Proper Orthogonal Decomposition Method to Nonlinear Filtering Problems in Medium-High Dimension

Zhongjian Wang, Xue Luo, *Senior Member, IEEE*, Stephen S.-T. Yau, *Fellow, IEEE*, and Zhiwen Zhang

***Abstract*—In this paper, we investigate the proper orthogonal decomposition (POD) method to numerically solve the forward Kolmogorov equation (FKE). Our method aims to explore the low-dimensional structures in the solution space of the FKE and to develop efficient numerical methods. As an important application and our primary motivation to study the POD method to FKE, we solve the nonlinear filtering (NLF) problems with a real-time algorithm proposed by Yau and Yau combined with the POD method. This algorithm is referred as POD algorithm in this paper. Our POD algorithm consists of offline and online stages. In the offline stage, we construct a small number of POD basis functions that capture the dynamics of the system and compute propagation of the POD basis functions under the FKE operator. In the online stage, we synchronize the coming observations in a real-time manner. Its convergence analysis has also been discussed. Some numerical experiments of the NLF problems are performed to illustrate the feasibility of our algorithm and to verify the convergence rate. Our numerical results show that the POD algorithm provides considerable computational savings over existing numerical methods.**

***Index Terms*—Duncan–Mortensen–Zakai equation, nonlinear filtering (NLF) problems, proper orthogonal decomposition (POD), real-time algorithm.**

Manuscript received October 31, 2018; revised April 4, 2019; accepted June 22, 2019. Date of publication July 9, 2019; date of current version March 27, 2020. The work of Z. Wang is supported by the Hong Kong Ph.D. Fellowship Scheme. The work of X. Luo is supported in part by the National Natural Science Foundation of China (NSFC) under Grant 11871003 and in part by the Fundamental Research Funds for the Central Universities under Grant YWF-19-BJ-J-269. The work of S. S.-T. Yau is supported by Tsinghua University Education Foundation fund under Grant 042202008, NSFC under Grant 11471184, and in part by the start-up fund from Tsinghua University. The work of Z. Zhang is supported in part by the Hong Kong RGC under Grant 27300616, Grant 17300817, and Grant 17300318), in part by the NSFC under Grant 11601457, in part by the Seed Funding Programme for Basic Research (HKU), and in part by the RAE Improvement Fund (HKU). Recommended by Associate Editor Z. Gao. *(Xue Luo is co-first author.) (Corresponding authors: Stephen S.-T. Yau; Zhiwen Zhang.)*

Z. Wang and Z. Zhang are with the Department of Mathematics, The University of Hong Kong, Hong Kong (e-mail: aris1992@outlook.com; zhangzw@hku.hk).

X. Luo is with the School of Mathematical Sciences, Beihang University (Shahe campus), Changping District, Beijing 102206, China (e-mail: xluo@buaa.edu.cn).

S. S.-T. Yau is with the Department of Mathematical Sciences, Tsinghua University, Beijing 100084, China (e-mail: yau@uic.edu).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2019.2927322

I. Introduction

NONLINEAR filtering (NLF) problem is originated from the problem of tracking and signal processing. The fundamental problem in the NLF is to give an instantaneous and accurate estimation of the states based on the noisy observations [20], [38]. Further investigation on the filtering of backward and forward–backward stochastic differential equations can refer to [36]. In this paper, we proposed an efficient numerical method to solve the forward Kolmogorov equation (FKE) arising from the NLF problem [19]. Our method is based on the proper orthogonal decomposition (POD) method [7], [35], [37], which is an effective tool in exploring the intrinsic low-dimensional structures of high-dimensional solutions. We start from the following signal based model:

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

where $x_t \in \mathbb{R}^n$ is the state of the system at time t, the initial state x_0 satisfying some initial distribution, $y_t \in \mathbb{R}^m$ is the observation at time t with $y_0 = 0$, and v_t and w_t are vector-valued Brownian motion processes with covariance matrices $E[dv_t dv_t^T] = Q(t)dt \in \mathbb{R}^{n\times n}$ and $E[dw_t dw_t^T] = S(t)dt \in \mathbb{R}^{m\times m}$, $S(t) > 0$, respectively. Furthermore, we assume that x_0, dw_t, and dv_t are independent. The most popular method so far to solve (1) is the particle filter (PF), see [2], [3], [13] and references therein. However, the main drawback of the PF is that it is hard to be implemented as a real-time solver due to its nature of the Monte Carlo simulation.

In 1960s, Duncan [11], Mortensen [29], and Zakai [41] independently derived the so-called Duncan–Mortensen–Zakai (DMZ) equation (or Zakai equation), which asserts that the unnormalized conditional density function of the states x_t, denoted by $\sigma(x,t)$, satisfies the following Ito stochastic partial differential equation (SPDE)

$$\begin{cases} d\sigma(x,t) = \mathcal{L}\sigma(x,t)dt + \sigma(x,t)h^T(x,t)S^{-1}dy_t \\ \sigma(x,0) = \sigma_0(x) \end{cases} \tag{2}$$

where $\sigma_0(x)$ is the density of the initial states x_0, and

$$\mathcal{L}(\cdot) := \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j}\left((gQg^T)_{ij}\,\cdot\right) - \sum_{i=1}^{n} \frac{\partial(f_i\cdot)}{\partial x_i}. \tag{3}$$

0018-9286 © 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Tsinghua University. Downloaded on April 16,2020 at 04:05:48 UTC from IEEE Xplore. Restrictions apply.

The DMZ equation laid down the solid foundation to study the NLF problem from the viewpoint of SPDE. However, the DMZ equation cannot be solved analytically in general. Many efforts have been made to develop efficient numerical methods. One of the commonly used method is the splitting-up method originated from the Trotter product formula, which was first introduced in [6] and has been extensively studied later, see [14], [18], and [30]. In [24], the so-called S^3 algorithm was developed based on the Wiener chaos expansion, by separating the computations involving the observations from those dealing only with the system parameters. The boundedness of the drift term f, diffusion term g, and observation term h is required for the technical proof.

To overcome this restriction, the third author and his coworker [39] developed a novel algorithm to solve the DMZ equation. Specifically, for each given realization of the observation process denoted by y_t, they made an invertible exponential transformation

$$\sigma(x,t) = \exp\big(h^T(x,t)S^{-1}(t)y_t\big)u(x,t) \tag{4}$$

and transformed the DMZ (2) into a deterministic partial differential equation with stochastic coefficients

$$\begin{cases} \frac{\partial}{\partial t}u(x,t) + \frac{\partial}{\partial t}(h^T S^{-1})y_t u(x,t) \\ = \exp\big(-h^T(x,t)S^{-1}(t)y_t\big)\big(\mathcal{L} - \frac{1}{2}h^T S^{-1}h\big) \\ \quad \cdot \big(\exp\big(-h^T(x,t)S^{-1}(t)y_t\big)u(x,t)\big) \\ \quad u(x,0) = \sigma_0(x) \end{cases}. \tag{5}$$

Equation (5) is the so-called pathwise robust DMZ equation. The boundedness of the drift term f (contained in the operator $\mathcal{L}(\cdot)$) and observation term h is replaced by some mild growth conditions in this case. Nevertheless, they still make the assumption that the drift term, the observation term, and the diffusion term are time invariant, which means that f, h, and g in (1) cannot explicitly depend on time. Later on, in [26], the second and the third author of this paper generalized the algorithm in [39] to more general settings of the NLF problems, namely, the time-dependent ones. In this paper, we refer it as offline and online method.

