Extended Direct Method: A Time-Varying Kolmogorov Equation Approach for Infinite-Dimensional Optimal Filtering Problems

Xiaopei Jiao, Member, IEEE
Beijing Institute of Mathematical Sciences and Applications (BIMSA), Beijing, China

Ji Shi, Member, IEEE
Capital Normal University, Beijing, China

Stephen S.-T. Yau, Life Fellow, IEEE
Beijing Institute of Mathematical Sciences and Applications (BIMSA), Beijing, China,
Tsinghua University, Beijing, China

***Abstract*— One of the most important breakthroughs was the propose of the Kalman filter in the 1960s, which motivated a large number of studies on nonlinear filtering. Among these efforts, a special class of systems known as Yau filters has attracted attention. For the finite-dimensional case, the Direct Method (DM) was proposed to address systems with quadratic potential quantity [1, 2]. To generalize the DM, we shall focus on a class of infinite-dimensional filtering systems in a statistical sense. Our approach is capable of handling the potential quantity of arbitrary polynomial degree of filtering systems, meanwhile accommodating systems with nonlinear observation terms expressed as arbitrary-degree polynomials. The core idea is to apply several transformations to simplify the fundamental kernel corresponding to the filtering system. By leveraging a series expansion, we derive an explicit series solution from which optimal state estimates are obtained via convolution. Building on this series solution, we introduce a truncation technique and develop the Extended Direct Method (EDM) filtering algorithm which is supported by strong theoretical guarantees of uniqueness and existence through Euler operator theory. Numerical experiments demonstrate that our method not only generalizes the classical Direct Method to a wider class of systems but also enables fast, real-time implementation. It achieves near-optimal performance in challenging filtering scenarios compared to the Galerkin-Spectral method with less computational cost time, significantly outperforming conventional algorithms, including the Particle Filter, Extended Kalman Filter etc.**

***Index Terms*— State estimation, Infinite dimensional filter, Time-varying Kolmogorov equation, Fundamental solution, Kernel method.**

Xiaopei Jiao and Ji Shi equally contributed to this paper. This work is supported by the National Natural Science Foundation of China Funding (No.12501613, No.12101426, No.42450242) grant and Tsinghua University Education Foundation.

Authors' addresses: Xiaopei Jiao is with Beijing Institute of Mathematical Sciences and Applications (BIMSA), Beijing 101400, China (e-mail: jiaoxiaopei@bimsa.cn); Ji Shi is with Academy for Multidisciplinary Studies, Capital Normal University, Beijing 100048, China (e-mail: shiji@cnu.edu.cn); Stephen S.-T. Yau is Beijing Key Laboratory of Topological Statistics and Applications for Complex Systems, Beijing Institute of Mathematical Sciences and Applications (BIMSA), Beijing 101408, P. R. China, Department of Mathematical Sciences, Tsinghua University, Beijing 100084, P. R. China (e-mail: yau@uic.edu). (Corresponding author: Stephen S.-T. Yau.)

0018-9251 © IEEE

I. Introduction

Filtering methods are usually employed to estimate the state of a stochastic system by utilizing noisy observational data. Up to now, there have been intensive applications across various areas such as orbit determination, satellite positioning, etc. A significant breakthrough emerged in the early 1960s, when Kalman and collaborators introduced the Kalman Filter (KF) [3, 4], offering optimal estimation techniques for linear dynamical systems. However, for the general nonlinear system, things are not easy. In the minimal variance sense, the optimal state estimation is described by the expectation value of the state variable conditioned on past observation data. The conditional density provides complete information about the system's state, which is the central goal of the filtering problem.

Through more than 60 years of efforts by numerous researchers in this area, we have plenty of effective methods at hand to solve various application problems nowadays. These methods can be categorized into three classes. Among the most important and widely used approaches are the various Kalman Filter extensions, including the well-known Extended Kalman Filter (EKF)[5], the Unscented Kalman Filter (UKF)[6], and the Ensemble Kalman Filter (EnKF)[7], along with other variants presented in[8–10]. Since these algorithms are based on KF, they all share the same assumption that the conditional density $\rho(t,x)$ follows a Gaussian distribution, which brings the merit of easy implementation and the drawback of insufficient approximation ability for complex nonlinear systems. To overcome this strict limitation, some works switch to approximate $\rho(t,x)$ instead, such as the famous particle filter (PF)-like methods and Gaussian sum filters (GSF). The former ones utilize the principle of Monte-Carlo simulation by evolving a set of particles to reconstruct the real density. The latter ones use a mixture of Gaussian densities to approximate $\rho(t,x)$. The performance of these two algorithms strongly depends on the number of particles and Gaussian components. The larger the number of components is, the higher the accuracy the algorithms will attain. However, the Monte Carlo-based methods suffer from "particle degeneracy" and the GSF faces the challenge of optimizing a bunch of parameters in a short period. Notably, there are also well-established methods that aim to overcome the "particle degeneracy" from the feedback control perspective [11].

The third class focuses on computing the density function of the state directly. In recent years, many studies have been following this approach. The first work describing the dynamics of density was the Kushner-Stratonovich equation (K-S) proposed in the late 1960s. To address the complexities of solving the K-S equation, Independently, Duncan [12], Mortensen [13], and Zakai [14] introduced what is now known as the Duncan-Mortensen-Zakai (DMZ) equation, which governs the evolution of the unnormalized conditional density $\sigma(t,x)$ [15]. The DMZ equation is more tractable than the original K-S equations since it is a linear stochastic partial differential equation (PDE). In [16], Yau et al. studied the general Kolmogorov equation under the assumption of zero-valued observations and derived its corresponding semigroup kernel. In the early 21st century, Yau et al. established the Yau-Yau filtering framework, which has had a significant impact in the field of nonlinear filtering through a series of seminal works [17, 18]. Briefly, this framework transforms the computation of DMZ equations into a series of Kolmogorov equations and separates the filtering process into online and offline parts. The advantages of this framework lie in that it is widely applicable, real-time, and memoryless, and its convergence is theoretically guaranteed. Under this framework, several practical algorithms [19–22] were proposed.

Focusing on Yau filters, in [23, 24], the fundamental solution of the Kolmogorov forward equation was expressed as the transition probability amplitude of an Euclidean quantum mechanical system up to a gauge transformation. However, this approach holds rigorously only when the drift functions are bounded. Hu et al. [25] considered a gauge transformation to connect the DMZ equation and the Kolmogorov equation with additional sets of ordinary differential equations (ODEs). Later, Shi et al. extended this strategy to the Yau system with nonlinear observation of linear growth and introduced a Gaussian approximation strategy. When restricting ourselves to Yau filters, FKE can be reduced to the time-varying Kolmogorov equation (TVKE) with less computational complexity. Finite-dimensional filters (FDF) represent a special class of systems described by finite statistics. Under this framework of TVSE, they correspond to quadratic potential functions. From the fundamental semigroup kernel, explicit iterative filter algorithms have been proposed for time-invariant and time-variant systems, respectively, [2, 26]. This series of works involving explicit representation of the Zakai equation is named "Direct method (DM)". One important limitation of DM so far is that it only works for density estimation of FDF.

The algorithm proposed in this paper will generalize DM to handle a class of infinite-dimensional filters (IDFs). The basic setting is assumed as the Yau filtering system in which the drift term consists of an affine function and a gradient term. The potential in TVKE $q(t,x)$ satisfies the **Assumptions** in section II.C, allowing us to incorporate a larger class of systems with nonlinear observation of arbitrary degree. Starting from the fundamental kernel of TVKE, we shall take a series of equivalent transformation and get the explicit series solution. By truncating this series in a different order, we shall design a novel and very efficient filtering method, named "Extended Direct Method" (EDM), to solve state conditional density. Besides, we have provided rigorous theoretical results on the existence and uniqueness of the fundamental kernel of the filtering systems.

The novelty of this paper consists of the following aspects:

1) The proposed EDM can deal with systems with infinite-dimensional statistics, which is more general than the traditional Direct Method.
2) The EDM method has a flexible architecture for implementation. According to the selection of truncation order, the fundamental kernel solver can provide solutions with different precision for practical use.
3) Numerical experiments have demonstrated the efficiency of the EDM method. While attaining the same high accuracy as the optimal estimate provided by the spectral method (SM), it significantly outperform the traditional EKF and PF, and it can save a lot of computational loads compared to the SM and PF.
4) The rigorous theoretical guarantee has been provided in terms of the existence and uniqueness of the fundamental kernel of the filters.

The structure of this paper is listed as follows. Basic concepts of nonlinear filter, Yau-Yau framework, and necessary assumptions are introduced in Section II. In Section III, an explicit series representation of the TVKE is constructed. Following this, a novel numerical algorithm and detailed theoretical guarantee are shown. In Section IV, numerical experiments and three typical nonlinear filtering examples are implemented to show the high precision and efficiency of our method. Section V states the conclusion of the whole paper. Detailed proofs of theoretical results will be put in the Appendix.

