain an approximation to the conditional pdf π. Yet, the main drawback of PDE methods are the intensive computation. It is almost impossible to achieve the real time performance. To overcome this shortcoming, the splitting-up algorithm is introduced to move the heavy computation off-line. It is like the Trotter product formula from semigroup theory. This operator splitting algorithm is proposed for the DMZ equation by Bensoussan, et al. [**9**]. More research articles follow this direction are [**31**, **60**, **39**] etc. In 1990s, Lototsky, et al. [**55**] developed a new algorithm (so-called S^3-algorithm) based on the Cameron-Martin version of Wiener chaos expansion. However, both the splitting-up method and the S^3-algorithm require the boundedness of the drifting term and the observation term (f and h in (2.2)), which leaves out even the linear case. To overcome this restriction, Yau and Yau [**75**] developed a real-time novel algorithm, called Yau-Yau's method, to solve the robust DMZ equation, where the boundedness of the drift term and observation term is replaced by some mild growth conditions on f and h. This algorithm has been further validated and applied to time-varying system in [**56**, **57**] i.e. f, h and g can be explicitly time-dependent. We report this method in this section.

Let us assume that we know the observation time sequence a-prior, and denote it as $\mathcal{P}_k = \{0 = \tau_0 < \tau_1 < \cdots < \tau_k = T\}$. But the observation data $\{\mathbf{y}_{\tau_i}\}$ at each sampling time τ_i, $i = 0, \cdots, k$ are unknown until the on-line experiment runs. We call the computation off-line if it can be performed without any on-line experimental data; otherwise, it is called on-line computations.

The robust DMZ equation of the model (2.2) in general form is given as following:

$$\begin{cases} \dfrac{\partial u}{\partial t}(t,\mathbf{x}) + \dfrac{\partial}{\partial t}\left(\mathbf{h}^T\mathbf{R}^{-1}\right)^T \mathbf{y}_t u(t,\mathbf{x}) \\ \qquad\qquad = \exp\left(-\mathbf{h}^T\mathbf{R}^{-1}\mathbf{y}_t\right)\left[L - \dfrac{1}{2}\mathbf{h}^T\mathbf{R}^{-1}\mathbf{h}\right]\exp\left(\mathbf{h}^T\mathbf{R}^{-1}\mathbf{y}_t\right)u(t,\mathbf{x}) \,, \\ u(0,\mathbf{x}) = \pi_0(\mathbf{x}). \end{cases} \tag{3.25}$$

where L is defined as

$$L(*) \equiv \frac{1}{2}\sum_{i,j=1}^{N_x} \frac{\partial^2}{\partial x_i \partial x_j}\left[\left(\mathbf{G}\mathbf{Q}\mathbf{G}^T\right)_{ij} *\right] - \sum_{i=1}^{N_x}\frac{\partial (f_i *)}{\partial x_i}, \tag{3.26}$$

by using the exponential transformation [**19**]

$$u(t,\mathbf{x}) = \exp\left[\mathbf{h}^T(t,\mathbf{x})\mathbf{R}^{-1}(t)\mathbf{y}_t\right]\pi(t,\mathbf{x}). \tag{3.27}$$

More explicitly, (3.25) can be expanded as

$$\begin{cases} \dfrac{\partial u}{\partial t}(t,\mathbf{x}) = \dfrac{1}{2}D_w^2 u(t,\mathbf{x}) + \mathbf{F}(t,\mathbf{x})\cdot\nabla u(t,\mathbf{x}) + J(t,\mathbf{x})u(t,\mathbf{x}) \\ u(0,\mathbf{x}) = \pi_0(\mathbf{x}), \end{cases} \tag{3.28}$$

where

$$D_w^2 = \sum_{i,j=1}^{N_x}(\mathbf{G}\mathbf{Q}\mathbf{G}^T)_{ij}\frac{\partial^2}{\partial x_i \partial x_j}, \tag{3.29}$$

$$\mathbf{F}(t,\mathbf{x}) = \left[\sum_{j=1}^{N_x}\frac{\partial}{\partial x_j}\left(\mathbf{G}\mathbf{Q}\mathbf{G}^T\right)_{ij} + \sum_{j=1}^{N_x}(\mathbf{G}\mathbf{Q}\mathbf{G}^T)_{ij}\frac{\partial K}{\partial x_j} - f_i\right]_{i=1}^{N_x}, \tag{3.30}$$