Let us assume that the observation time sequences $0 = t_0 < t_1 < \cdots < t_{N_t} = T$ are given. Notice that the observation data $\{y_{t_j}\}$ at each observation time t_j, $j = 1, \ldots, N_t$ are unknown until the online experiment runs. Therefore, in each time interval $t_{j-1} \le t < t_j$, one freezes the stochastic coefficient y_t to be $y_{t_{j-1}}$ in (5) and makes the exponential transformation

$$u_j(x,t) = \exp\big(h^T(x,t)S^{-1}(t)y_{t_{j-1}}\big)u(x,t).$$

It is easy to deduce that u_j satisfies the FKE

$$\frac{\partial}{\partial t}u_j(x,t) = \left(\mathcal{L} - \frac{1}{2}h^T S^{-1}h\right)u_j(x,t) \tag{6}$$

where the operator $\mathcal{L}$ is defined in (3). In [27], the offline and online method has been put into practice by the second and the third author of this paper. They investigated the Hermite spectral method to numerically solve the one-dimensional (1-D) FKE (6) and analyzed the convergence rate of the proposed method.

Although for the extremely low-dimensional state, the offline and online method [27] is very efficient as to accuracy and CPU time compared with PF, the bottle-neck of the algorithm in [26] is to solve high-dimensional FKE accurately and compute a huge amount of numerical integrations online, if the state in NLF problems is high dimensional. The real-time manner can be heavily deteriorated by the so-called curse of dimensionality. Yueh *et al.* [40] proposed a numerical scheme based on the quasi-implicit Euler method for solving the high-dimensional FKE, which took more than 131 min to solve a 6-D problem in time interval [0, 20] with observation time step $\Delta\tau = 0.01$ on a desktop computer, which is far from being real time.

This motivates us to investigate the possible low-dimensional structures in the high-dimensional FKEs arising from NLF problems, so that we can design more efficient numerical methods. In fact, many high-dimensional problems have certain low-dimensional structures, which suggest the existence of reduced-order models (ROMs) and better formulations for efficient numerical methods. Inspired by the recent work of the last author in this paper who has developed problem-dependent basis functions to solve SPDEs [8]–[10], we propose to use the POD method to explore the low-dimensional structures of the solutions to FKE. This in turn will help us obtain an efficient numerical method to solve the NLF problems.

The POD algorithm in this paper constructs the basis from the snapshots of some reference solutions, obtained either by finite difference method (FDM) [40] or by spectral method [27]. The basis is called POD basis in the sequel, which represents the most energetic structures of the FKE and provides an efficient way to explore the low-dimensional structures of the FKE solutions. With the POD basis at hand, the offline and online method can be applied by replacing the prescribed basis, say generalized Hermite functions (GHF) in [27].

The advantage of the POD algorithm is that with much fewer POD basis, most dynamics of the system can be captured well. Therefore, the POD algorithm significantly reduces the degree of freedom (DOF) in the online computation. For example, in Section IV, the number of the POD basis for the 2-D almost linear problem and 2-D cubic sensor problem are only 70 and 100, respectively. If simply tensor-producting the GHF in [27] for 2-D problems, it would be 625 basis for almost linear problem and 2025 basis for cubic sensor problems, respectively. Hence, by exploring low-dimensional structures in the solution space, the POD algorithm helps us alleviate the curse of dimensionality to a certain extent.

We should point out that the number of the POD basis depends on the decay speed of the eigenvalues of the correlation matrix (14) and is problem-dependent. Due to its energy-minimizing property in the sense that the POD basis minimizes the total mean squared error and gives the optimal representation of solution snapshots, our POD algorithm provides considerable computational savings over existing numerical methods. In other words, the POD algorithm can be viewed as the optimization of the offline and online method by elaborately constructing the problem-dependent basis. After the POD basis functions have been constructed, we only need to solve a much smaller-scaled FKE in the offline stage and much fewer numerical integrations in the online stage. We shall demonstrate the performance of our algorithm through numerical experiments in Section IV.

Authorized licensed use limited to: Tsinghua University. Downloaded on April 16,2020 at 04:05:48 UTC from IEEE Xplore. Restrictions apply.

The rest of this paper is organized as follows. In Section II, we recall the basic idea of the POD method and the established facts of the well-posedness of the pathwise robust DMZ equation. In Section III, we describe the POD algorithm in detail. The convergence and effectiveness analysis of the proposed algorithm will also be discussed. In Section IV, we present numerical results to demonstrate the accuracy and efficiency of our method. Conclusions are drawn in Section V.

II. Preliminaries

In this section, we shall first introduce the POD method for solving a deterministic and nonparametric dynamical system. To make this paper self-contained, we also introduce the existence and uniqueness of the solution to DMZ equation and the offline and online algorithm for time-varying NLF problems.

A. Proper Orthogonal Decomposition

The POD, also known as Karhunen–Loève expansion in stochastic process and signal analysis [21], [23], or the principal component analysis in statistics [1], or singular value decomposition in linear algebra, or the method of empirical orthogonal functions in geophysical fluid dynamics [15], [31], has first been introduced in solving the turbulence in fluid dynamics. It aims to generate optimally ordered orthonormal basis functions in the least squares sense for a given set of theoretical, experimental, or computational data. ROMs or surrogate models are then obtained by truncating this optimal basis functions, which provide considerable computational savings over the original high-dimensional problems. We refer the interested readers to [5], [7], [17], [35], [37], [42], and references therein for more details.

Let X be a Hilbert space equipped with the inner product $(\cdot,\cdot)_X$ and norm $||\cdot||_X$. Let $u(\cdot,t)\in X, t\in[0,T]$ the solution of a dynamic system. In practice, we approximate the space X with a linear finite dimensional space V with $\dim V = d$, where d represents the DOF of the solution space. We should point out that d can be extremely large for high-dimensional problem. Given a set of snapshot of solutions, a linear spaces V can be spanned

$$V = \text{span}\{u(\cdot,t_0), u(\cdot,t_1), \ldots, u(\cdot,t_{N_t})\} \tag{7}$$

where $t_0,\ldots,t_{N_t}\in[0,T]$ are different time instances. The POD method aims to build a low-dimensional orthonormal basis functions $\{\phi_k\}_{k=1}^{N_{\text{pod}}}$ with $N_{\text{pod}} \ll \min(d,(N_t+1))$ that optimally approximates the solution snapshots. Specifically, the POD basis functions $\{\phi_k\}_{k=1}^{N_{\text{pod}}}$ minimize the following error:

$$\frac{1}{N_t+1}\sum_{i=0}^{N_t}\left|\left|u(\cdot,t_i)-\sum_{k=1}^{N_{\text{pod}}}(u(\cdot,t_i),\phi_k(\cdot))_X\phi_k(\cdot)\right|\right|_X^2 \tag{8}$$

subject to the constrains that $(\phi_m(\cdot),\phi_n(\cdot))_X=\delta_{mn},\ 1\le m,n\le N_{\text{pod}}$.

In this paper, we shall use the method of snapshots [35] to construct POD basis from the training solution snapshots and generate a low-dimensional subspace to approximate the solutions of FKE in our NLF problems. More details will be provided in the Section III-A.

B. Pathwise Robust DMZ Equation

As we briefly mentioned in the introduction that the solution of the DMZ (2) is the key to solve the NLF problems completely. However, it is impractical to be solved in an efficient way. With a given observation path, one can derive the pathwise robust DMZ (5) easily with an exponential transform (4). The existence and uniqueness of (5) has been investigated by many researchers. The well-posedness is guaranteed when the drift term $f\in C^1$ and the observation term $h\in C^2$ are bounded in [32]. Later on, similar results were obtained under weaker conditions. For instance, the well-posedness results on the pathwise robust DMZ equation with a special class of unbounded coefficients were obtained in [4] and [12], but the results were for 1-D case. Moreover their results cannot even cover the linear case. In [39], the third author of this paper and his collaborator established the well-posedness result under the condition that f and g have at most linear growth. The second and third author of this paper used more delicate analysis to give a time-varying analogous well-posedness result to the pathwise-robust DMZ equation under some mild growth conditions on f and h in [26].