II. Basic concepts and preliminaries

A. Nonlinear filtering system

This work focuses on continuous nonlinear filtering systems, which are governed by the following stochastic differential equations:

$$\begin{cases} dx(t) = f(x(t))dt + g(x(t))dw(t), \ x(0) = x_0 \in \mathbb{R}^n, \\ dy(t) = h(x(t))dt + dv(t), \ y(0) \in \mathbb{R}^m, \end{cases} \tag{1}$$

where $x(t) = (x_1, \cdots, x_n)^\top$ and $y(t) = (y_1, \cdots, y_m)^\top$ denote state and observation vectors respectively. $f(x) : \mathbb{R}^n \to \mathbb{R}^n$ and $h : \mathbb{R}^n \to \mathbb{R}^m$ are usually called the drift and observation (measurement) terms. $g : \mathbb{R}^n \to \mathbb{R}^{n\times p}$ describes the diffusion of the state influenced by stochastic noise, and it is assumed as an orthogonal matrix. Brownian motions $w(t) \in \mathbb{R}^p, v(t) \in \mathbb{R}^m$ are assumed independent and equiped with positive definite covariance

I_n and S_n. We denote $\mathcal{Y}_t := \{y_s : 0 \le s \le t\}$ as collection of history measurements.

B. The time-varying Kolmogorov equation for Yau filtering system

In a continuous filtering system, the primary objective is to compute the conditional expectation $\mathbb{E}[x_t \mid \mathcal{F}_t]$, which provides an optimal estimate for state under least square error criterion. Therefore, once we get the conditional density, optimal state estimation and all its moment statistics are computable. In the following, we shall recall the Yau-Yau filtering procedure and its relation with the time-varying Kolmogorov equation briefly.

To begin with, the filtering system admits an unnormalized conditional density $\sigma(t,x)$, which satisfies the corresponding DMZ equation. By the following gauge transformation

$$u(t,x) = \exp[-h^T(x,t)S_t^{-1}y_t]\sigma(t,x). \tag{2}$$

It can be demonstrated that $u(t,x)$ satisfies the robust DMZ equation, a deterministic PDE.

Second, we divide the time interval into a partition $\mathcal{P} := 0 = t_0 < t_1 < \cdots < t_N = T$ and approximate the observation y_t by holding it constant as $y_{t_{k-1}}$ over each subinterval $t \in [t_{k-1}, t_k)$. Let $u_k(t,x)$ denote the solution to the robust DMZ equation under this frozen observation within the given interval. By making an invertible exponential transformation

$$\tilde{u}_k(t,x) = \exp[h^T(x,t)S_t^{-1}y_{t_{k-1}}]u_k(t,x), \tag{3}$$

$u_k(t,x)$ is associated with $\tilde{u}_k$ which satisfies an observation-independent Kolmogorov equation. Inspired by this property, the Yau-Yau framework splits the filtering process into two parts: the Kolmogorov equation is solved offline in advance, and then state estimation is finished online with offline results.

The first important work relating the filtering problem with the Kolmogorov equation was done by Yau and Yau [26]. They showed that for the Yau filtering system, $\tilde{u}_k(t,x)$ can be simplified by further transformation:

$$\tilde{u}_k(t,x)e^{-\phi(x)} = v_k(t,\tilde{x}), \tag{4}$$

where $\tilde{x} = B(t)x + b(t)$ and $B(t) = e^{-Lt}, b(t) = -\int_0^t e^{-Ls} l ds$. Here $v_k(t,\tilde{x})$ satisfies time-varying Kolmogorov equation (TVKE):

$$\begin{cases} \dfrac{\partial v_k}{\partial t}(t,\tilde{x}) = \dfrac{1}{2}\displaystyle\sum_{i=1}^n \dfrac{\partial^2}{\partial \tilde{x}_i^2} v_k(t,\tilde{x}) - \dfrac{1}{2}q(t,\tilde{x})v_k(t,\tilde{x}), \\ v_k(t_{k-1},\tilde{x}) = \exp((y_{t_{k-1}} - y_{t_{k-2}})S_{t_{k-1}}^{-1}\tilde{h}(\tilde{x}))v_{k-1}(t_{k-1},\tilde{x}), \end{cases} \tag{5}$$

where $\tilde{h}(\tilde{x}) = h(B(t)^{-1}\tilde{x} - B(t)^{-1}b(t))$ and $q(t,\tilde{x}) = q(t, B(t)^{-1}\tilde{x} - B(t)^{-1}b(t))$ with

$$q(\cdot,x) := \Delta\phi + |\nabla\phi|^2 + 2(Lx+l)\cdot\nabla\phi + h^T S^{-1} h + 2tr(L). \tag{6}$$

After obtaining the solution $v_k(t,\tilde{x})$ to equation (5), the state estimate over the interval $[t_{k-1}, t_k]$ can be approximated by

$$\hat{x}_t := \mathbb{E}[x_t|\mathcal{F}_t] \approx \frac{\int x\tilde{u}_k(t,x)dx}{\int \tilde{u}_k(t,x)dx}. \tag{7}$$

Note that the effectiveness of the approximation has been theoretically proved by Yau and Yau in [17, 18].

C. Assumptions

In this paper, we make the following assumptions:

1) For simplicity, we consider the diffusion coefficient $g(\cdot)$ be an orthogonal matrix, i.e., $g \cdot g^\top = I_n$ is an identity matrix.
2) The drift function is given by $f(x) = Lx + l + \nabla\phi(x)$, where $L \in \mathbb{R}^{n\times n}$ and $l \in \mathbb{R}^n$ are constant matrix and vector, respectively. The function ϕ is differentiable.
3) The potential function $q(t,x)$ in (6) is a polynomial of degree m in variable x.

III. Method

The central objective of this paper is to develop an efficient algorithm for solving the TVKE equation (5) defined over each time interval, enabling the construction of a real-time filtering method. To streamline the notation, we drop the subscript and represent $v_k(t,\tilde{x})$ simply as $v(t,x)$.

$$\begin{cases} \dfrac{\partial v}{\partial t}(t,x) = \dfrac{1}{2}\displaystyle\sum_{i=1}^n \dfrac{\partial^2}{\partial x_i^2} v(t,x) - \dfrac{1}{2}q(t,x)v(t,x), \\ v(0,x) = v_0(x), \end{cases} \tag{8}$$

where $(t,x) \in [0,T]\times\mathbb{R}^n$, and $q(t,x) : [0,T]\times\mathbb{R}^n \to \mathbb{R}$ is called potential function. As a powerful tool to solve the preceding Cauchy initial value problem (8), we shall consider the corresponding fundamental solution:

$$\begin{cases} \dfrac{\partial K}{\partial t}(t,x,y) = \dfrac{1}{2}\displaystyle\sum_{i=1}^n \dfrac{\partial^2}{\partial x_i^2} K(t,x,y) - \dfrac{1}{2}q(t,x)K(t,x,y), \\ \displaystyle\lim_{t\to 0}\int_{\mathbb{R}^n} K(t,x,y)\psi(y)dy = \psi(x), \quad \text{for any } \psi \in C^\infty(\mathbb{R}^n). \end{cases} \tag{9}$$

where $(t,x,y) \in [0,T]\times\mathbb{R}^n\times\mathbb{R}^n$. Note that the second equation is often expressed as $\lim_{t\to 0} K(t,x,y) = \delta(x-y)$, where δ denotes the standard Dirac delta function.

Once a fundamental solution is obtained, the solution can be directly solved using convolution.

$$v(t,x) = \int_{\mathbb{R}^n} K(t,x,y)v_0(y)dy. \tag{10}$$

A. Construction of explicit representation

In the following, our task is to simplify the computation of the original fundamental kernel $K(t,x,y)$ by introducing a series of reversible transformations.

An important point to note is that the initial distribution of $K(t,x,y)$ contains a singularity. Due to the challenge of representing the δ function in a numerical context, this poses significant difficulties in designing any accurate numerical implementation. To simplify the solution of (9) and avoid the direct representation of the δ function, we first apply the following reversible exponential transformation.

$$K(t,x,y) = (2\pi t)^{-\frac{n}{2}} \times \exp(W(t,x,y)). \tag{11}$$

A direct computation gives the dynamics of $W(t,x,y)$ as follows. Detailed proof will be found in the appendix.

PROPOSITION III.1. *Evolution of $W(t,x,y)$ satisfies*

$$\begin{aligned}
&\frac{\partial W}{\partial t}(t,x,y) = \frac{1}{2}\Delta_x W + \frac{1}{2}|\nabla_x W|^2 - \frac{1}{2}q(t,x) + \frac{n}{2t},\\
&\lim_{t\to 0}(2\pi t)^{-\frac{n}{2}}\exp(W(t,x,y)) = \delta(x-y).
\end{aligned} \tag{12}$$

where $\nabla_x = [\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \cdots, \frac{\partial}{\partial x_n}]^\top$.

Next, we apply the following transformation by separating an additional singular term.

$$W(t,x,y) = -\frac{\|x-y\|^2}{2t} + \tilde{W}(t,x,y). \tag{13}$$

Consequently, one can obtain the governing equation for the evolution of $\tilde{W}(t,x,y)$. Detailed proof will be found in the Appendix.

PROPOSITION III.2. *Evolution of $\tilde{W}(t,x,y)$ satisfies*

$$\begin{aligned}
\frac{\partial \tilde{W}}{\partial t}(t,x,y) =& \frac{1}{2}\Delta_x \tilde{W} + \frac{1}{2}|\nabla_x \tilde{W}|^2\\
&+ \frac{1}{t}(y-x)\cdot\nabla_x\tilde{W} - \frac{1}{2}q(t,x),\\
\lim_{t\to 0+} \tilde{W}&(t,x,y) = 0.
\end{aligned} \tag{14}$$

In what follows, we proceed by applying the inverse of the exponential transformation.

$$\phi(t,x,y) = \exp(\tilde{W}(t,x,y)), \tag{15}$$

we can transform the nonlinear PDE (14) into a linear parabolic PDE. A detailed proof is provided in the appendix.