$$\begin{aligned} J(t,\mathbf{x}) = &-\frac{\partial}{\partial t}\left(\mathbf{h}^T\mathbf{R}^{-1}\right)^T\mathbf{y}(t) + \frac{1}{2}\sum_{i,j=1}^{N_x}\frac{\partial^2}{\partial x_i\partial x_j}\left(\mathbf{G}\mathbf{Q}\mathbf{G}^T\right)_{ij} + \sum_{i,j=1}^{N_x}\frac{\partial}{\partial x_i}\left(\mathbf{G}\mathbf{Q}\mathbf{G}^T\right)_{ij}\frac{\partial K}{\partial x_j} \\ &+ \frac{1}{2}\sum_{i,j=1}^{N_x}(\mathbf{G}\mathbf{Q}\mathbf{G}^T)_{ij}\left[\frac{\partial^2 K}{\partial x_i\partial x_j} + \frac{\partial K}{\partial x_i}\frac{\partial K}{\partial x_j}\right] - \sum_{i=1}^{N_x}\frac{\partial f_i}{\partial x_i} - \sum_{i=1}^{N_x} f_i\frac{\partial K}{\partial x_i} - \frac{1}{2}\left(\mathbf{h}^T\mathbf{R}^{-1}\mathbf{h}\right), \end{aligned} \tag{3.31}$$

in which

$$K(t,\mathbf{x}) = \mathbf{h}^T(t,\mathbf{x})\mathbf{R}^{-1}(t)\mathbf{y}_t. \tag{3.32}$$

Let $u_i(t,\mathbf{x})$ be the solution of the robust DMZ equation (3.12) with $\mathbf{y}_t$ freezed as the observation $\mathbf{y}_{\tau_{i-1}}$ on the interval $\tau_{i-1} \le t \le \tau_i$, $i = 1, 2, \cdots, k$

$$
\begin{cases}
\dfrac{\partial u_i}{\partial t}(t,\mathbf{x}) + \dfrac{\partial}{\partial t}\left(\mathbf{h}^T\mathbf{R}^{-1}\right)^T \mathbf{y}_{\tau_{i-1}} u_i(t,\mathbf{x}) \\
\qquad\qquad = \exp\left(-\mathbf{h}^T\mathbf{R}^{-1}\mathbf{y}_{\tau_{i-1}}\right)\left[L - \dfrac{1}{2}\mathbf{h}^T\mathbf{R}^{-1}\mathbf{h}\right]\exp\left(\mathbf{h}^T\mathbf{R}^{-1}\mathbf{y}_{\tau_{i-1}}\right) u_i(t,\mathbf{x}) \\
u_1(0,\mathbf{x}) = \pi_0(\mathbf{x}), \\
\text{or} \\
u_i(\tau_{i-1},\mathbf{x}) = u_{i-1}(\tau_{i-1},\mathbf{x}), \quad \text{for } i = 2,3,\cdots,k.
\end{cases} \tag{3.33}
$$

Define the norm of $\mathcal{P}_k$ by $|\mathcal{P}_k| = \sup_{1\le i\le k}(\tau_i - \tau_{i-1})$. It is shown in [**75**, **56**] that as $|\mathcal{P}_k| \to 0$, we have

$$\sum_{i=1}^{k} \chi_{[\tau_{i-1},\tau_i]}(t) u_i(t,\mathbf{x}) \to u(t,\mathbf{x})$$

in some sense, for all $0 \le t \le T$, where $u(t,\mathbf{x})$ is the exact solution of (3.25). For the conciseness of notation, let us denote

$$N(t,\mathbf{x}) \equiv -\frac{\partial}{\partial t}\left(\mathbf{h}^T\mathbf{R}^{-1}\right)\mathbf{y}_t - \frac{1}{2}D_w^2 K + \frac{1}{2}D_w K \cdot \nabla K - \mathbf{f}\cdot\nabla K - \frac{1}{2}\left(\mathbf{h}^T\mathbf{R}^{-1}\mathbf{h}\right), \tag{3.34}$$

The proof consists of two steps:

1) The exact solution $u(t,\mathbf{x})$ of the robust DMZ equation (3.25) is well approximated by u_R as $R \to \infty$, for any $t \in [0,T]$, where u_R is the solution to (3.25) restricted on B_R (the ball centered at the origin with the radius R) with Dirichlet boundary condition.

Theorem 3.9 ([**56**]). *For any $T > 0$, let $u(t,\mathbf{x})$ be a solution of the robust DMZ equation (3.28) in $[0,T]\times\mathbb{R}^n$. Let $R \gg 1$ and $u_R(t,\mathbf{x})$ be the solution to (3.25) restricted on B_R. Assume the following conditions are satisfied, for all $(t,\mathbf{x}) \in [0,T]\times\mathbb{R}^n$:*

- $N(t,\mathbf{x}) + \frac{3}{2}N_x \left|\left|\mathbf{GQG}^T\right|\right|_\infty + |\mathbf{f} - D_w K| \le C$,
- $e^{-\sqrt{1+|\mathbf{x}|^2}}\left[14N_x\left|\left|\mathbf{GQG}^T\right|\right|_\infty + 4|\mathbf{f} - D_w K|\right] \le \tilde{C}$,

where N and K are defined in (3.34) and (3.32), respectively, D_w is defined as

$$D_w * = \left[\sum_{j=1}^{N_x}\left(\mathbf{GQG}^T\right)_{ij}(t,\mathbf{x})\frac{\partial *}{\partial x_j}\right]_{i=1}^{N_x}, \tag{3.35}$$

and C, $\tilde{C}$ are generic constants possibly depending on T. Let $v = u - u_R$, then $v \ge 0$ for all $(t,\mathbf{x}) \in [0,T]\times B_R$ and

$$\int_{B_{\frac{R}{2}}} v(T,\mathbf{x}) \le \bar{C} e^{-\frac{9}{16}R}\int_{\mathbb{R}^{N_x}} e^{\sqrt{1+|\mathbf{x}|^2}}\pi_0(\mathbf{x}), \tag{3.36}$$

where $\bar{C}$ is some constant, which may depend on T.