Although compared to the DMZ (2), the pathwise robust DMZ (5) should be easier to solve, since the stochastic term has been transformed into the coefficients, it is still difficult to obtain an analytic solution in general. Many efforts have been devoted to develop efficient and robust numerical methods to solve the FKE (6), see [6], [14], [18], [24], [30], and references therein.

C. Offline and Online Algorithm

In 2013, the second and third author of this paper developed the offline and online algorithm for the time-varying NLF problems, where the GHF are used as the prescribed orthogonal basis in solving the cubic sensor problem [27]. The offline computing means that it can be performed without any online observation or experimental data. On the contrary, the online computing needs the observation data that is only available during the experiment. We briefly recall this algorithm and summarize it in Algorithm 1. Let $\{\psi_k\}_{k=1}^{N_b}$ denote the prescribed orthogonal basis, say the generalized Hermite functions in [27], and $I^{[t_{j-1},t_j]}$, the propagator defined by solving (9) on a domain D, where the initial data are chosen as each basis ψ_k.

Remark 2.1: If (9) is time-invariant and the observation intervals are uniform, i.e., $t_{j+1}=t_j+\Delta t$, $\forall j$, we only need to calculate the propagator (9) once in the ofline stage. That is, the first for-loop in Algorithm 1 is unnecessary.

III. Our POD Algorithm to Solve the NLF Problems

The novelty of our algorithm is to construct a set of problem-dependent orthogonal basis in the offline stage by the POD algorithm, where the solution snapshots are chosen from reference solutions. The choice of a numerical method for the reference solution is not crucial since all the computation are implemented

Authorized licensed use limited to: Tsinghua University. Downloaded on April 16,2020 at 04:05:48 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1: Off- and On-Line Computing [27].

```
for $j = 1 \to N_t$ do        %Off-line stage
  for $k = 1 \to N_b$ do
    Solve the FKE on the domain $D$
      $\begin{cases} \frac{\partial \phi}{\partial t}(x,t) = \left(\mathcal{L} - \frac{1}{2}h^T S^{-1} h\right)\phi(x,t), \\ \phi(x, t_{j-1}) = \psi_k(x) \end{cases}$        (9)
    on $[t_{j-1}, t_j]$ by FDM, and get $I^{[t_{j-1},t_j]}\psi_k(x)$.
    Store $I^{[t_{j-1},t_j]}\psi_k$.
  end for
end for
Set up the initial distribution of $x_0$.        %on-line stage
for $j = 1 \to N_t$ do
  Project $u(\cdot, t_{j-1})$ onto the prescribed basis
  functions, and obtain the a priori solution at $t_j$:
    $u^-(x, t_j) = \sum_{k=1}^{N_b} (u(\cdot, t_{j-1}), \psi_k(\cdot))_{L^2(D)} I^{[t_{j-1},t_j]}\psi_k(x).$
  Assimilate the new observation data $y_{t_j}$ into the a
  priori solution $u^-(x, t_j)$:
    $u(x, t_j) = \exp[h^T(x, t_j) S^{-1}(t_j)(y_{t_j} - y_{t_{j-1}})] u^-(x, t_j).$        (10)
  Calculate related statistics by using $u(x, t_j)$ as the
  unnormalized density function at time $t_j$.
end for
```

in the offline stage. In the online stage, our algorithm provides huge savings since the number of POD basis is much smaller.

A. Description of Our POD Algorithm

In this section, we shall state the construction of the POD basis by the method of snapshots in detail. Again let X be a Hilbert space equipped with the inner product $(\cdot,\cdot)_X$ and norm $||\cdot||_X$. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Denote $u(\cdot, t, \omega) \in X, t \in [0, T], \omega \in \Omega$, be the solution of a random dynamic system, i.e., (5), where the randomness is associated with the observations in y_t. We aim to construct a set of POD basis, still denoted by $\{\phi_k\}_{k=1}^{N_{\text{pod}}}$, so that the projection error in X, i.e.,

$$u(\cdot, t, \omega) - \sum_{k=1}^{N_{\text{pod}}} (u(\cdot, t, \omega), \phi_k(\cdot))_X \phi_k(\cdot)$$

is minimized in the norm $L^2_F(0, T; X)$, which is defined as

$$||\circ||^2_{L^2_F(0,T;X)} = \mathbb{E}\left[\int_0^T ||\circ||^2_X dt\right]. \tag{11}$$

The POD basis has an optimal approximation property in the sense of minimizing the projection error. However, it is not easy to be obtained analytically. In the sequel, we shall use the method of snapshots [35] to construct the POD basis numerically.

We approximate the space X by a linear finite dimensional space V with $\dim V = d$, where d is the DOF of the physical space. We choose the time instances as $0 = t_0 < t_1 < \cdots < t_{N_t} = T$ and generate a set of Monte Carlo realizations of the random observations $\{y_{t_i}(\omega_j)\}$ with $0 \le i \le N_t, 1 \le j \le N_{\text{mc}}$.

To obtain solution snapshots, we use FDM to solve (5) along each sample path of the random observation. This procedure provides us with sufficient amount of snapshots $\{u(\cdot, t_i, \omega_j)\}$, with the cardinality $N = (N_t + 1)N_{\text{mc}}$. These solution snapshots are assumed to capture the information of the solution space (or manifold) of the (5) well. We remark that Monte Carlo realizations $\{y_{t_i}(\omega_j)\}$ are served as training purpose, which can be replaced by the historical collected observations data.

Given the set of snapshots of solutions, a linear space V can be spanned, denoted as

$$\begin{aligned} V = \text{span}\{u(x, t_i, \omega_j) : \ & x \in \mathbb{R}^n, \ t_i \in [0, T], \omega_j \in \Omega \\ & i = 0, \ldots, N_t, \ j = 1, \ldots, N_{\text{mc}}\}. \end{aligned} \tag{12}$$

To construct the POD basis $\{\phi_k\}_{k=1}^{N_{\text{pod}}}$, we need to find the minimizers of the following minimization problem

$$\min_{\{\phi_k\}_{k=1}^{N_{\text{pod}}}} \frac{1}{(N_t + 1)N_{\text{mc}}} \sum_{j=1}^{N_{\text{mc}}} \sum_{i=0}^{N_t} \left|\left| u(\cdot, t_i, \omega_j) - \sum_{k=1}^{N_{\text{pod}}} \Big(u(\cdot, t_i, \omega_j), \phi_k(\cdot) \Big)_X \phi_k(\cdot) \right|\right|^2_X \tag{13}$$

subject to the constrains that $(\phi_m(\cdot), \phi_n(\cdot))_X = \delta_{mn}, \ 1 \le m, n \le N_{\text{pod}}$.

By Sirovich [35], we know that the optimization problem (13) can be reduced to an eigenvalue problem

$$Kv = \lambda v \tag{14}$$

where $K \in R^{N \times N}$ is the correlation matrix with the $(i_1 N_{\text{mc}} + j_1, i_2 N_{\text{mc}} + j_2)$th element

$$K_{i_1 N_{\text{mc}} + j_1, i_2 N_{\text{mc}} + j_2} = \frac{1}{N}\left(u(\cdot, t_{i_1}, \omega_{j_1}), u(\cdot, t_{i_2}, \omega_{j_2})\right)_X.$$

We sort the eigenvalues in a decreasing order as $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_N > 0$ and denote the corresponding eigenvectors by v_k, $k = 1, \ldots, N$. It can be shown that the POD basis $\{\phi_k\}_{k=1}^{N_{\text{pod}}}$ are constructed by

$$\phi_k(\cdot) = \frac{1}{\sqrt{\lambda_k}} \sum_{j=1}^{N_{\text{mc}}} \sum_{i=0}^{N_t} (v_k)_{i N_{\text{mc}} + j} u(\cdot, t_i, \omega_j) \tag{15}$$

for $1 \le k \le N$, where $(v_k)_l$ is the lth component of the eigenvector v_k. In addition, we have the follow estimate for the projection error.