PROPOSITION III.3. *$\phi(t,x,y)$ satisfies the following linear equation without singular initial,*

$$\begin{aligned}
&\frac{\partial \phi}{\partial t}(t,x,y) = \frac{1}{2}\Delta_x\phi - \frac{1}{2}q(t,x)\phi + \frac{1}{t}(y-x)\cdot\nabla_x\phi,\\
&\phi(t,x,y)|_{t\to 0+} = 1.
\end{aligned} \tag{16}$$

REMARK 1. *It is particularly noteworthy that the singularity term contained in the initial condition of (9) has been transferred to the gradient term in (16) as $t \to 0^+$. This transformation enables the design of more efficient numerical algorithms. Eq. (16) can be referred to as the transformed TVKE with a singular gradient term.*

In summary, the relationship of these intermediate functions can be illustrated as the following computational flow:

$$\sigma \xleftarrow[\text{Eq. (2)-(4)}]{} v_k \xleftarrow[\text{Eq. (8)}]{} K \xleftarrow[\text{Eq. (9)}]{} W \xleftarrow[\text{Eq. (11)}]{} \tilde{W} \xleftarrow[\text{Eq. (13)}]{} \phi \tag{15}$$

Up to now, computation of conditional density has been transformed to ϕ for which we shall design an explicit algorithm later.

Following Assumption (3) in Section II.C, this is equivalent to that $q(t,x)$ is a degree m polynomial with time-dependent coefficients. Since $B(t)$ and $b(t)$ are both analytic in the neighborhood of the origin $t=0$, the time-dependent coefficients in $q(t,x)$ are also analytic and can be expanded as a power series in t. This implies that

$$q(t,x) = q_0(x) + \sum_{n=1}^{\infty} q_n(x)t^n, \tag{17}$$

where $q_n(x)$ are polynomials of degree no greater than m.

In the following, $\phi(t,x,y)$ is assumed to have an asymptotic series expansion in terms of t.

$$\phi(t,x,y) = 1 + \sum_{k=1}^{\infty}\phi_k(x,y)t^k, \tag{18}$$

Substituting such series representation ϕ back to Eq. (16), we shall get constraints satisfied by components

$$\begin{aligned}
&\phi_1 + \sum_{k=1}^{\infty}(k+1)\phi_{k+1}(x,y)t^k = \left[-\frac{1}{2}q_0(x) + (y-x)\cdot\nabla_x\phi_1\right]\\
&+\sum_{k=1}^{\infty}\left[\frac{1}{2}\Delta_x\phi_k - \frac{1}{2}\sum_{i=0}^{k}q_i(x)\phi_{k-i} + (y-x)\cdot\nabla_x\phi_{k+1}\right]t^k.
\end{aligned} \tag{19}$$

By comparing the coefficients on both sides, we shall get

$$\begin{cases}
(x-y)\cdot\nabla_x\phi_1(x,y) + \phi_1(x,y) = -\frac{1}{2}q_0(x),\\
(x-y)\cdot\nabla_x\phi_{k+1}(x,y) + (k+1)\phi_{k+1}(x,y)\\
= \frac{1}{2}\Delta_x\phi_k(x,y) - \frac{1}{2}\sum_{i=0}^{k}q_i(x)\phi_{k-i}(x,y), \quad k\geq 1.
\end{cases} \tag{20}$$

These are several first-order PDEs satisfied by the series components ϕ_k. By assuming that these components are polynomials of an appropriate degree, the explicit form of ϕ_k is presented in the following result, with a full proof available in the Appendix.

THEOREM III.1. *Under assumptions in Section C, we shall assume*

$$-\frac{1}{2}q_k(z+y) := \sum_{|\alpha|=0}^{m} q_{k\alpha}(y)z^\alpha, \quad \forall k\geq 0, \tag{21}$$

where $z^\alpha := z_1^{\alpha_1}\cdots z_n^{\alpha_n}$ and $\alpha = (\alpha_1,\alpha_2,\cdots,\alpha_n)$ denotes the multi-index, with the norm defined as $|\alpha| := \sum_{i=1}^n \alpha_i$. The term $q_{k\alpha}(y)$ is a polynomial of y with a degree no greater than $m-|\alpha|$ for $k\geq 0$. The explicit solution of Eq. (20) can be expressed as

$$\phi_k(x,y) := \sum_{|\alpha|=0}^{km}\phi_{k\alpha}(y)(x-y)^\alpha, \quad k\geq 1 \tag{22}$$

where the coefficients of each component are given as

$$\phi_{1\alpha}(y) = \frac{q_{0\alpha}(y)}{|\alpha|+1}, \quad 0\leq|\alpha|\leq m, \tag{23}$$

and

$$\phi_{(k+1)\alpha}(y) = \begin{cases} \frac{1}{|\alpha|+k+1}[\sum_{i=1}^{n}(\alpha_i+2)(\alpha_i+1) \\ \times\phi_{k(\alpha_1,\cdots,\alpha_i+2,\cdots,\alpha_n)} \\ +\sum_{|\beta|=|\alpha|-m}^{|\alpha|}\sum_{i=0}^{k} q_{i(\alpha-\beta)}\phi_{(k-i)\beta}], \\ for \quad 0\le|\alpha|\le km-2, \\ \frac{1}{|\alpha|+k+1}\left[\sum_{|\beta|=|\alpha|-m}^{|\alpha|}\sum_{i=0}^{k} q_{i(\alpha-\beta)}\phi_{(k-i)\beta}\right], \\ for \quad km-1\le|\alpha|\le(k+1)m. \end{cases} \tag{24}$$

Especially, the first-order component can be written down

$$\phi_1(x,y) := \sum_{|\alpha|=0}^{m} \frac{q_{0\alpha}(y)}{|\alpha|+1}(x-y)^\alpha \tag{25}$$

In summary, we shall get the following closed-form solution:

$$\phi(t,x,y) = 1+\sum_{k=1}^{\infty}\sum_{|\alpha|=0}^{km}\phi_{k\alpha}(y)(x-y)^\alpha t^k. \tag{26}$$

with coefficients $\phi_{k\alpha}(\cdot)$ given by Eq. (23) and Eq. (24). Recall we have the computational flow (Eq. (15)), the original fundamental solution can be determined explicitly.

$$\begin{aligned} K(t,x,y) &= (2\pi t)^{-\frac{n}{2}}\exp\left(-\frac{\|x-y\|^2}{2t}\right) \\ &\times(1+\sum_{k=1}^{\infty}\sum_{|\alpha|=0}^{km}\phi_{k\alpha}(y)(x-y)^\alpha t^k). \end{aligned} \tag{27}$$

REMARK 2. *It is worth emphasizing that this closed-form expression characterizes the asymptotic behavior as $t\to 0$, which aids in the efficient numerical approximation of the fundamental solution over short time intervals.*

B. Numerical algorithm

After obtaining an explicit series solution of the transformed TVKE, the truncation technique can be applied to derive an efficient approximation of the fundamental solution to the TVKE. In this subsection, we introduce a new filtering algorithm constructed using the proposed truncation approach.

More precisely, as stated in Theorem III.1, the first-order truncation of Eq. (25) can be used to approximate the fundamental solution:

$$\phi(t,x,y) = 1+\sum_{|\alpha|=0}^{m}\frac{q_{0\alpha}(y)}{|\alpha|+1}(x-y)^\alpha t+\mathcal{O}(t^2), \tag{28}$$

and

$$\begin{aligned} K(t,x,y) &= (2\pi t)^{-\frac{n}{2}}\exp\left(-\frac{\|x-y\|^2}{2t}\right) \\ &\times(1+\sum_{|\alpha|=0}^{m}\frac{q_{0\alpha}(y)}{|\alpha|+1}(x-y)^\alpha t+\mathcal{O}(t^2)). \end{aligned} \tag{29}$$

Once we have obtained an efficient approximation of the fundamental solution, the TVKE (5) can be solved quickly by convolving it with the time-varying initial distribution. By applying the equivalent relation between the forward Kolmogorov equation and the TVKE through a reversible gauge transformation, nonlinear filtering can be successfully performed. Since our new method extends the traditional "Direct Method", we will refer to it as the "Extended Direct Method". The detailed numerical implementation and procedure are provided in the following pseudo-code.

Algorithm 1 Extended Direct Method (EDM)

Input: Time interval partition $\mathcal{P} := \{0=\tau_0<\tau_1<\cdots<\tau_N=T\}$. Initial state distribution σ_0. **Output:** State estimation $\hat{x}_t := \mathbb{E}[x_t|\mathcal{F}_t]$ at observation time instants $\{\tau_i\}|_{i=1}^N$.

1: Calculate fundamental kernel $K(t,x,y)$ based on (6) (17) (23) and (29) sequentially and save in an offline manner.

2: For: $i=1,2,...,N$ **do**

3: Prediction by applying convolution $v_i(\tau_i,\tilde{x}) = \int v_i(\tau_{i-1},\tilde{y})K_t(\tilde{x},\tilde{y})d\tilde{y}$.