2) $u_R(\tau,\mathbf{x})$ is well approximated by $u_{k,R}(\tau,\mathbf{x})$, as $k \to +\infty$, in the L^1 sense, where $u_{k,R}$ is described in detail in the theorem below.

Theorem 3.10 ([**56**]). *Let Ω be a bounded domain in $\mathbb{R}^n$. Assume that*

- $|N(t,\mathbf{x})| \le C$,
- *There exists some $\alpha \in (0,1)$, such that*

$$|N(t,\mathbf{x}) - N(t,\mathbf{x};\bar{t})| \le \tilde{C}|t-\bar{t}|^\alpha, \tag{3.37}$$

for all $(t,\mathbf{x}) \in [0,T]\times\Omega$, $\bar{t} \in [0,T]$, where $N(t,\mathbf{x})$ is in (3.34), and $N(t,\mathbf{x};\bar{t})$ denotes $N(t,\mathbf{x})$ with the observation $\mathbf{y}_t = \mathbf{y}_{\bar{t}}$. Let $u_\Omega(t,\mathbf{x})$ be the solution of (3.28) on $[0,T]\times\Omega$ with zero-Dirichlet boundary condition. For any $0 \le \tau \le T$, let $\mathcal{P}_k^\tau = \{0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_k = \tau\}$ be a partition of $[0,\tau]$, where $\tau_i = \frac{i\tau}{k}$. Let $u_{i,\Omega}(t,\mathbf{x})$ be the approximate solution obtained by

our algorithm restricted on $[\tau_{i-1},\tau_i]\times\Omega$. *That is,* $u_{i,\Omega}(t,\mathbf{x})$ *is the solution on* $\Omega\times[\tau_{i-1},\tau_i]$ *of the equation*

$$\begin{cases} \dfrac{\partial u_{i,\Omega}}{\partial t}(t,\mathbf{x}) = \dfrac{1}{2}D_w^2 u_{i,\Omega}(t,\mathbf{x}) + \mathbf{F}(t,\mathbf{x};\tau_{i-1})\cdot\nabla u_{i,\Omega}(t,\mathbf{x}) + J(t,\mathbf{x};\tau_{i-1})u_{i,\Omega}(x,t) \\ u_{i,\Omega}(\tau_{i-1},\mathbf{x}) = u_{i-1,\Omega}(\tau_{i-1},\mathbf{x}) \\ u_{i,\Omega}(t,\mathbf{x})|_{\partial\Omega} = 0, \end{cases} \tag{3.38}$$

for $i=1,2,\cdots,k$, *with* $u_{1,\Omega}(0,\mathbf{x})=\pi_{0,\Omega}(\mathbf{x})$. *Here,* $\mathbf{F}(t,\mathbf{x};\tau_{i-1})$, $J(t,\mathbf{x};\tau_{i-1})$ *denote* $\mathbf{F}(t,\mathbf{x})$, $J(t,\mathbf{x})$ *with the observation* $\mathbf{y}_t=\mathbf{y}_{\tau_{i-1}}$, *respectively. Then*

$$u_\Omega(\tau,\mathbf{x}) = \lim_{k\to\infty} u_{k,\Omega}(\tau,\mathbf{x}),$$

in the L^1 *sense in space and the following estimate holds:*

$$\int_\Omega |u_\Omega - u_{k,\Omega}|(\tau,\mathbf{x}) \le \frac{\bar{C}}{k^\alpha}, \tag{3.39}$$

where $C,\tilde{C},\bar{C}$ *are generic constants, possibly depending on* T, $\int_\Omega \sigma_{0,\Omega}$. *The right-hand side of* (3.39) *tends to zero as* $k\to\infty$.

Generally speaking, it is impractical to solve (3.33) in the real-time manner, since the on-line data $\{\mathbf{y}_{\tau_i}\}$, $i=1,\cdots,k$, are in the coefficients of (3.33). We have to numerically solve the time-consuming PDE on-line, every time after the new observation data coming in. Yet, the proposition below helps to move the heavy computations off-line. This is the key ingredient of the Yau-Yau's method in [**75**, **56**].