Proposition 3.1 (Sec. 3.3.2, [16] or p. 502, [5]): Let $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_N > 0$ denote the positive eigenvalues of K in (14). Then, $\{\phi_k\}_{k=1}^{N_{\text{pod}}}$ constructed according to (15) are the POD

Authorized licensed use limited to: Tsinghua University. Downloaded on April 16,2020 at 04:05:48 UTC from IEEE Xplore. Restrictions apply.

Algorithm 2: Construction of the Basis.

```
for $j = 1 \to N_{mc}$ do % Generation of the solution snapshots
    Generate random observations $\{y_{t_i}(\omega_j)\}$ with $0 \le i \le N_t, 1 \le j \le N_{mc}$.
    Compute the solution $u(x, t, \omega_j)$ of the pathwise robust DMZ (5) by FDM on a domain $D$.
    Store the snapshots of $u$ as $\mathcal{U} = \{u(\cdot, t_i, \omega_j)\}_{i,j}$, $i = 0, \dots, N_t$.
end for
Compute the eigen-decomposition of the correlation matrix $K$ in (14), where the eigen-pairs are denoted by $(\lambda_k, v_k)$, $k = 1, \dots, (N_t + 1)N_{mc}$. % The method of snapshots
Set a tolerance $\text{tol}_\rho$, and let $\rho = 0$.
while $\rho < tol_\rho$ do
    Increase $N_{\text{pod}}$ and calculate $\rho = \frac{\sum_{k=1}^{N_{\text{pod}}} \lambda_k}{\sum_{k=1}^{N} \lambda_k}$.
end while
Store the first $N_{\text{pod}}$ eigen-pairs $\{\lambda_k, v_k\}_{k=1}^{N_{\text{pod}}}$.
for $k = 1 \to N_{\text{pod}}$ do
    Construct the basis $\{\varphi_k\}_{k=1}^{N_{\text{pod}}}$ as in (15).
end for
```

basis and we have the following error formula:

$$\frac{1}{N}\sum_{j=1}^{N_{\text{mc}}}\sum_{i=0}^{N_t}\left\|u(\cdot, t_i, \omega_j) - \sum_{k=1}^{N_{\text{pod}}}\Big(u(\cdot, t_i, \omega_j), \phi_k(\cdot)\Big)_X \phi_k(\cdot)\right\|_X^2$$
$$= \frac{\sum_{k=N_{\text{pod}}+1}^{N} \lambda_k}{\sum_{k=1}^{N} \lambda_k}\left(\frac{1}{N}\sum_{j=1}^{N_{\text{mc}}}\sum_{i=0}^{N_t} ||u(\cdot, t_i, \omega_j)||_X^2\right). \tag{16}$$

In our POD algorithm, the snapshots of solutions $u(x, t_i, \omega_j)$ s are obtained by numerical methods. Therefore, the POD basis $\{\phi_k\}_{k=1}^{N_{\text{pod}}}$ are computed and represented based on the numerical solutions $u(x, t_i, \omega_j)$.

To determine the number of POD basis N_{pod}, we use the decay property of eigenvalues in λ_k and choose the first N_{pod} dominant eigenvalues such that the ratio

$$\rho = \frac{\sum_{k=1}^{N_{\text{pod}}} \lambda_k}{\sum_{k=1}^{N} \lambda_k} \tag{17}$$

is big enough so that $1 - \rho$ is less than a prescribed error threshold tol_ρ, say $\text{tol}_\rho = 1\%$. One would prefer the eigenvalues decaying as fast as possible so that the fewer basis can ensure the higher accuracy. We refer the interested reader to [33] for some estimates on the rate of decay of the eigenvalues in the Karhunen–Loève expansion, which are essentially the eigenvalues in POD algorithm. Finally, we summarize the construction of the POD basis in Algorithm 2.

In our numerical experiments in Section IV-B, we observed that in the asymptotic regime, the accumulated ratio (17) obtained using our constructed basis approaches one exponentially fast as N_{pod} increases, i.e.,

$$1 - \rho \sim \mathrm{e}^{-cN_{\text{pod}}}, \quad c > 0. \tag{18}$$

This can significantly reduce the number of the basis involving in the online computation.

Remark 3.1: In Section III-B, we shall prove that under mild assumptions the error between the solution obtained by the POD basis $\{\phi_k\}_{k=1}^{N_{\text{pod}}}$ and the reference solution has exponential decay property.

Remark 3.2: We assume that the number of time instances $(N_t + 1)$ and the number of sample paths N_{mc} are chosen in such a way that the solution snapshots capture the information of the solution space (or manifold) of the (5). The choice of parameter sample points is a critical question that arises in the POD method to compute the basis, especially for systems with time-dependent and/or stochastic parameters; see [5, Sec. 6]. Our strategy in choosing the sample points may not be optimal though, the numerical results reveal that the (5) has certain low dimensional structures in the solution space. Our result can be viewed as a recent progress in the POD method for solving stochastic dynamic problems.

Remark 3.3: If the stochastic dynamic problems possess some kind of ergodicity property, one can choose any one sample path of the observation y_t and replace the norm $L_F^2(0, T; X)$ by $L^2(0, T; X)$ in computing the projection error of the POD basis. That is, given any $\omega_0 \in \Omega$, the POD basis $\{\phi_k\}_{k=1}^{N_{\text{pod}}}$ minimizes the error

$$\int_0^{T_{\text{mix}}}\left\|u(\cdot, t, \omega_0) - \sum_{k=1}^{N_{\text{pod}}}(u(\cdot, t, \omega_0), \phi_k(\cdot))_X \phi_k(\cdot)\right\|_X^2 dt$$

where T_{mix} should be beyond the mixing time.

B. Convergence Analysis

Our POD algorithm significantly improves the performance of the offline and online algorithm developed in [26]. In this section, we shall first discuss the connection between the offline and online algorithm in [26] and the splitting-up method in [6], so that the convergence of the DMZ (2) in $L_F^2(0, T; H^1(\mathbb{R}^n))$ is applicable in our POD algorithm, where $H^1(\mathbb{R}^n)$ denotes Sobolev space $W^{1,2}(\mathbb{R}^n)$. Furthermore, under the assumption of certain ergodicity property in Remark 3.3, we can also show the convergence of the POD algorithm to the pathwise robust DMZ (5) in L^2 norm without the boundedness condition on f, g, and h as in Assumption [As-1]-[As-2].