4: Update by using new arrived observation $v_i(\tau_{i-1},\tilde{x}) = v_{i-1}(\tau_{i-1},\tilde{x})\exp((y_{\tau_{i-1}}-y_{\tau_{i-2}})S^{-1}_{\tau_{i-1}}\tilde{h}(\tilde{x}))$.

5: Return state estimation $\hat{x}_t = \frac{\int(B(t)^{-1}\tilde{x}-B(t)^{-1}b(t))e^{\phi(B(t)^{-1}\tilde{x}-B(t)^{-1}b(t))}v_i(t,\tilde{x})d\tilde{x}}{\int e^{\phi(B(t)^{-1}\tilde{x}-B(t)^{-1}b(t))}v_i(t,\tilde{x})d\tilde{x}}$ at time $t=\tau_i$.

6: End for.

REMARK 3. *The Extended Direct Method (EDM) is not based on discretizing a PDE (such as the Zakai or Kushner equation) in time and space, which would indeed require a separate stability analysis of the numerical scheme. Instead, EDM is derived from a closed-form analytical series expansion of the fundamental solution (kernel) evaluated at a set of quadrature/sampling points. In this sense, EDM is closer to a semianalytical or functional approximation method rather than a numerical PDE solver (like finite difference or finite element methods). The evolution is embedded in the analytical structure of the kernel itself.*

C. Theoretical properties of TVKE

In this subsection, we examine the theoretical property of the transformed Kolmogorov equation (16), under specific assumptions on the potential function $q(t,x)$. Note that this parabolic PDE depends on the gradient term with unbounded coefficients. As a result, conventional analytical techniques for PDEs are not directly applicable. It is important to recognize that (20) defines a system of recursive relations. The sequence ϕ_k can be iteratively determined by solving a series of first-order PDEs. Once ϕ_k is determined, we will obtain $\phi(t,x,y)$ in closed form. Therefore, the existence of $\phi_t(x,y)$ is equivalent to the existence of the function series ϕ_k.

Notice that $(x-y)\cdot\nabla_x$ is an Euler operator in the variable x. The system of first-order PDEs (20) can be referred to as "Euler operator-type PDEs". This class of first-order partial differential equations has been extensively studied by Yau et al through multiple publications on finite-dimensional filtering theory [27–29]. In the analysis that follows, we employ the Taylor asymptotic expansion along with the Euler operator to address and overcome the challenges introduced by the unbounded gradient term.

In the following, we present an intermediate result from Euler operator theory that proves instrumental in determining the degree of the solution polynomial. A detailed proof is provided in the appendix.

LEMMA III.2 ([30]). *Let $E_S := \sum_{l\in S} x_l \frac{\partial}{\partial x_l}$ be an Euler operator, where S is a subset of index $\{1,2,\cdots,n\}$. $P_k(x)$ denotes the set of polynomials of degree no more than k in variable $x_1,\cdots,x_n$. Assume $\zeta \in C^\infty(\mathbb{R}^n)$ and m is a positive constant. If $E_S(\zeta)+m\zeta \in P_k(x)$, then $\zeta \in P_k(x)$.*

The next two results establish the existence and uniqueness of solutions to the transformed TVKE (16). Complete proofs are provided in the Appendix.

THEOREM III.3. *Consider a polynomial function $f(x,y)$ of degree m (where $m \geq 1$), and let s denote a strictly positive constant. The following PDE,*

$$(x-y)\cdot\nabla_x\zeta(x,y)+s\zeta(x,y)=f(x,y), \qquad (30)$$

admits a unique smooth solution $\zeta(x,y)$ which is also a polynomial of degree m.

THEOREM III.4. *Under assumptions in Section C, Eq. (20) admits a unique solution with the polynomial form. More precisely, $\phi_k(x,y)$ is a polynomial of degree no greater than km.*

IV. Numerical Experiments

A. Fundamental kernel solution

In this subsection, a (1+2)-dimensional PDE will be solved using our newly proposed extended direct method to examine its effectiveness in solving the fundamental equation.

$$\begin{cases}\dfrac{\partial K}{\partial t}(t,x,y)=\dfrac{1}{2}\dfrac{\partial^2}{\partial x^2}K-\dfrac{1}{2}(a_0+a_1x+a_2x^2)K,\\ K(0,x,y)=\delta(x-y)\end{cases} \qquad (31)$$

The coefficients are chosen as $[a_0,a_1,a_2] = [0.15,-0.1,1]$. The spatial computational region is set to $[-5,5]^2$.

The direct evaluation of the fundamental kernel implemented by the EDM method (i.e., Eq. (29)) is shown in Table I. We shall find that the RMSE of our method shall converge further with Δt decreasing. It is noted that for a relatively small observation interval $\Delta t \sim$1e-1, our method already has enough accuracy of $\sim$1e-3.

TABLE I: Convergence behavior of EDM method on solving fundamental solution. Here standard value is computed by the Legendre-Galerkin method [31]. ($a=0.15, b=-0.1, c=1$, # sample = 1500, # basis = 15)

Δt	RMSE (EDM)
0.1	7.113e-03
0.06	3.160e-03
0.03	1.478e-03
0.01	9.918e-04
0.006	9.573e-04
0.003	9.460e-04

B. Numerical convergence of TVKE

In order to verify the convergence behavior over time, we implement our method to solve the TVKE.

$$\frac{\partial v}{\partial t}(t,x)=\frac{1}{2}\frac{\partial^2}{\partial x^2}v(t,x)-\frac{1}{2}(1+x^6)v(t,x)$$

where the initial $v(0,x)$ is generated by applying **random generalized orthogonal polynomials**. Let $M>0$ and $P_i(x)$ denote i-th generalized orthogonal polynomial basis. Therefore, the N_v-dimensional generalized orthogonal polynomial space is defined as

$$V_{\text{org}}=\left\{\sum_{i=0}^{N_v-1}\lambda_iP_i(x)\middle||\lambda_i|\leq M\right\}$$

We generate the samples of initial $v(0,x)$ from the space V_{org} by using random sampling of the coefficient $\lambda_i \in [-M,M]$. In the following, we shall take $P_i(x)$ as generalized Legendre polynomials. We specify the domain of the coefficient $M=1$ and dimension $N_v=15$ for the test.

Convergence behavior with different observation interval $T=\Delta t$ is also investigated, including computational cost and RMSE in Table II. Corresponding log-log illustration of cost time and RMSE of our EDM method is shown in Fig. 1. Compared with time-stepping type methods [32, 33], our EDM method takes minimal computational time of $\sim$1e-3, which is not sensitive to Δt. This is due to the EDM method, which presents an explicit kernel structure and does not need any further time-stepping procedure. The time stepping method such as Krylov subspace, Magnus method and Finite difference method have less computational complexity with a smaller simulation time $T=\Delta t$.

In order to quantify the convergence order of our method, we present log-log plot of RMSE in Fig. 1. RMSE implemented by our EDM method with $N=1$ shows that basically EDM method has linear convergence order indicated by red reference line, i.e.,

$$|v(T,\cdot)-v_{\text{EDM}}(T,\cdot)|\leq C\Delta t$$

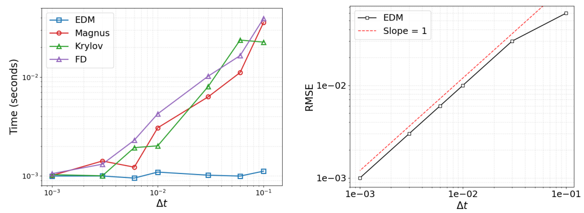


Fig. 1: Time cost and Convergence behavior on observation interval. (a) Consumed computational time of EDM method, Magnus approximation, and Krylov subspace algorithm. (b) RMSE behavior of the EDM method. Red dashed line denotes the reference of linear convergence order.

TABLE II: Convergence behavior (RMSE) over different observation intervals for EDM method, Krylov subspace algorithms, and Magnus approximation. (Spatial discretization 150×150) The true solution is implemented by Finite-difference method.

Δt	RMSE(EDM)	RMSE(Magnus)	RMSE(Krylov)
0.1	5.153e-02	2.755e-03	6.892e-04
0.06	1.607e-02	9.439e-04	2.133e-04
0.03	8.074e-03	1.261e-03	2.870e-04
0.01	6.929e-03	6.895e-04	1.264e-04
0.006	6.673e-03	1.152e-04	1.510e-05
0.003	6.675e-03	1.128e-04	1.171e-05
0.001	6.669e-03	3.276e-05	1.546e-06

C. Nonlinear filtering examples

This section presents numerical experiments on three filtering examples to evaluate the effectiveness and accuracy of the proposed "extended direct method". The first example is a one-dimensional nonlinear system with a Benes-type drift term and cubic measurement. The second example consists of a quadratic polynomial as the drift function and a two-dimensional highly oscillatory observation term. In the third example, we will implement a two-dimensional cubic system with linear drift functions and cubic measurement functions. All three examples considered here can be classified as infinite-dimensional systems in the sense of sufficient statistics, implying that the conditional density cannot be fully characterized using a finite set of statistical parameters. The potential functions in all examples are polynomials of degree greater than 2, which cannot be effectively addressed by the traditional Gaussian approximation-based "Direct Method" [1].

Through these three typical infinite-dimensional filters, our novel method demonstrates greater generality than the "Direct Method." More importantly, the newly proposed EDM not only approaches the nearly optimal estimate provided by the Legendre-Galerkin SM but also has a significantly lower computational load compared to the SM. All 3 benchmark filter systems will be tested by Extended Kalman Filter (EKF), Sequential Importance Resampling Particle Filter (PF), Legendre-Galerkin Spectral Method (SM), and Extended Direct Method (EDM). For the state track, the domain within one standard deviation (STD) is presented as a bandwidth region.