Proposition 3.1 ([**75**, **56**]). *For each* $\tau_{i-1}\le t<\tau_i$, $i=1,2,\cdots,k$, $u_i(t,\mathbf{x})$ *satisfies* (3.33) *if and only if*

$$\rho_i(t,\mathbf{x}) = \exp\left[\mathbf{h}^T(t,\mathbf{x})\mathbf{R}^{-1}(t)\mathbf{y}_{\tau_{i-1}}\right]u_i(t,\mathbf{x}), \tag{3.40}$$

satisfies the Kolmogorov forward equation (KFE)

$$\frac{\partial\rho_i}{\partial t}(t,\mathbf{x}) = \left(L - \frac{1}{2}\mathbf{h}^T\mathbf{R}^{-1}\mathbf{h}\right)\rho_i(t,\mathbf{x}), \tag{3.41}$$

where L *is defined in* (3.26).

It is clear that (3.41) is independent of the observation path $\{\mathbf{y}_{\tau_i}\}_{i=0}^k$, and the transformation between u_i and ρ_i is one-to-one. It is also not hard to see that (3.41) could be numerically solved beforehand. Let us denote $U(t) := L - \frac{1}{2}\mathbf{h}^T\mathbf{R}^{-1}\mathbf{h}$ for short to emphasize its time-dependence. Under certain conditions, $\{U(t)\}_{t\in[0,T]}$ forms a family of strong elliptic operators. Furthermore, the operator $U(t): D(U(t))\subset L^2\left(\mathbb{R}^{N_x}\right)\to L^2\left(\mathbb{R}^{N_x}\right)$ is the infinitesimal generator of the two-parameter semigroup $\mathcal{U}(t,\tau)$, for $t\ge\tau$. In particular, with the observation time sequence known $\{\tau_i\}_{i=1}^k$, we obtain a sequence of two-parameter semigroup $\{\mathcal{U}(t,\tau_{i-1})\}_{i=1}^k$, for $\tau_{i-1}\le t<\tau_i$. Let us take the initial conditions of KFE (3.41) at $t=\tau_i$ as a set of complete orthonormal base in $L^2\left(\mathbb{R}^{N_x}\right)$, say $\{\phi_l(x)\}_{l=1}^\infty$. We pre-compute the solutions of (3.41) at time $t=\tau_{i+1}$, denoted as $\{\mathcal{U}(\tau_{i+1},\tau_i)\phi_l\}_{l=1}^\infty$. These data should be stored in preparation of the on-line computations.

The on-line computation in our algorithm consists of two parts at each time step τ_{i-1}, $i=1,\cdots,k$.

- Project the initial condition $\rho_i(\tau_{i-1},\mathbf{x})\in L^2\left(\mathbb{R}^{N_x}\right)$ at $t=\tau_{i-1}$ onto the base $\{\phi_l(\mathbf{x})\}_{l=1}^\infty$, i.e., $\rho_i(\tau_{i-1},\mathbf{x})=\sum_{l=1}^\infty\hat{\rho}_{i,l}\phi_l(\mathbf{x})$. Hence, the solution to (3.41) at $t=\tau_i$ can be expressed as

$$\rho_i(\tau_i,\mathbf{x}) = \mathcal{U}(\tau_i,\tau_{i-1})\rho_i(\tau_{i-1},\mathbf{x}) = \sum_{l=1}^\infty \hat{\rho}_{i,l}\left[\mathcal{U}(\tau_i,\tau_{i-1})\phi_l(\mathbf{x})\right], \tag{3.42}$$

  where $\{\mathcal{U}(\tau_i,\tau_{i-1})\phi_l(\mathbf{x})\}_{l=1}^\infty$ have already been computed off-line.

- Update the initial condition of (3.41) at τ_i with the new observation $\mathbf{y}_{\tau_i}$. Let us specify the observation updates (the initial condition of (3.41)) for each time step. For $0 \le t \le \tau_1$, the initial condition is $\rho_1(x,0) = \pi_0(x)$. At time $t = \tau_1$, when the observation $\mathbf{y}_{\tau_1}$ is available,

$$\rho_2(\tau_1, \mathbf{x}) \overset{(3.40)}{=} \exp\left[\mathbf{h}^T(\tau_1, \mathbf{x})\mathbf{R}^{-1}(\tau_1)\mathbf{y}(\tau_1)\right] u_2(\tau_1, \mathbf{x})$$
$$\overset{(3.40),(3.33)}{=} \exp\left[\mathbf{h}^T(\tau_1, \mathbf{x})\mathbf{R}^{-1}(\tau_1)\mathbf{y}(\tau_1)\right] \rho_1(\tau_1, \mathbf{x}),$$