1) Analysis Based on the Splitting-Up Method: Let us assume that the observation time sequences are uniform, namely $t_{j+1} - t_j = \Delta t$, $j = 0, \dots, N_t - 1$. The observation data at time t_j is denoted by y_{t_j} and $\Delta y_j = y_{t_j} - y_{t_{j-1}}$. Let us recall the splitting-up method briefly. To be consistent with the settings in [6], we assume in this section that $S = I$, the identity matrix. The DMZ (2) has been decomposed into two processes U and U^- in the time intervals $[t_{i-1}, t_i)$, $i = 1, \dots, N_t$, which satisfy

Authorized licensed use limited to: Tsinghua University. Downloaded on April 16,2020 at 04:05:48 UTC from IEEE Xplore. Restrictions apply.

$$
\begin{aligned}
dU(t) &= \left(\mathcal{L}U - \frac{\mu}{2}U\right) dt \\
U(t_{i-1}) &= \begin{cases} U^-(t_{i-1}), & \text{if } i = 2, 3, \ldots, N_t \\ \sigma_0, & \text{if } i = 1 \end{cases}
\end{aligned}
\tag{19}
$$

and

$$
\begin{aligned}
dU^-(t) + \frac{\mu}{2}U^- dt &= U^- h^T dw_t \\
&= U^- h^T dy_t - U^- h^T h dt \\
U^-(t_{i-1}) &= U(t_i)
\end{aligned}
\tag{20}
$$

respectively, where $\mathcal{L}$ is the operator in (3) and σ_0 is the unnormalized conditional density function of the initial state x_0. Notice the following two important facts:

1) U satisfies FKE (6) or (9) in Algorithm 1 with $\mu = h^T h$ in (19);
2) U^- can be solved explicitly, i.e.,

$$U^-(t) = U^-(t_{i-1}) e^{\int_{t_{i-1}}^{t} h^T dy_s + \frac{1}{2}\int_{t_{i-1}}^{t}\left(h^T h - \mu\right) ds}.$$

If $\mu = h^T h$, then

$$
\begin{aligned}
U^-(t) &= U^-(t_{i-1}) e^{\int_{t_{i-1}}^{t} h^T(s) dy_s} \\
&\approx U^-(t_{i-1}) e^{h^T(t_{i-1})\Delta y_i}.
\end{aligned}
$$

This is exactly the initial update (10) in Algorithm 1.

Now, we recall the convergence result in [6]. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let us make the following generic assumptions on the drift and observation terms as those in [6].

[As-1] The drift term and the diffusion term are bounded, i.e.,

$$
\begin{aligned}
f &\in L^\infty(\mathbb{R}^n \times (0, \infty); \mathbb{R}^n) \\
g &\in L^\infty(\mathbb{R}^n \times (0, \infty); L(\mathbb{R}^n, \mathbb{R}^n))
\end{aligned}
$$

with f and g be Lipschitz in x, uniformly in t.

[As-2] The observation term is also bounded, i.e., $h \in L^\infty(\mathbb{R}^n \times (0, \infty); \mathbb{R}^m)$.

[As-3] The operator gQg^T is uniformly elliptic, i.e., for all $\xi \in \mathbb{R}^n$, there exists a constant $\alpha > 0$ such that

$$\xi^T(gQg^T)\xi \geq \alpha|\xi|^2.$$

Remark 3.4: Although [As-1] and [As-2] seem to be very restrictive, as [6] claimed in the end, "this limitation is purely technical" for the mathematical proof. For further discussions on the growth of f and h, we refer the interested readers to [26] and references therein.

Proposition 3.2 (Th. 3.1, [6]): Assume [As-1]-[As-3] hold, then we have

1) $U, U^- \to \sigma$ as $\Delta t \to 0$ in $L_F^2(0, T; H^1(\mathbb{R}^n))$ and $L_F^2(0, T; L^2(\mathbb{R}^n))$, respectively;
2) $U(t), U^-(t) \to \sigma(t)$ as $\Delta t \to 0$ in $L^2(\Omega, \mathcal{A}, \mathbb{P}; L^2(\mathbb{R}^n))$, $\forall t \in [0, T]$;

where σ is the solution to the DMZ (2), the norms of $L_F^2(0, T; V)$ and $L^2(\Omega, \mathcal{A}, \mathbb{P}; V)$ are defined as in (11) and

$$||\sigma||^2_{L^2(\Omega, \mathcal{A}, \mathbb{P}; V)}(t) = \mathbb{E}||\sigma||^2_V(t)$$

respectively, where V is some function space in concern.

Let us denote $U^{N_{\text{pod}}}$ the approximation solution to U obtained by the Galerkin method in the linear space spanned by the POD basis $\{\phi_k\}_{k=1}^{N_{\text{pod}}}$. We expect that $U^{N_{\text{pod}}} \to U$ in $L_F^2(0, T; H^1(\mathbb{R}^n))$, as $N_{\text{pod}} \to \infty$.

Recall the estimate of the projection error in the linear space spanned by the GHF [28]. We define the n-dimensional GHF as

$$\mathcal{H}_{\mathbf{k}}^{\alpha,\beta}(\mathbf{x}) := \prod_{j=1}^{n} \mathcal{H}_{k_j}^{\alpha_j,\beta_j}(x_j)$$

where $\mathbf{x} = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$, $\mathbf{k} = (k_1, \ldots, k_n)$, and

$$\mathcal{H}_k^{\alpha,\beta}(x) = \left(\frac{\alpha}{2^k k! \sqrt{\pi}}\right)^{\frac{1}{2}} H_k(\alpha(x - \beta)) e^{-\frac{1}{2}\alpha^2(x-\beta)^2}$$

with $H_n(x)$ be the univariate physical Hermite polynomials, α, β are two parameters.

Suppose the prescribed orthonormal basis are $\{\mathcal{H}_{\mathbf{k}}^{\alpha,\beta}(\mathbf{x})\}_{\mathbf{k} \in \Omega_{N_h}}$, where N_h is the total number of the basis and $\Omega_{N_h} := \{\mathbf{k} : \; |\mathbf{k}|_\infty \leq N_h^{\frac{1}{n}}\}$ with $|\mathbf{k}|_\infty = \max_{i \in \{1, \ldots, n\}} k_i$. Then, we have the following error estimate on the projection error.

Proposition 3.3 (Th. 2.1, [28]): Given $U \in W_{\alpha,\beta}^r(\mathbb{R}^n)$, we have for any $0 \leq l \leq r$

$$\left|\left|P_{N_h}^{\alpha,\beta} U - U\right|\right|_{W_{\alpha,\beta}^l(\mathbb{R}^n)} \lesssim N_h^{\frac{l-r}{2n}} |U|_{W_{\alpha,\beta}^r(\mathbb{R}^n)} \tag{21}$$

where $P_{N_h}^{\alpha,\beta}$ is the projection operator

$$P_{N_h}^{\alpha,\beta} : \; W_{\alpha,\beta}^l(\mathbb{R}^n) \to \text{span}\left\{\mathcal{H}_{\mathbf{k}}^{\alpha,\beta}, \mathbf{k} \in \Omega_{N_h}\right\}$$

defined as

$$P_{N_h}^{\alpha,\beta} U(\mathbf{x}) := \sum_{\mathbf{k} \in \mathbf{\Omega_{N_h}}} (\mathcal{H}_{\mathbf{k}}^{\alpha,\beta}, U)_{W_{\alpha,\beta}^l(\mathbb{R}^n)} \mathcal{H}_{\mathbf{k}}^{\alpha,\beta}(\mathbf{x})$$

and the norm and seminorm of $W_{\alpha,\beta}^r(\mathbb{R}^n)$ are defined as

$$
\begin{aligned}
||U||^2_{W_{\alpha,\beta}^r(\mathbb{R}^n)} &:= \sum_{0 \leq |\mathbf{k}|_1 \leq r} ||\mathcal{D}_{\mathbf{x}}^{\mathbf{r}} U||^2 \\
|U|^2_{W_{\alpha,\beta}^r(\mathbb{R}^n)} &:= \sum_{j=1}^{n} ||\mathcal{D}_{x_j}^{r} U||^2
\end{aligned}
$$

with $\mathcal{D}_{\mathbf{x}}^{\mathbf{r}} := \prod_{i=1}^{n} \mathcal{D}_{x_i}^{r_i}$ and $\mathcal{D}_{x_i} = \partial_{x_i} + \alpha_i^2(x_i - \beta_i)$.