To assess the accuracy of various algorithms, standard evaluation metrics such as the mean square error (MSE) and the averaged mean square error (MMSE) over multiple experimental trials are commonly used.

$$\begin{aligned} \mathrm{MSE}(t_k) &:= \frac{1}{M}\sum_{i=1}^{M}(\hat{X}^i_{t_k} - X^i_{t_k})^2, \\ \mathrm{MMSE} &:= \frac{1}{N}\sum_{k=1}^{N}\mathrm{MSE}(t_k), \end{aligned} \tag{32}$$

where $\hat{X}^i_t$ denotes the state estimation given by the algorithm at the i-th trial at observation time t_k, where $k = 1, \dots, N$. X^i_t represents the corresponding true state value. The total number of simulated trials is denoted as M. The mean time (MT) is defined as MT $= \frac{1}{M}\sum_{i=1}^{M} T_i$, where T_i is the computational time for the i-th trial.

EXAMPLE IV.1 (1D cubic system).

$$\begin{cases} dX_t = \tanh(X_t)dt + dW_t, \quad \mathbb{E}[(dW_t)^2] = dt, \\ dZ_t = H_0 X_t^3 dt + dV_t, \quad \mathbb{E}[(dV_t)^2] = sdt, \\ X_0 \sim \sigma_0 := 0.3\mathcal{N}(0.5, 0.1) + 0.7\mathcal{N}(-0.5, 0.2). \end{cases} \tag{33}$$

The basic setting of the parameters is listed as follows: The observation variance is $S = 0.015$, and the total simulation time is $T = 2$ s. The observation coefficient is $H_0 = 0.5$. The observation time increment is $\Delta t = 0.005$. To test the robustness and accuracy, we repeat the same state tracking simulation 20 times, then calculate the MMSE and MT. For the SM, the basis functions are chosen to be Legendre polynomials, with the order up to $m = 8$. The scaling factor is $C = 3.0$. For the PF, we use 20 particles to approximate the empirical density. For the extended direct method (EDM), the computational domain is set as $(x, y) \in [-3, 3] \times [-3, 3]$. A first-order approximation is used to calculate the transformed fundamental solution $\phi(t, x, y)$. The potential function $q(x)$ in this example is a polynomial of degree 6, taking the following form:

$$q(x) = 1 + \frac{H_0^2 x^6}{s} \tag{34}$$

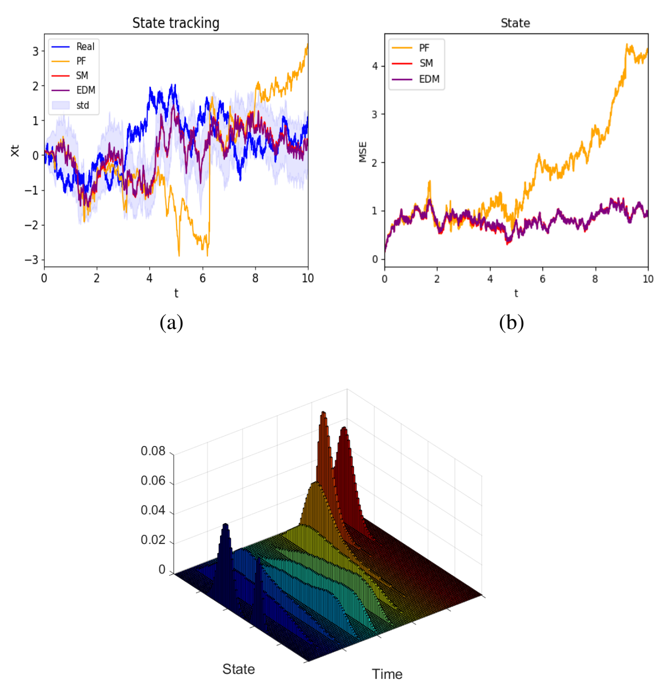
The corresponding forward Kolmogorov equation is expressed as a (1+1)-dimensional parabolic PDE:

$$\frac{\partial \rho}{\partial t} = \frac{1}{2}\frac{\partial^2 \rho}{\partial x^2} - \tanh(x)\frac{\partial \rho}{\partial x} - (\mathrm{sech}(x) + \frac{1}{2}\frac{H_0^2 x^6}{s})\rho(t, x) \tag{35}$$

In order to verify the performance of handling multi-modal distribution, we design the initial density to be a Mixed Gaussian. EDM can successfully track the two different modes during iteration (Fig. 2).

It is observed that the EKF either fails to provide valid state estimations or suffers from numerical overflow in this example. One fundamental reason for EKF's failure is that it is designed based on the first two moments of the density, whereas the cubic system has infinite-dimensional statistics. The PF generally provides reasonable estimations, but with lower precision compared to the SM, despite having a shorter runtime. According to Table III, both the SM and EDM significantly outperform the PF, with the MMSE being 32.5% lower than that of the PF. In terms of mean computational time (MT), the EDM demonstrates the lowest computational cost, being 84% faster than the SM and 73.6% faster than the PF. Fig. 2 shows the state estimate and one STD reliable region together. The nonlinear structure present in the drift and observation makes recovering the true state from the corresponding observation data a moderately challenging task. For such problems, the EDM method exhibits superior performance in handling multi-modal state distribution in both computational efficiency and MMSE.

REMARK 4. *Our method (EDM) is fundamentally not a Gaussian-assuming filter like the EKF. By construction, its solution is derived from a polynomial series expansion of the fundamental kernel, allowing it to represent and track multiple modes in the posterior density, provided they are expressible within the approximation order.*



(a) (b) (c)

Fig. 2: State track with 1 STD (upper left), mean square error (upper right), and conditional density (lower middle) in Example IV.1.

For the convergence behavior of Legendre-Galerkin Spectral method, we show the corresponding results of computational time and RMSE (filter) under different

Algorithms	PF	SM	EDM
MT	3.143	2.559	**0.830**
MMSE	1.703	**0.806**	0.814

TABLE III: Performance of all simulated algorithms in Example IV.1. The bold numbers represent the lowest value in terms of the two measurement metrics, MT and MMSE.

basis numbers in Table IV. It is clear to see that even though computational time will significantly increase with the greater basis number, RMSE will converge.

TABLE IV: Convergence behavior of Legendre-Galerkin Spectral method.

# basis	Time	RMSE(filter)
8	1.590	0.841
12	8.629	0.773
16	55.902	0.709
20	910.624	0.702

EXAMPLE IV.2 (2D highly oscillatory system).

$$\begin{cases} dX_t = (a_0 + a_1 X_t + a_2 X_t^2)dt + dW_t, \ \mathbb{E}[dW_t dW_t^T] = dt \\ dZ_t = H_0 \begin{bmatrix} X_t \sin(X_t) \\ X_t \cos(X_t) \end{bmatrix} dt + dV_t, \ \mathbb{E}[dV_t dV_t^T] = sI_2 dt \\ X_0 \sim \sigma_0 := \mathcal{N}(\mu_1 = 0.1, \sigma_1 = 0.05). \end{cases} \tag{36}$$

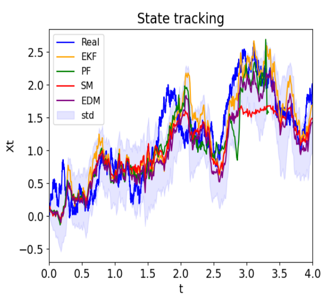
The coefficients are set as $[a_0, a_1, a_2] = [0.3, -0.2, 0.1]$ and $H_0 = 1.0$. The covariance coefficient in the observation process is $s = 0.08$. The total time is $T = 2$. The simulation runs over a time horizon of $T = 4$, and the observation time interval is set to $\Delta t = 0.01$. Other default parameters related to estimating the MMSE and MT are the same as in Example IV.1. For the Legendre SM, the spatial scaling factor is chosen as $C = 3.0$, and the number of Legendre basis polynomials is $m = 12$. For the extended direct method (EDM), we restrict the spatial computational domain to $(x, y) \in [-3.0, 3.0] \times [-3.0, 3.0]$ for the fundamental equation $K(t, x, y)$. The number of particles is set to $N = 50$ when simulating the PF. The potential function $q(x)$ is a polynomial of degree 4, given by the following form:

$$\begin{aligned} q(x) =& (a_0^2 + a_1) + (2a_0 a_1 + 2a_2)x + (a_1^2 + 2a_0 a_2 + \frac{H_0^2}{s})x^2 \\ &+ 2a_1 a_2 x^3 + a_2^2 x^4. \end{aligned} \tag{37}$$

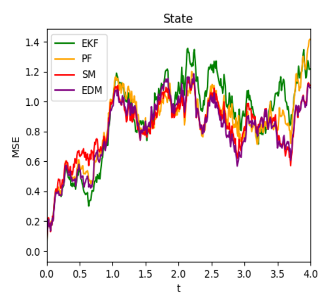
The corresponding forward Kolmogorov equation can also be written as

$$\begin{aligned} \frac{\partial \rho}{\partial t}(t, x) =& \frac{1}{2}\frac{\partial^2 \rho}{\partial x^2} - (a_0 + a_1 x + a_2 x^2)\frac{\partial \rho}{\partial x} \\ &- (a_1 + 2a_2 x + \frac{1}{2}\frac{H_0^2}{s}x^2)\rho. \end{aligned} \tag{38}$$

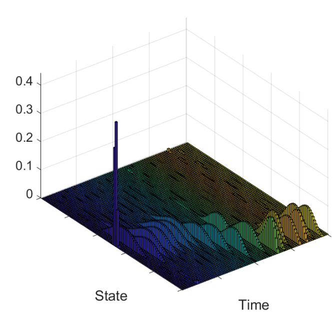
Figure 3 (top left) illustrates the state estimation simulated using four different filtering algorithms. Among these, the simulation performed by EDM is closest to that of SM, compared to EKF and PF. Throughout the simulated period, EKF generally provides reasonable estimates, though occasionally unstable values are observed. In contrast, PF produces less accurate estimates, especially during the latter part of the simulation, where noticeable oscillations occur. As shown in Table V, both SM and EDM outperform the other two methods, with the MMSE of EDM being 5% lower than that of PF and 9.6% lower than that of EKF. In terms of mean computational time (MT), EDM demonstrates significantly higher computational efficiency, with a computational time that is 87% lower than SM and 91% lower than PF. Finally, Figure 3 also depicts the conditional density of the state.