with the fact $\mathbf{y}_0 = 0$. Here, $\rho_1(\tau_1, \mathbf{x}) = \sum_{l=1}^{\infty} \hat{\rho}_{1,l} \left[\mathcal{U}(\tau_1, 0)\phi_l(\mathbf{x})\right]$, where $\{\hat{\rho}_{1,l}\}_{l=1}^{\infty}$ is computed in the previous step, and $\{\mathcal{U}(\tau_1, 0)\phi_l(\mathbf{x})\}_{l=1}^{\infty}$ are prepared by off-line computations. Hence, we obtain the initial condition $\rho_2(\tau_1, \mathbf{x})$ of (3.41) for the next time interval $\tau_1 \le t \le \tau_2$. Recursively, the initial condition of (3.41) for $\tau_{i-1} \le t \le \tau_i$ is

$$\rho_i(\tau_{i-1}, \mathbf{x}) = \exp\left[\mathbf{h}^T(\tau_{i-1}, \mathbf{x})\mathbf{R}^{-1}(\tau_{i-1})(\mathbf{y}_{\tau_{i-1}} - \mathbf{y}_{\tau_{i-2}})\right] \cdot \rho_{i-1}(\tau_{i-1}, \mathbf{x}), \tag{3.43}$$

for $i = 2, 3, \cdots, k$, where $\rho_{i-1}(\tau_{i-1}, \mathbf{x}) = \sum_{l=1}^{\infty} \hat{\rho}_{i-2,l} \left[\mathcal{U}(\tau_{i-1}, \tau_{i-2})\phi_l(\mathbf{x})\right]$.

The approximation of $u(t, \mathbf{x})$, denoted as $\tilde{u}(t, \mathbf{x})$, is obtained

$$\tilde{u}(t, \mathbf{x}) = \sum_{i=1}^{k} \chi_{[\tau_{i-1}, \tau_i]}(t) u_i(t, \mathbf{x}), \tag{3.44}$$

where $u_i(t, \mathbf{x})$ is obtained from $\rho_i(t, \mathbf{x})$ by (3.40). And $\pi(x, t)$ could be recovered by (3.27).

In [**57**], the algorithm suggested in [**75**, **56**] has been applied to several 1D NLF problems, and the results have been compared with the EKF and the PF both in accuracy and in real-time manner. The basis functions of $L^2(\mathbb{R})$ are chosen to be the generalized Hermite functions $\{H_n^{\alpha,\beta}(x)\}_{n=0}^{N}$, where $\alpha > 0$ and β are the scaling factor and the translating factor, respectively. When applying to the high-dimensional NLF problems, the curse of dimensionality is arisen. To tackle this difficulty in some degree, Yau and the author [**58**] investigate to solve the KFE by using the sparse-grid algorithm [**66**]. This shed a light on applying the Yau-Yau's method to high-dimensional NLF problems.

4. Conclusion and future work

In this survey, starting from the KF, we briefly go through the local approaches, including EKF, GSF, QKF and etc. The Bayesian interpretation of KF is somewhat clear from [**35**]. In this survey, we briefly sketch the re-derivation of KF from DMZ equation according to [**26**]. Emphases have been put on the existing three major global approaches: finite-dimensional filter, sequential Monte Carlo methods (particle filter) and the Yau-Yau's method.

The study of finite-dimensional filter starts from 1980s. It is well-known that there exits finite-dimensional estimator for certain type of NLF problem, say [**8**] and there also exists essentially infinite-dimensional one [**34**]. Thus, the natural question is to ask for the borderline. From the viewpoint of estimation algebra, Yau gave the complete classification for the estimation algebra with maximal rank [**69**, **70**] and some partial results on those with non-maximal rank [**73**, **15**, **68**]. The complete classification of general estimation algebra is still wide open. The greatest benefit from the classfication is to construct numerous novel finite-dimensional filters.

The sequential Monte Carlo methods is nowadays one of the most popular methods in industry. The derivation from the prediction-correction recursion has been included in this survey. Also, the convergence of the PF with multinomial and residual resampling strategy has been stated rigorously. The performance of PF can be improved further with carefully chosen the instruction distribution and experienced resampling strategies. However, the PF can never achieve the real-time performance due to its nature of Monte Carlo simulations.

The Yau-Yau's method is the most recent algorithm [**75**, **56**, **57**] in solving directly the posterior pdf. The real-time performance is guaranteed for NLF problems with medium low dimensional states [**56**, **76**], and no further assumptions on the function's type (say Gaussian). The further investigations can be carried on how to apply to high-dimensional state NLF problem and break the so-called "curse of dimension" in certain degree. A possible way-out is to combine the sparse-grid algorithm [**58**]. More efforts are needed in this direction, if in aim to solve real applications.

Acknowledgements

The author greatly acknowledges the support of Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry, the Fundamental Research Funds for the Central Universities (YWF-14-RSC-026). And this work is also partially supported by Beijing Natural Science Foundation (1154011) and National Natural Science Foundation of China (11471184).

References

[1] D. F. Allinger and S. K. Mitter, *New results in innovations problem for nonlinear filtering*, Stochastics **4**, pp. 339-348, 1981.

[2] D. L. Alspach and H. W. Sorenson, *Nonlinear Bayesian estimation using Gaussian sum approximation*, IEEE Trans. Automat. Control **17**(4), pp. 439-448, 1972.