Remark 3.5: The space $W_{\alpha,\beta}^0(\mathbb{R}^n) = L^2(\mathbb{R}^n)$ and $W_{\alpha,\beta}^r(\mathbb{R}^n) \subset H^r(\mathbb{R}^n)$ by [28, Corollary 3.2], where $H^r(\mathbb{R}^n)$ denotes Sobolev space $W^{r,2}(\mathbb{R}^n)$.

It is clear to see that if the solution U is extremely smooth, then the projection error decreases faster than polynomials of any degree N_h. That is, it may present exponential convergence with respect to N_h. Notice that the basis here is prescribed without any information of the solution U. One can expect that the elaborately selected N_{pod} POD basis according to the solution snapshots of U will yield a smaller projection error.

Proposition 3.4: If for almost all $w \in \Omega$, $t \in [0, T]$, $U \in L_F^2(0, T; W_{\alpha,\beta}^r(\mathbb{R}^n))$, then we get

Authorized licensed use limited to: Tsinghua University. Downloaded on April 16,2020 at 04:05:48 UTC from IEEE Xplore. Restrictions apply.

$$\begin{aligned}&\left|\left|U^{N_{\text{pod}}}-U\right|\right|^2_{L^2_F(0,T;H^1(\mathbb{R}^n))}\\&\qquad\lesssim N_{\text{pod}}^{\frac{1-r}{n}}\,(1+T)\,||U||^2_{L^2_F(0,T;W^r_{\alpha,\beta}(\mathbb{R}^n))}\,. \qquad (22)\end{aligned}$$

Proof: It is clear to see that

$$\begin{aligned}&\left|\left|U^{N_{\text{pod}}}-U\right|\right|^2_{L^2_F(0,T;H^1(\mathbb{R}^n))}\\
&\lesssim\left|\left|U^{N_{\text{pod}}}-U\right|\right|^2_{L^2_F(0,T;W^1_{\alpha,\beta}(\mathbb{R}^n))}\\
&\le\left|\left|P^{\alpha,\beta}_{\Phi}U-U\right|\right|^2_{L^2_F(0,T;W^1_{\alpha,\beta}(\mathbb{R}^n))}\\
&\quad+\left|\left|P^{\alpha,\beta}_{\Phi}U-U^{N_{\text{pod}}}\right|\right|^2_{L^2_F(0,T;W^1_{\alpha,\beta}(\mathbb{R}^n))}\\
&\le\left|\left|P^{\alpha,\beta}_{N_{\text{pod}}}U-U\right|\right|^2_{L^2_F(0,T;W^1_{\alpha,\beta}(\mathbb{R}^n))}\\
&\quad+\left|\left|P^{\alpha,\beta}_{\Phi}U-U^{N_{\text{pod}}}\right|\right|^2_{L^2_F(0,T;W^1_{\alpha,\beta}(\mathbb{R}^n))}\\
&\lesssim N_{\text{pod}}^{\frac{1-r}{n}}\,\mathbb{E}\left\{\int_0^T|U|^2_{W^r_{\alpha,\beta}(\mathbb{R}^n)}dt\right.\\
&\qquad\left.+\int_0^T\int_0^t|U|^2_{W^r_{\alpha,\beta}(\mathbb{R}^n)}dsdt\right\} \qquad (23)\end{aligned}$$

where the first inequality follows by Remark 3.5, the third inequality is due to the minimizer property of POD basis and the last inequality is obtained by (21) and the similar argument in [28, Th. 3.3]. ■

Remark 3.6: The similar result in Proposition 3.4 can be obtained for bounded domain $D\subset\mathbb{R}^n$. We refer the interested readers to [34].

From Propositions 3.2 and 3.4, we get

Theorem 3.5: Assume [As-1]–[As-3] hold, then we have

1) $U^{N_{\text{pod}}}\to\sigma$ in $L^2_F(0,T;H^1(\mathbb{R}^n))$, as $N_{\text{pod}}\to\infty$ and $\Delta t\to 0$ subsequently;
2) $U^{N_{\text{pod}}}(t)\to U(t)$ in $L^2(\Omega,\mathcal{A},\mathbb{P};H^1(\mathbb{R}^n))$, for all $t\in[0,T]$.

2) *Analysis Based on the Offline and Online Algorithm of the Pathwise Robust DMZ Equation:*

Suppose the assumption in Remark 3.3 holds, then we obtain the POD basis $\{\phi_k\}_{k=1}^{N_{\text{pod}}}$ from one given sample path $\omega_0\in\Omega$. The boundedness condition in [As-1]–[As-3] can be replaced by some mild growth condition.

For all $(x,t)\in\mathbb{R}^n\times[0,T]$:

[As-4]

$$N(x,t)+\frac{3}{2}n||gQg^T||_\infty+|f-D_wK|\le C_1$$

[As-5]

$$e^{-\sqrt{1+|x|^2}}\left[14n||gQg^T||_\infty+4|f-D_wK|\right]\le C_2$$

where

$$K(x,t):=h^T(x,t)S^{-1}(t)y_t(w_0)$$

$$\begin{aligned}N(x,t):=&-\frac{\partial}{\partial t}(h^TS^{-1})y_t(\omega_0)-\frac{1}{2}D^2_wK\\&+\frac{1}{2}D_wK\cdot\nabla K-f\cdot\nabla K-\frac{1}{2}(h^TS^{-1}h)\end{aligned}$$

with

$$D^2_w(\circ):=\sum_{i,j=1}^n(gQg^T)_{ij}\frac{\partial^2\circ}{\partial x_i\partial x_j}$$

$$D_w(\circ):=\left[\sum_{j=1}^n(gQg^T)_{ij}\frac{\partial\circ}{\partial x_j}\right]_{i=1}^n.$$

Moreover, let $B_R\in\mathbb{R}^n$ be the ball centered at the origin with the radius $R>0$, we assume further that on any B_R, we have

[As-6] $N(x,t)\le C_3$,

[As-7] There exists some $\alpha\in(0,1)$ such that

$$|N(x,t)-N(x,t;\bar{t})|\le C_4|t-\bar{t}|^\alpha$$

for all $(x,t)\in D\times[0,T]$, $\bar{t}\in[0,T]$, where $N(x,t;\bar{t})$ is $N(x,t)$ with $y_t=y_{\bar{t}}$.

In [26, Ths. 3.1 and 3.2], the second and the third author of this paper show that for any $T>0$, the solution of the pathwise robust DMZ (5) in $\mathbb{R}^n\times[0,T]$, denoted as $u(x,t)$, can be approximated by the solution to (5) restricted on the ball B_R with Dirichlet boundary condition, denoted as u_R. Moreover, u_R can be approximated in L^1 sense by the solution of (5) freezing y_t in $[t_j,t_{j+1}]$ to be y_{t_j}, $j=0,\ldots,N_t-1$.

Our POD algorithm is performed on some bounded domain $D\subset\mathbb{R}^n$. In [25], the second author and her coworker showed if we solve (9) by ofline and online algorithm with generalized Jacobi polynomial for scalar NLF problem, then the error of the approximate solution decays exponentially fast with respect to the number of the basis, if the solution of (9) is smooth enough. We claim that this result is also valid for $x\in\mathbb{R}^n$, $n\ge1$, under the assumption that

[As-8]

$$\frac{1}{2}\sum_{i,j=1}^n\frac{\partial^2}{\partial x_i\partial x_j}(gQg^T)_{ij}-\nabla\cdot f-\frac{1}{2}\,h^TS^{-1}h\le C_5$$

[As-9]

$$\frac{1}{2}(gQg^T)\nabla\left(\ln\left(\frac{g^Tg}{\mathbf{w}_{-1,-1}}\right)\right)-f\le C_6$$

where $\mathbf{w}_{-1,-1}$ is the weight associated with the Jacobi polynomials, $\mathbf{w}_{-1,-1}=\prod_{i=1}^n(1-x_i)^{-1}(1+x_i)^{-1}$.