(a)



(b)



(c)

Fig. 3: State track with 1 STD (upper left), mean square error (upper right), and conditional density (lower middle) in Example IV.2.

Algorithms	EKF	PF	SM	EDM
MT	**0.0089**	1.969	1.305	0.1697
MMSE	0.8893	0.8456	0.8238	**0.8036**

TABLE V: Performance of all simulated algorithms in Example IV.2. The bold numbers represent the lowest value in terms of the two measurement metrics, MT and MMSE.

EXAMPLE IV.3 (2D cubic system).

$$\begin{cases} dX_t = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} dt + dW_t, \quad \mathbb{E}[dW_t dW_t^{\top}] = I dt, \\ dZ_t = \begin{bmatrix} X_1^3 \\ X_2^3 \end{bmatrix} dt + dV_t, \quad \mathbb{E}[dV_t dV_t^{\top}] = sI dt, \\ X_0 \sim \mathcal{N}([0,0]^{\top}, 0.05 I_2) \end{cases} \tag{39}$$

The observation covariance is given by $S = 0.1$. The drift coefficients are specified as $a_{11} = -0.3$, $a_{12} = a_{21} = 0.1$, and $a_{22} = 0.2$. The simulation is carried out over a total time horizon of $T = 4$, with an observation time step of $dt = 0.02$. For SM, the number of basis polynomials is set to 45, considering the basis order up to 9. The scaling factor is $C = 3.0$. For the PF, we shall use 100 particles to approximate the empirical density in this 2D filter. For the extended direct method (EDM), the computational domain is set as $(x_1, x_2, y_1, y_2) \in [-3, 3]^4$. The potential function $q(x)$ in this example is a polynomial of degree 6, given by the following form:

$$\begin{aligned} q(x_1, x_2) =& (a_{11} + a_{22}) + (a_{11}x_1 + a_{12}x_2)^2 \\ &+ (a_{12}x_1 + a_{22}x_2)^2 + \frac{H_0^2(x_1^6 + x_2^6)}{s}. \end{aligned} \tag{40}$$

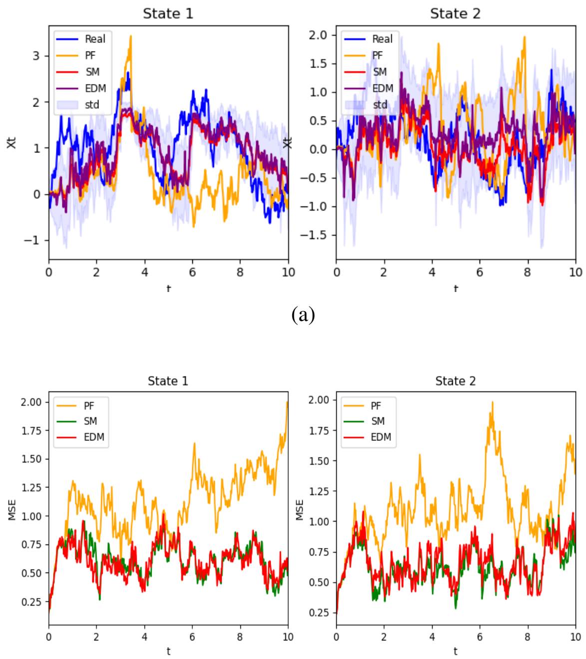
The corresponding forward Kolmogorov equation is written as follows.

$$\begin{aligned} \frac{\partial \rho}{\partial t}(t, x_1, x_2) =& \frac{1}{2}(\rho_{x_1x_1} + \rho_{x_2x_2}) - (a_{11}x_1 + a_{12}x_2)\rho_{x_1} \\ &- (a_{21}x_1 + a_{21}x_2)\rho_{x_2} \\ &- (a_{11} + a_{22})\rho - \frac{1}{2}(x_1^6 + x_2^6)s^{-1}\rho, \end{aligned} \tag{41}$$

The nonlinear property in this example arises from the observation terms with cubic polynomials, making it a typically challenging 2D filter. Simulation results for state tracking and the mean square errors of different filter algorithms are shown in Fig. 4. In the upper two subfigures of Fig. 4, EKF fails to provide effective estimates for the true state on both state dimensions x_1 and x_2. The PF provides a reasonable estimate for state x_2, while the estimation error in state x_1 is relatively large, meaning the PF can only capture partial useful information. Regarding the MMSE shown in Table VI, SM can be considered the optimal estimation method, with an MMSE lower than that of PF by 54%. Our newly proposed algorithm, EDM, not only approaches the optimal estimation provided by SM but also significantly reduces the computational cost compared to SM by 48.8% and compared to PF by 45%. In terms of precision, both SM and EDM can effectively track the true state on both dimensions.

V. Conclusion

This paper aims to introduce a real-time filtering algorithm designed to effectively handle systems characterized by infinite-dimensional statistical structures. More precisely, we have investigated the time-varying



(a)

(b)

Fig. 4: State track with 1 STD (a) and mean square error (b) in Example IV.3. The left and right columns represent corresponding values on the first state variable and second one respectively.

Algorithms	PF	SM	EDM
MT	8.8377	9.4968	**4.853**
MMSE	0.9496	**0.4347**	0.4538

TABLE VI: Performance of all simulated algorithms in Example IV.3. The bold numbers represent the lowest value in terms of the two measurement metrics, MT and MMSE.

Kolmogorov equation with polynomial potential of any degree. Through a series of invertible exponential transformations, explicit fundamental solutions can be written down. Existence and uniqueness can be well guaranteed by the Euler operator theory. Based on the first-order truncation of linear parabolic PDE, the extended direct method was successfully proposed. There are several advantages attained by EDM compared to other filter algorithms through typical numerical simulations. EDM is more flexible than the direct method for filters with infinite-dimensional statistics. It is easy to implement by separating the online update from the offline computation. EDM has better precision than EKF and PF, which achieves a comparative accuracy level as the Legendre-Galerkin spectral method but with much fewer computational efforts.

By utilizing a derived analytical kernel solution, in the moderate dimension case, state estimation can be obtained by applying a traditional equidistant grid. However, for high-dimensional numerical integrals, a more efficient sampling method is needed, such as sparse quadrature or targeted Monte-Carlo sampling. However, there are still several bottlenecks in using Monte-Carlo sampling, such as slow convergence, high variance, and sampling inefficiency. One thing that should be noted here is that our method is applicable to high dimensions. Kernel value on arbitrary sparse points can be immediately obtained by our efficient approximation scheme. This would motivate us to propose a sparse sampling-based kernel filter, which will be our future work.

REFERENCES

[1] Ji Shi, Zhiyu Yang, and Stephen S-T Yau. Direct method for Yau filtering system with nonlinear observations. *International Journal of Control*, 91(3):678–687, 2018.

[2] Xiuqiong Chen, Xue Luo, and Stephen S-T Yau. Direct method for time-varying nonlinear filtering problems. *IEEE Transactions on Aerospace and Electronic Systems*, 53(2):630–639, 2017.

[3] R. E. Kalman. A new approach to linear filtering and prediction problem. *Journal of Basic Engineering*, 82:35–45, 1960.

[4] Rudolph Emil Kalman and Richard S Bucy. New results in linear filtering and prediction theory. *Journal of Basic Engineering*, 83(1):95–108, 1961.

[5] Brian D. O. Anderson and John B. Moore. *Optimal filtering*. Courier Corporation, 2012.

[6] Simon Julier, Jeffrey Uhlmann, and Hugh F Durrant-Whyte. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Transactions on automatic control*, 45(3):477–482, 2000.

[7] Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162, 1994.

[8] Ienkaran Arasaratnam and Simon Haykin. Cubature kalman filters. *IEEE Transactions on automatic control*, 54(6):1254–1269, 2009.

[9] Silvere Bonnable, Philippe Martin, and Erwan Salaün. Invariant extended kalman filter: theory and application to a velocity-aided attitude estimation problem. In *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, pages 1297–1304. IEEE, 2009.

[10] Kazufumi Ito and Kaiqi Xiong. Gaussian filters for nonlinear filtering problems. *IEEE transactions on automatic control*, 45(5):910–927, 2000.