[3] J. L. Anderson, *An ensemble adjustment Kalman filter for data assimilation*, Month. Weather Rev. **129**, pp. 2884-2903, 2001.

[4] I. Arasaratnam and S. Haykin, *Cubature Kalman filters*, IEEE Trans. Automat. Control **54**(6), pp. 1254-1269, 2009.

[5] I. Arasaratnam, S. Haykin and R. J. Elliott, *Discrete-time nonlinear filtering algorithms using Gauss-Hermite quadrature*, Proceedings of the IEEE **95**(5), pp. 953-977, 2007.

[6] I. Arasaratnam, S. Haykin and T. R. Hurd, *Cubature Kalman filtering for continuous-discrete systems: theory and simulations*, IEEE Trans. Signal Process. **58**(10), pp. 4977-4993, 2010.

[7] T. R. Bayes, *Essay towards solving a problem in the doctrine of chances*, Phil. Trans. Roy. Soc. Lond. **53**, pp. 370-418, 1763. Repreinted in Biometrika **45**, 1958.

[8] V. Beneš, *Exact finite-dimensional for certain diffusions with nonlinear drift*, Stochastics **5**(1/2), pp. 65-92, 1981.

[9] A. Bensoussan, R. Glowinski and A. Răşcanu, *Approximation of the Zakai equation by the splitting up method*, SIAM J. Control Optim. **28**(6), pp. 1420-1431, 1990.

[10] D. Brigo, B. Hanzon and F. LeGland, *A differential geometric approach to nonlinear filtering: the projection filter*, IEEE Trans. Automat. Control **43**(2), pp. 247-252, 1998.

[11] R. W. Brockett and J. M. C. Clark, *The geometry of the conditional density functions*. In Analysis and Optimization of Stochastic Systems, Jacobs et al., eds., Academic Press, New York, pp. 299-309, 1980.

[12] J. Carpenter, P. Clifford and P. Fearnhead, *Improved particle filter for nonlinear problems*, IEE Proc. Radar Sonar Navigation **146**, pp. 2-7, 1999.

[13] Z. Chen, *Bayesian Filtering: From Kalman filters to particle filters, and beyond*, Technical report, McMaster University, Canada, 2003.

[14] J. Chen, *On uniquity of Yau filters*, Proceedings of the American Control Conference(ACC), Baltimore, MD, pp. 252-254, 1994.

[15] W.-L. Chiou, W.-R. Chiueh and S. S.-T. Yau, *Structure theorem for five-dimensional estimation algebras*, Syst. Contr. Lett. **55**, pp. 275-281, 2006.

[16] D. Crisan and A. Doucet, *A survey of convergence results on particle filtering methods for practioners*, IEEE Trans. Signal Processing **50**(3), pp. 736-746, 2002.

[17] D. Crisan and T. Lyons, *Minimal entropy approximations and optimal algorithms*, Monte Carlo Methods Appl. **8**, pp. 343-355, 2002.

[18] N. Chopin, *Central limit theorem for sequential monte carlo methods and its application to bayesian inference*, Ann. Statist. **32**(6), pp. 2385-2411, 2004.

[19] M. H. A. Davis, *On the multiplicative functional transformation arising in nonlinear filtering theory*, Z. Wahrsch. Verw. Gebiete **54**, pp. 125-139, 1980.

[20] A. Doucet, N. de Freitas and N. J. Gordon, eds., Sequential Monte Carlo Methods in Practice. Springer, New York, 2001.

[21] P. Del Moral, *Nonlinear filtering: interacting particle solution*, Markov Process. Related Fields **2**, pp. 555-279, 1996.

[22] P. Del Moral and J. Jacod, *Interacting particle filtering with discrete-time observations: Asymptotic behaviour in the Gaussian case*, in Stochastics in Finite and Infinite Dimensions: In Honor of Gopinath Kallianpur, T. Hida, R. L. Karandikar, H. Kunita, B. S. Rajput, S. Watanabe and J. Xiong, eds., Birkhäuser, Boston, MA, pp. 101-122, 2001.

[23] J. L. Doob, Stochastic Processes. Wiley, New York, 1953.

[24] T. E. Duncan, *Probability densityies for diffusion processes with applications to nonlinear filtering theory*, Ph. D. thesis, Stanford University, 1967.

[25] R. J. Elliott and V. Krishnamurthy, *New finite dimensional filters for parameter estimation of discrete time linear Gaussian models*, IEEE Trans. Automat. Control **44**, pp. 938-951, 1998.

[26] R. J. Elliott and S. Haykin, *A Zakai equation derivation of the extended Kalman filter*, Automatica J. IFAC **46**, pp. 620-624, 2010.

[27] G. Evensen, *Sequential data assimilation with a noninear quasi-geostrophic model using Monte Carlo methods to forecast error statistics*, J. Geophys. Res. **99**, pp. 143-162, 1994.