If the POD basis $\{\phi_k\}_{k=1}^{N_{\text{pod}}}$ obtained by minimizing the projection error in the norm $X=W^r_{\mathbf{w}_{-1,-1}}(I)$ in (13), $I=[-1,1]$, with Remark 3.3, one can argue similarly as in (23) to get that

$$\left|\left|u^{N_{\text{pod}}}-u\right|\right|^2_{L^2\left(0,T;W^1_{\mathbf{w}_{-1,-1}}\right)}\le C^*N_{\text{pod}}^{2(1-r)} \qquad (24)$$

where $u^{N_{\text{pod}}}$ is the numerical solution obtained by the Galerkin method in the spanned linear space span$\{\phi_k(x),$

Authorized licensed use limited to: Tsinghua University. Downloaded on April 16,2020 at 04:05:48 UTC from IEEE Xplore. Restrictions apply.

$k = 1, \dots, N_{\text{pod}}\}$, and

$$\| \circ \|^2_{W^r_{\mathbf{w}_{-1,-1}}(I)} = \sum_{k=0}^{r} | \circ |^2_{W^r_{\mathbf{w}_{-1,-1}}(I)}$$

$$| \circ |^2_{W^r_{\mathbf{w}_{-1,-1}}(I)} = \langle \partial^r u, \partial^r u \rangle_{\mathbf{w}_{-1+r,-1+r}} .$$

IV. Numerical Results

In this section, we are interested in investigating the approximation properties of our POD algorithm and the computational savings over existing methods. The experiments are performed in 2-D NLF problems. We shall clarify the settings of these two problems first.

Example 1. Almost linear problem: This problem is modeled by an SDE in the Ito as follows:

$$\begin{cases} dx_1 = dv_1 \\ dx_2 = dv_2 \\ dy_1 = x_1(1 + 0.2\cos(x_2))dt + dw_1 \\ dy_2 = x_2(1 + 0.2\cos(x_1))dt + dw_2 \end{cases} \tag{25}$$

where $E[dw_t dw_t^T] = I_2 dt$, $E[dv_t dv_t^T] = 0.1 I_2 dt$, with $w = [w_1, w_2]^T$, $v = [v_1, v_2]^T$, I_2 be the identity matrix of size 2×2. The states are two independent standard Brownian motions. The initial state is $x(0) = [x_1(0), x_2(0)]^T = [1, 1.2]^T$. We shall denote the state in vector form $x(t) = [x_1(t), x_2(t)]^T$. The total experimental time is $T = 20$.

Example 2. Cubic sensor problem: The observations in cubic sensor problem have higher nonlinearity than those in (25), which may cause problem when using the conventional extended Kalman filter (EKF). It is modeled in the following equation:

$$\begin{cases} dx_1 = (-0.4x_1 + 0.1x_2)dt + dv_1 \\ dx_2 = -0.6x_2 dt + dv_2 \\ dy_1 = (x_1^3 + x_2)dt + dw_1 \\ dy_2 = (x_2^3 + x_1)dt + dw_2 \end{cases} \tag{26}$$

where $E[dw_t dw_t^T] = I_2 dt$ and $E[dv_t dv_t^T] = 0.1 I_2 dt$. The initial state is $x(0) = [x_1(0), x_2(0)]^T = [0.1, 0.05]^T$. The total experimental time is $T = 10$.

Remark 4.1: It can be easily verified that Examples 1 and 2 satisfy the Assumptions [As-4]–[As-9].

A. Comparison With Existing Methods

In this section, we shall mainly compare the estimation performance and real-time manner of our POD algorithm with the reference solutions, EKF and PF in two examples (25) and (26), respectively.

In both examples, the real state is generated by solving the SDE (25) or (26) for x in the time interval $[0, T]$ (T $= 20$ or 10) with time step $dt = 0.01$ using the Euler–Maruyama method [22]. This provides us the values of the real state at discrete times $t_j = jdt$, $j = 1, \dots, 2000$ (or 1000).

Example 1. Almost linear problem: To obtain sufficient amount of snapshots, as described in Algorithm 2, we partition the time interval [0,20] with observation time step $\Delta t = 0.2$, generate $N_{\text{mc}} = 500$ random observations $\{y_{t_j}(\omega_i)\}$ with $1 \leq j \leq 100, 1 \leq i \leq 500$, and use FDM to solve FKE (6) with initial density function $\sigma_0(x) = \exp(-2|x|^2)$ along each sample path ω_i within the spacial domain $[-5, 5]^2$ and 1-D mesh size $\Delta x = 10/128$. The Courant–Friedrichs–Lewy (CFL) stability condition of FDM is satisfied by choosing the time step as $\frac{dt}{10}$. The POD basis functions are constructed as in (15).

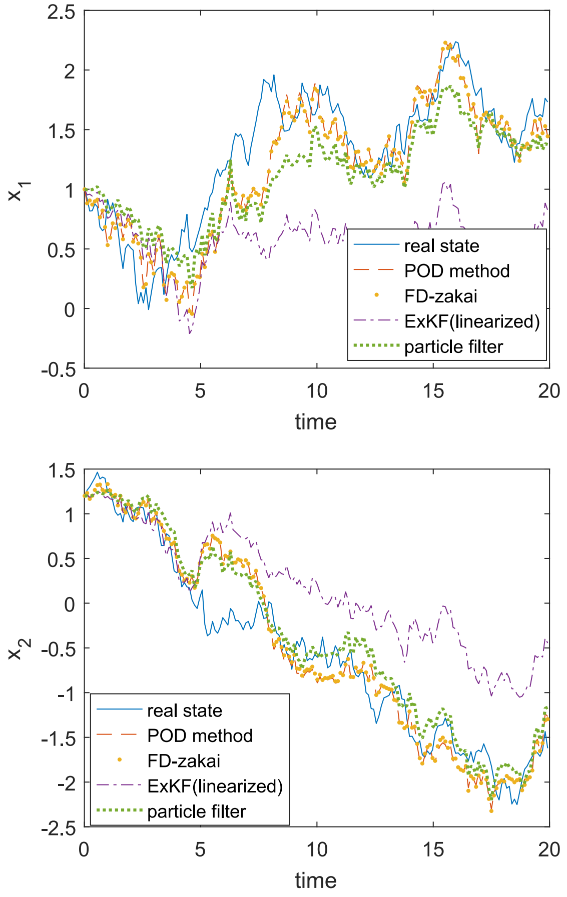


Fig. 1. Estimations of the almost linear problem (25) obtained by our POD algorithm (in red dashed line), the reference solution (in orange dot), the EKF (in purple dashed line), and the PF with 5000 particles (in green dot) versus time have been plotted. The blue line is the true state generated by one realization.

In Fig. 1, we plot the estimations of both two states in one realization obtained by our POD algorithm with the number of the POD basis $N_{\text{pod}} = 70$. The reference solution is obtained by solving FKE (6) directly online by FDM. It seems that all methods give acceptable experimental results except for the EKF. Yet our algorithm gives significant computational savings. The CPU time of the reference method is 41.54 s, that of PF is 156.58 s, while that of our algorithm is only 2.97 s (the time for online computing), which is almost $\frac{1}{14}$ of the reference method and $\frac{1}{53}$ of the PF. We remark that the efficiency of the PF is closely related to the number of particles. In this example, we use 5000 particles to avoid explosion in the tracking, which might happen frequently if only 1000 particles are used.