[11] Tao Yang, Prashant G Mehta, and Sean P Meyn. Feedback particle filter. *IEEE transactions on Automatic control*, 58(10):2465–2480, 2013.

[12] Tyrone Edward Duncan. *Probability densities for diffusion processes with applications to nonlinear filtering theory and detection theory*. Stanford Uni-

versity, 1967.
[13] R. E. Mortensen. *Optimal control of continuous time stochastic systems*. PhD thesis, University of California, Berkley, California, Aug. 1966.
[14] M. Zakai. On the optimal filtering of diffusion process. *Z. Wahrsch. verw. Gebiete*, 11(3):230–243, 1969.
[15] Moshe Zakai. On the optimal filtering of diffusion processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 11(3):230–243, 1969.
[16] S T Yau and SS T Yau. Explicit solution of a kolmogorov equation. *Applied Mathematics and Optimization*, 34:231–266, 1996.
[17] Shing-Tung Yau and Stephen S-T Yau. Real time solution of nonlinear filtering problem without memory I. *Mathematical Research Letters*, 7(6):671–693, 2000.
[18] S.-T. Yau and S. S.-T. Yau. Real time solution of the nonlinear filtering problem without memory II. *SIAM Journal on Control and Optimization*, 47(1):163–195, 2008.
[19] Xue Luo and Stephen S-T Yau. Complete real time solution of the general nonlinear filtering problem without memory. *IEEE Transactions on Automatic Control*, 58(10):2563–2578, 2013.
[20] Wenhui Dong, Xue Luo, and Stephen S-T Yau. Solving nonlinear filtering problems in real time by legendre galerkin spectral method. *IEEE Transactions on Automatic Control*, 66(4):1559–1572, 2020.
[21] Ji Shi, Xiuqiong Chen, and Stephen Shing-Toung Yau. A novel real-time filtering method to general nonlinear filtering problem without memory. *IEEE Access*, 9:119343–119352, 2021.
[22] Ji Shi, Xiaopei Jiao, and Stephen S-T Yau. Dglg: A novel deep generalized legendre-galerkin approach to optimal filtering problem. *IEEE Transactions on Automatic Control*, 2024.
[23] Bhashyam Balaji. Euclidean quantum mechanics and universal nonlinear filtering. *Entropy*, 11(01):42–58, 2009.
[24] Bhashyam Balaji. Feynman path integral discretization and its applications to nonlinear filtering. In *Signal Processing, Sensor Fusion, and Target Recognition XXII*, volume 8745, pages 206–217. SPIE, 2013.
[25] Guo-Qing Hu and SS-T Yau. Finite-dimensional filters with nonlinear drift. xv. new direct method for construction of universal finite-dimensional filter. *IEEE Transactions on Aerospace and Electronic Systems*, 38(1):50–57, 2002.
[26] S. T. Yau and S. S.-T. Yau. Nonlinear filtering and time varying Schrödinger equation I. *IEEE Transactions on Aerospace & Electronic Systems*, 40:284–292, 2004.
[27] Xiao-pei Jiao and Stephen S-T Yau. New classes of finite dimensional filters with nonmaximal rank estimation algebra on state dimension n and linear rank n-2. *SIAM Journal on Control and Optimization*, 58(6):3413–3427, 2020.
[28] Xiaopei Jiao and Stephen S-T Yau. Finite-dimensional estimation algebra on arbitrary state dimension with nonmaximal rank: linear structure of wong matrix. *International Journal of Control*, 97(11):2669–2676, 2024.
[29] Hongyu Yu, Xiaopei Jiao, and Stephen S-T Yau. Complete classification of finite dimensional estimation algebras with state dimension n, linear rank n-1, and constant wong matrix. *IEEE Transactions on Automatic Control*, 69(1):295–302, 2023.
[30] Stephen S-T Yau, Xiuqiong Chen, Xiaopei Jiao, Jiayi Kang, Zeju Sun, and Yangtianze Tao. *Principles of Nonlinear Filtering Theory*, volume 33. Springer Nature, 2024.
[31] Jie Shen, Tao Tang, and Li-Lian Wang. *Spectral methods: algorithms, analysis and applications*, volume 41. Springer Science & Business Media, 2011.
[32] C González, A Ostermann, and Mechthild Thalhammer. A second-order magnus-type integrator for nonautonomous parabolic problems. *Journal of computational and applied mathematics*, 189(1-2):142–156, 2006.
[33] Mechthild Thalhammer. A fourth-order commutator-free exponential integrator for nonautonomous differential equations. *SIAM journal on numerical analysis*, 44(2):851–864, 2006.

APPENDIX

Proof of Proposition III.1. Direct computation yields that

$$\begin{aligned}\frac{\partial K}{\partial t}(t,x,y) &= \left(-\frac{n}{2t}+\frac{\partial W}{\partial t}\right)K(t,x,y),\\ \nabla_x K(t,x,y) &= K(t,x,y)\nabla_x W(t,x,y),\\ \Delta_x K(t,x,y) &= K(t,x,y)(|\nabla_x W(t,x,y)|^2+\Delta_x W(t,x,y)).\end{aligned} \tag{42}$$

□

Proof of Proposition III.2. Direct computation yields that

$$\begin{aligned}\frac{\partial W}{\partial t}(t,x,y) &= \frac{\|x-y\|^2}{2t^2}+\frac{\partial \tilde{W}}{\partial t}(t,x,y),\\ \nabla_x W(t,x,y) &= \frac{y-x}{t}+\nabla_x\tilde{W}(t,x,y),\\ \Delta_x W(t,x,y) &= -\frac{n}{t}+\Delta_x\tilde{W}(t,x,y).\end{aligned} \tag{43}$$

□

Proof of Proposition III.3. Direct computation implies that

$$\begin{aligned}\frac{\partial \tilde{W}}{\partial t}(t,x,y) &= \frac{1}{\phi}\frac{\partial\phi}{\partial t}(t,x,y),\\ \nabla_x\tilde{W}(t,x,y) &= \frac{1}{\phi}\nabla_x\phi(t,x,y),\\ \Delta_x\tilde{W} &= -\frac{1}{\phi^2}|\nabla_x\phi(t,x,y)|^2+\frac{1}{\phi}\Delta_x\phi(t,x,y).\end{aligned} \tag{44}$$

□

Proof of Theorem III.1. First we transform iterative equations (20) to the following equations by using change of variable $z = x - y$,

$$\begin{aligned} z \cdot \nabla_z \phi_1 + \phi_1 &= -\frac{1}{2} q_0(z+y), \\ z \cdot \nabla_z \phi_{k+1} + (k+1)\phi_{k+1} &= \frac{1}{2}\Delta_z \phi_k \\ &- \frac{1}{2}\sum_{i=0}^{k} q_i(z+y)\phi_{k-i}, k \geq 1, \end{aligned} \tag{45}$$

where relations $\nabla_x = \nabla_z$ and $\Delta_x = \Delta_z$ are used.

Next, we start with the first equation to solve ϕ_1. The same techniques from Theorem III.3 shows

$$\phi_1(z) = \sum_{0 \leq |\alpha| \leq m} \frac{q_{0\alpha}(y)}{|\alpha|+1} z^\alpha. \tag{46}$$

We assume the proposed solution form based on the theorem.III.4,

$$\phi_k = \sum_{0 \leq |\alpha| \leq km} \phi_{k\alpha}(y) z^\alpha, \tag{47}$$

for $k \geq 2$. By induction, if we have obtained a solution of ϕ_i for $i \leq k$. Direct computation will show that

$$\begin{aligned} &\frac{1}{2}\Delta_z \phi_k - \frac{1}{2}\sum_{i=0}^{k} q_i(z+y)\phi_{k-i} \\ &= \sum_{|\gamma|=0}^{km-2}\left[\sum_{i=1}^{n}(\gamma_i+2)(\gamma_i+1)\phi_{k(\gamma_1,\cdots,\gamma_i+2,\cdots,\gamma_n)}\right. \\ &\quad \left. + \sum_{|\beta|=|\gamma|-m}^{|\gamma|}\sum_{i=0}^{k} q_{i(\gamma-\beta)}\phi_{(k-i)\beta}\right] z^\gamma \\ &\quad + \sum_{|\gamma|=km-1}^{(k+1)m}\left[\sum_{|\beta|=|\gamma|-m}^{|\gamma|}\sum_{i=0}^{k} q_{i(\gamma-\beta)}\phi_{(k-i)\beta}\right] z^\gamma. \end{aligned} \tag{48}$$

and

$$z \cdot \nabla_z \phi_{k+1} + (k+1)\phi_{k+1} = \sum_{|\gamma|=0}^{(k+1)m} (|\gamma|+k+1)\phi_{(k+1)\gamma}(y) z^\gamma. \tag{49}$$

Substitute the form (47) into equation (45), and we compare the coefficients on both sides of (45). Finally we get the components in series solution:

$$\phi_{(k+1)\alpha}(y) = \begin{cases} \frac{1}{|\gamma|+k+1}\left[\sum_{i=1}^{n}(\gamma_i+2)(\gamma_i+1)\phi_{k(\gamma_1,\cdots,\gamma_i+2,\cdots,\gamma_n)}\right. \\ \left. + \sum_{|\beta|=|\gamma|-m}^{|\gamma|}\sum_{i=0}^{k} q_{i(\gamma-\beta)}\phi_{(k-i)\beta}\right], \\ \text{for} \quad 0 \leq |\gamma| \leq km-2, \\ \frac{1}{|\gamma|+k+1}\left[\sum_{|\beta|=|\gamma|-m}^{|\gamma|}\sum_{i=0}^{k} q_{i(\gamma-\beta)}\phi_{(k-i)\beta}\right], \\ \text{for} \quad km-1 \leq |\gamma| \leq (k+1)m. \end{cases} \tag{50}$$

□

Proof of Lemma III.2. The strategy used in this proof is basically the same as those provided in [30]. For the convenience for readers to follow, we summarize and reduce the original proof in the following. For simplificty of expression, we let $S = \{1, 2, \cdots, l\} \subset \{1, 2, \cdots, n\}$. Our proof includes two parts.