[28] R. Frey and W. Runggaldier, *A nonlinear filtering approach to volatility estimation with a view towards high frequency data*, Internat. J. Theoret. Appl. Finance **4**, pp. 199–210, 2001.

[29] A. Gelb, Applied Optimal Estimation. MIT Press, Cambridge, MA, 1974.

[30] N. Gordon, D. Salmond and A. F. Smith, *Novel approach to nonlinear/non-Gaussian Bayesian state estimation*, IEE Proc. F, Radar Signal Process. **140**, pp. 107-113, 1993.

[31] I. Gyongy and N. Krylov, *On the splitting-up method and stochastic partial differential equation*, Ann. Probab. **31**, pp. 564-591, 2003.

[32] I. Gyongy, *Approximations of stochastic partial differential equations*, in Stochastic Partial Differential Equations. Vol. 277, Lecture Notes in Pure and Applied Mathematics, pp. 287-307, Dekker, New York, 2002.

[33] J. Handschin, *Monte Carlo techniques for prediction and filtering of non-linear stochastic processes*, Automatica J. IFAC **6**, pp. 555-563, 1970.

[34] M. Hazewinkel, S. I. Marcus and H. J. Sussmann, *Nonexistence of finite-dimensional filters for conditional statistics of the cubic sensor problem*, Syst. Contr. Lett. **3**, pp. 331-340, 1983.

[35] Y. C. Ho and R. C. K. Lee, *A Bayesian approach to problems in stochastic estimation and control*, IEEE Trans. Automat. Control **9**, pp. 333-339, 1964.

[36] P. L. Houtekamer and H. L. Mitchell, *Data assimilation using an ensemble Kalman filter technique*, Month. Weather Rev. **126**, pp. 796-811, 1998.

[37] G.-Q. Hu and S. S.-T. Yau, *Finite-dimensional filters with nonlinear drift XV: New direct method for construction of universal finite-dimensional filter*, IEEE Trans. Aerospace Electron. Systems **38**(1), pp. 50-57, 2001.

[38] K. Ito and K. Xiong, *Gaussian filters for nonlinear filtering problems*, IEEE Trans. Automat. Control **45**(5), pp. 910-927, 2000.

[39] K. Ito, Approximation of the Zakai equation for nonlinear filtering, SIAM J. Contr. Optim. **34**, pp. 620-634, 1996.

[40] A. H. Jazwinski, Stochastic Processes and Filtering Theory. Academic Press, New York, 1970.

[41] B. Jia, M. Xin and Y. Cheng, *Sparse-grid quadrature nonlinear filtering*, Automatica J. IFAC **48**, pp. 327-341, 2012.

[42] S. Julier and J. Uhlmann, *Unscented filtering and nonlinear estimation*, Proceedings of the IEEE **92**(3), pp. 401-422, 2004.

[43] S. Julier, J. Uhlmann and H. F. Durrant-Whyte, *A new method for nonlinear transformation of means and covariances in filters and estimators*, IEEE Trans. Automat. Control **45**(3), pp. 477-482, 2000.

[44] T. Kailath, *A view of three decades of linear filtering theory*, IEEE Trans. Inform. Theory **20**(2), pp. 146-181, 1974.

[45] R. E. Kalman, *A new approach to linear filtering and prediction problem*, Trans. ASME, Ser. D, J. Basic Eng. **82**, pp. 34-45, 1960.

[46] R. E. Kalman and R. S. Bucy, *New results in linear filtering and prediction theory*, Trans. ASME, Ser. D, J. Basic Eng. **83**, pp. 95-107, 1961.

[47] G. Kitagawa, *Monte-Carlo filter and smoother for non-Gaussian nonlinear state space models*, J. Comput. Graph. Statist. **1**, pp. 1-25, 1996.

[48] A. N. Kolmogorov, *Interpolation and extrapolation of stationary random sequences*, Izv. Akad. Nauk USSR, ser. Math. **5**(5), pp. 3-14, 1941.

[49] R. Kulhavy, *Recursive nonlinear estimation: a geometric approach*, Automatica J. IFAC **26**(3), pp. 545-555, 1990.

[50] H. J. Kusher, *Dynamical equations for optimal nonlinear filtering*, J. Differential Equations **3**, pp. 179-190, 1967.

[51] H. J. Kusher and A. S. Budhiraja, *A nonlinear filtering algorithm based on an approximation of the conditional distribution*, IEEE Trans. Automat. Control **45**(3), pp. 580-585, 2000.

[52] J.-R. Künsch, *State space and hidden Markov models*. In Complex Stochastic Systems, O. E. Barndorff-Nielsen, D. R. Cox and C. Klüppelberg, eds., Chapman and Hall, London, pp. 109-173, 2001.

[53] H. Kunsch, S. Geman and A. Kehagias, *Hidden Markov random fields*, Ann. Appl. Probab. **5**, pp. 577-602, 1995.

[54] J. Liu and R. Chen, *Sequential Monte Carlo methods for dynamic systems*, J. Amer. Statist. Assoc. **93**, pp. 1032-1044, 1998.

[55] S. V. Lototsky, R. Mikulevicius and B. L. Rozovskii, *Nonlinear filtering revisited: A spectral approach*, SIAM J. Control Optim. **35**, pp. 435-461, 1997.

[56] X. Luo and S. S.-T. Yau, *Complete real time solution of the general nonlinear filtering problem without memory*, IEEE Trans. Automat. Control **58**(10), pp. 2563-2578, 2013.

[57] X. Luo and S. S.-T. Yau, *Hermite spectral method to 1d forward Kolmogorov equation and its application to nonlinear filtering problems*, IEEE Trans. Automat. Control **58**(10), pp. 2495-2507, 2013.

[58] X. Luo and S. S.-T. Yau, *Hermite spectral method with hyperbolic cross approximations to high-dimensional parabolic PDEs*, SIAM J. Numer. Anal. **51**(6), pp. 3186-3212, 2013.

[59] R. E. Mortensen, *Optimal control of continuous time stochastic systems*, Ph. D. thesis, Univeristy of California, Berkeley, CA, 1966.

[60] N. Nagase, *Remarks on nonlinear stochastic partial differential equations: An application of the splitting-up method*, SIAM J. Control Optim. **33**, pp. 1716-1730, 1995.

[61] D. Ocone, *Topics in nonlinear filtering theory*, Ph. D. thesis, Massachusetts Institute of Technology, MA, 1980.

[62] B. Oksendal, Stochastic Differential Equations, 5th edition. Springer-Verlag, Berlin, 1998.

[63] A. Rasoulian and S. S.-T. Yau, *Finite-dimensional filters with nonlinear drift IX: Construction of finite-dimensional estimation algebra of non-maximal rank*, Syst. Contr. Lett. **30**, pp. 109-118, 1997.

[64] H. E. Rauch, T. Tung and T. Striebel, *Maximum likelihood estimations of linear dynamic system*, AIAA J. **3**, pp. 1445-1450, 1965.

[65] D. B. Rubin, *A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when the fraction of missing information is modest: the SIR algorithm*, J. Am. Statist. Assoc. **82**, pp. 543-546, 1987.

[66] S. A. Smolyak, *Quadrature and interpolation formulas for tensor products of certain classes of functions*, Soviet Math. Dokl. **4**, pp. 240-243, 1963.

[67] N. Wiener, Extrapolation, Interpolation and Smoothing of Time Series, with Engineering Applications. Wiley, New York, 1949.

[68] X. Wu and S. S.-T. Yau, *Classification of estimation algebras with state dimension 2*, SIAM J. Control Optim. **45**(3), pp. 1039-1073, 2006.

[69] S. S.-T. Yau, *Finite dimensional filters with nonlinear drift I: A class of filters including both Kalman-Bucy filters and Beneš filters*, J. Math. Systems Estim. Control **4**, pp. 181-203, 1994.

[70] S. S.-T. Yau, *Complete classification of finite-dimensional estimation algebras of maximal rank*, Internat. J. Control **76**(7), pp. 657-677, 2003.

[71] S. S.-T. Yau and G.-Q. Hu, *Finite-dimensional filters with nonlinear drift X: Explicit solution of DMZ equation*, IEEE Trans. Automat. Control **46**(1), pp. 142-148, 2001.

[72] S. S.-T. Yau and G.-Q. Hu, *Classification of finite-dimensional estimation algebras of maximal rank with arbitrary state-space dimension and Mitter conjecture*, Internat. J. Control **78**(10), pp. 689-705, 2005.

[73] S. S.-T. Yau and A. Rasoulian, *Classification of four-dimensional estimation algebras*, IEEE Trans. Automat. Control **44**, pp. 2312-2318, 1999.

[74] S. S.-T. Yau and S. T. Yau, *Finite dimensional filters with nonlinear drift III. Duncan-Mortensen-Zakai equation with arbitrary initial condition for linear filtering system and the Beneš filtering system*, IEEE Trans. Aerospace Electron. Systems **33**, pp. 1277-1294, 1997.

[75] S. T. Yau and S. S.-T. Yau, *Real time solution of nonlinear filtering problem with out memory II*, SIAM J. Control Optim. **47**(1), pp. 163-195, 2008.

[76] M.-H. Yueh, W.-W. Lin and S. T. Yau, *An efficient algorithm of Yau-Yau method for solving nonlinear filtering problems*, Commun. Inf. Syst. **14**(2), pp. 111-134, 2014.

[77] M. Zakai, *On the optimal filtering of diffusion processes*, Z. Wahrsch. Verw. Gebiete **11**(3), pp. 230-243, 1969.

School of Mathematics and System Sciences, Beihang University, Beijing, P. R. China 100191. email: xluo@buaa.edu.cn