Step 1. We shall prove that the function ζ is a polynomial of degree k in the variables $x_1, \ldots, x_l$, with coefficients that are smooth functions of $x_{l+1}, \ldots, x_n$.

Next, we introduce the multi-index $\beta = (\beta_1, \ldots, \beta_l)$ with $|\beta| = k+1$, and define the differential operator $D := \left(\frac{\partial}{\partial x_1}\right)^{\beta_1} \cdots \left(\frac{\partial}{\partial x_l}\right)^{\beta_l}$. It then suffices to demonstrate that $D(\zeta) = 0$.

It is clear that $D[E_S(\zeta) + m\zeta] = 0$. To proceed, we aim to simplify the term $DE_S(\zeta)$ by interchanging the order of the operators D and E_S. As a first step, we establish the following identities through straightforward computation:

$$(\frac{\partial}{\partial x_j})^p E_S = E_S(\frac{\partial}{\partial x_j})^p + p(\frac{\partial}{\partial x_j})^p, \text{for } j \in S \text{ and } p \in \mathbb{Z}_+. \tag{51}$$

By applying the above identities and using the relation $D[E_S(\zeta) + m\zeta] = 0$, we obtain:

$$\begin{aligned} 0 =& D[E_S(\zeta) + m\zeta] \\ =& \left(\frac{\partial}{\partial x_{s_1}}\right)^{\beta_1} \cdots \left(\frac{\partial}{\partial x_{s_l}}\right)^{\beta_l} E_S(\zeta) + mD\zeta \\ =& E_S(D\zeta) + (|\beta| + m)D\zeta. \end{aligned} \tag{52}$$

Next, we observe the following identities involving differential operators.

$$\begin{aligned} E_S(x_j)^p =& x_j^p E_S + p x_j^p, \quad \text{for } j \in S \text{ and } p \in \mathbb{Z}_+ \\ =& x_j^p (E_S + p). \end{aligned} \tag{53}$$

Applying Eq. (53), we obtain

$$0 = x_1^{k+m+1}[E_S(D\zeta) + (k+m+1)D\zeta] = E_S(x_1^{k+m+1} D\zeta). \tag{54}$$

Next, we define $\phi(x) := x_1^{k+m+1} D\zeta$, with the goal of showing that $\phi \equiv 0$. This directly implies that $D\zeta = 0$. Note that

$$\begin{aligned} 0 =& \int_\varepsilon^1 \frac{1}{t} E_S(\phi(tx_1, \cdots, tx_l, x_{l+1}, \cdots, x_n))dt \\ =& \int_\varepsilon^1 \frac{d\phi}{dt}(tx_1, \cdots, tx_l, x_{l+1}, \cdots, x_n)dt \\ =& \phi(x) - \phi(\varepsilon x_1, \cdots, \varepsilon x_l, x_{l+1}, \cdots, x_n), \end{aligned} \tag{55}$$

which implies that $\phi(x) = \phi(0, \ldots, 0, x_{l+1}, \ldots, x_n) = 0$ as $\varepsilon \to 0$.

Step 2. We shall prove the function ζ is a polynomial whose degree is no greater than k.

Next we assume the representation

$$\zeta = \sum_{0 \leq |\alpha| \leq k} a_\alpha(x_{l+1}, \cdots, x_n) x_1^{\alpha_1} \cdots x_l^{\alpha_l}, \tag{56}$$

where

$$a_\alpha(x_{l+1}, \cdots, x_n) \in C^\infty(\mathbb{R}^{n-l}), \tag{57}$$

and

$$E_S(\zeta) + m\zeta = \sum_{0 \leq |\alpha| \leq k} b_\alpha(x_{l+1}, \cdots, x_n) x_1^{\alpha_1} \cdots x_l^{\alpha_l}, \tag{58}$$

where $b_\alpha(x_{l+1}, \ldots, x_n)$ denotes a polynomial of degree no more than $k - |\alpha|$. From Eq. (56) to Eq. (58), it follows that $a_\alpha(x_{l+1}, \ldots, x_n)$ is also a polynomial of degree no greater than $k-|\alpha|$, as established by comparing coefficients on both sides. $\square$

Proof of Theorem III.3. To find an explicit solution to the equation, we make a change of variable $z = x - y$. Noticing that $\nabla_x = \nabla_z$, it follows that

$$z \cdot \nabla_z \zeta + s\zeta = f(z+y, y), \tag{59}$$

Notice change of variable does not influence the degree of $f(z+y, y)$. Hence, we shall assume

$$f(z+y, y) = \sum_{0 \le |\alpha| \le m} f_\alpha(y) z^\alpha, \tag{60}$$

where $f_\alpha(y)$ is a polynomial of y with degree no more than $m - |\alpha|$.

Lemma III.2 can be applied that the solution ζ must be a polynomial of degree m for argument (y, z). Then we assume

$$\zeta = \sum_{0 \le |\alpha| \le m} b_\alpha(y) z^\alpha, \tag{61}$$

where $b_\alpha(y)$ is a polynomial of y with degree no more than $m - |\alpha|$. Substitute such a polynomial form to the Euler operator type PDE (59)

$$\sum_{0 \le |\alpha| \le m} (|\alpha| + s) b_\alpha(y) z^\alpha = \sum_{0 \le |\alpha| \le m} f_\alpha(y) z^\alpha, \tag{62}$$

and compare the coefficients on both sides, which implies

$$b_\alpha(y) = \frac{f_\alpha(y)}{|\alpha| + s}, \quad \forall 0 \le |\alpha| \le m. \tag{63}$$

Therefore,

$$\zeta = \sum_{0 \le |\alpha| \le m} \frac{f_\alpha(y)}{|\alpha| + s} (x - y)^\alpha. \tag{64}$$

$\square$

Proof of Theorem III.4. The induction method will be used in this proof. First, when $k = 1$, We observe that q_0 is a polynomial whose degree is less than or equal to m. Theorem III.3 implies that ϕ_1 has a unique polynomial solution of degree at most m. From induction, we shall assume $\deg(\phi_n) = nm$ holds for $n \le k$. Simple calculations demonstrate that $\deg(\frac{1}{2}\Delta_x \phi_k - \frac{1}{2}\sum_{i=0}^{k} q_i(x)\phi_{k-i}) = (k+1)m$, so $\deg(\phi_{k+1}) = (k+1)m$ again by Theorem III.3. The desired results are obtained. $\square$

Xiaopei Jiao (Member, IEEE) received B.S. degree in applied physics and a dual B.S. degree in computer science from Shanghai Jiao Tong University, China in 2017 and a Ph.D. degree in applied mathematics from Tsinghua University, China in 2022. After his Ph.D., he joined as a postdoctoral researcher at the Beijing Institute of Mathematical Sciences and Applications (BIMSA) and subsequently at the University of Twente, Netherlands. Now, he is an assistant professor at BIMSA, China. His research interests include nonlinear filter theory and numerical applications, Physics-Informed neural networks (PINNs), and numerical methods in partial differential equations.

Ji Shi (Member, IEEE) received the B.S. degree in Information and computing science in North China Electric Power University, Beijing, China, in 2013, and the Ph.D. degree in applied mathematics from the Department of Mathematical Sciences, Tsinghua University, Beijing, China. He is currently an associate professor in the Academy for Multidisciplinary Studies at Capital Normal University, Beijing, China. His research interests include nonlinear filtering, active disturbance rejection control, and deep learning.

Stephen S.-T. Yau (Life Fellow, IEEE) received a Ph.D. degree in mathematics from the State University of New York at Stony Brook, NY, USA in 1976. He was a Member of the Institute of Advanced Study at Princeton from 1976-1977 and 1981-1982, and a Benjamin Pierce Assistant Professor at Harvard University during 1977-1980. After that, he joined the Department of Mathematics, Statistics, and Computer Science (MSCS), University of Illinois at Chicago (UIC), and served for over 30 years. During 2005-2011, he became a joint Professor with the Department of Electrical and Computer Engineering at the MSCS, UIC. After his retirement in 2012, he joined Tsinghua University, Beijing, China, where he is a full-time professor in the Department of Mathematical Sciences. His research interests include nonlinear filtering, bioinformatics, complex algebraic geometry, CR geometry, and singularity theory.

Dr. Yau is the Managing Editor and founder of the *Journal of Algebraic Geometry* since 1991, and the Editor-in-Chief and founder of *Communications in Information and Systems* from 2000 to the present. He was the General Chairman of the IEEE International Conference on Control and Information, which was held at the Chinese University of Hong Kong in 1995. He was awarded the Sloan Fellowship in 1980, the Guggenheim Fellowship in 2000, and the AMS Fellow Award in 2013. In 2005, he was entitled the UIC Distinguished Professor.