Authorized licensed use limited to: Tsinghua University. Downloaded on April 16,2020 at 04:05:48 UTC from IEEE Xplore. Restrictions apply.

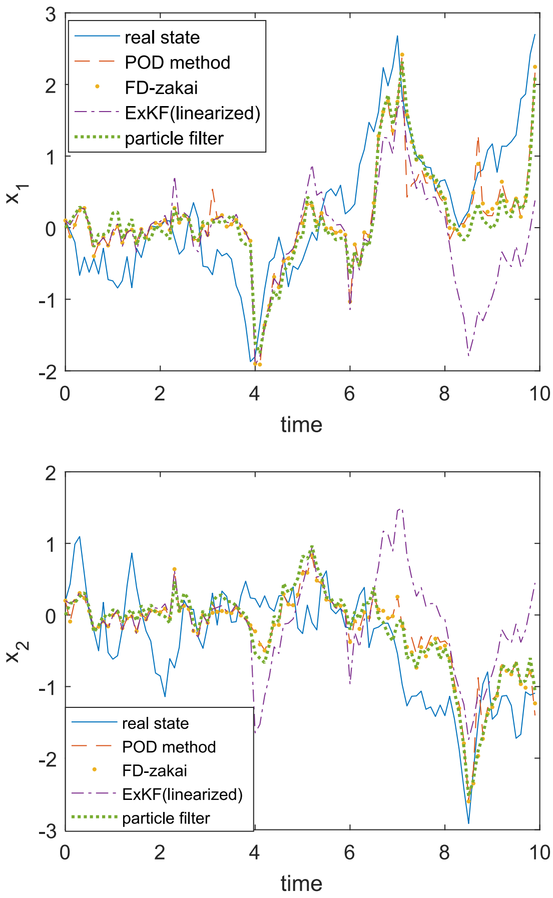


Fig. 2. Estimations of the cubic sensor problem (26) obtained by our POD algorithm (in red dashed line), the reference solution (in orange dot), and the EKF (in purple dashed-dot), and the PF with 1000 particles (in green dot) versus time have been plotted. The blue line is the true state generated by one realization.

Example 2. Cubic sensor problem: In this example, the reference solution is also obtained by using FDM to solve FKE (6) with initial density function $\sigma_0(x) = \exp(-|x|^4/4)$. The spacial domain is $[-3,3] \times [-3,3]$ with 1-D mesh size $\Delta x = 6/128$. The time step is chosen to be $\frac{dt}{40}$ satisfying the CFL stability condition. $N_{\text{pod}} = 100$ basis are constructed according to (15) after the similar training in Example 1.

In Fig. 2, we display the similar results as those in Fig. 1. The CPU time of the reference method is 99.83 s, that of PF with 1000 particles is 14.94 s, while that of our algorithm is only 2.21 s. Compared to Example 1, fewer particles are required to avoid the explosion. This reveals that in some sense the random walk is nontrivial, since it is harder to catch its behavior by Monte Carlo simulation.

Remark 4.2: The CPU time of the reference solution in Example 2 is significantly longer than that in Example 1, since the time discretization in Example 2 is 4 times finer than that in Example 1. We believe that the finer time discretization is due to the higher nonlinearity. Notice that all the computations of the reference solution are conducted in the online stage.

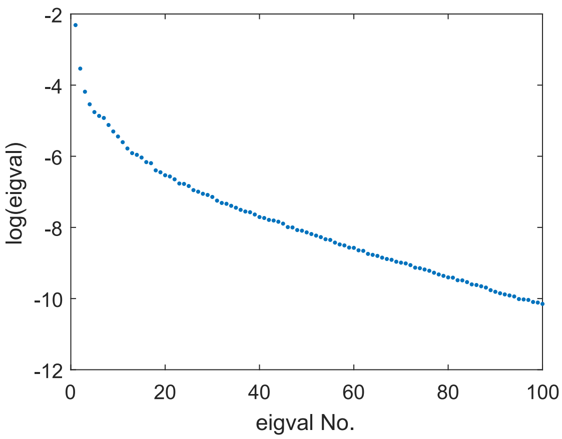


Fig. 3. Decay property of eigenvalues in the POD algorithm.

We repeat the experiments for $N_{\text{path}} = 300$ times and record the mean square errors averaged over 300 sample paths. The MSE for certain method is defined as

$$\text{MSE}(\mathbf{x}) := \frac{1}{300} \sum_{i=1}^{300} \frac{1}{N_t} \sum_{j=1}^{N_t} |\mathbf{x}^i(t_j) - \mathbf{x}^i_{\text{true}}(t_j)|$$

where $|\cdot|$ is the Euclidean distance, $\mathbf{x}^i$ is the numerical estimation obtained by different methods for the ith sample path of the true state $\mathbf{x}^i_{\text{true}}$. We find that the MSE of our POD algorithm is 0.8849, that of the reference method is 0.8523, while that of the PF with 1000 particles is 0.9188. Notice that the further compression by POD algorithm has comparable accuracy with the reference solution, which is only with the difference less than 4%.

B. More Discussions on Our POD Algorithm

In our POD algorithm, there are still some parameters to be tuned in, for example, the number of the constructed basis, the choice of the training snapshot solutions, etc. In this section, we shall do the numerical experiments on Example 2, since both examples present similar behaviors and low-dimensional structure of Example 2 seems to be more difficult to be captured due to the cubic observation function.

1) Decay Property of the Relative Errors Versus the Number of the Basis: It has been shown in Proposition 3.1 that the relative L^2 error of the training solutions can be represented by the quantity $1-\rho$, where ρ is defined in (17)

$$\frac{\sum_{k=N_{\text{pod}}+1}^{N} \lambda_k}{\sum_{k=1}^{N} \lambda_k} = 1 - \rho$$

where N and N_{pod} are the total number of snapshots and that of the POD basis. In Fig. 3, we plot the quantity of the eigenvalues in decreasing order. One can see that the eigenvalues decay exponentially fast, which implies that the quantity $1-\rho$ also decays exponentially fast with respect to the number of the basis. We use regression to fit the data and find the decay speed

Authorized licensed use limited to: Tsinghua University. Downloaded on April 16,2020 at 04:05:48 UTC from IEEE Xplore. Restrictions apply.

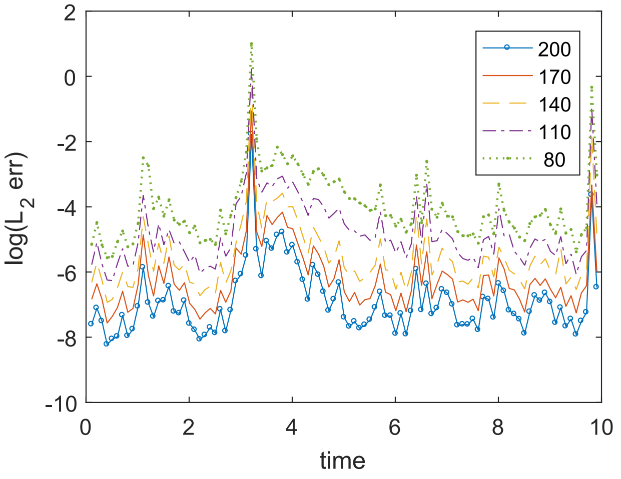


Fig. 4. Relative L^2 errors of one realization versus time are plotted with the number of the basis being $N_{\text{pod}} = 80, 110, 140, 170$, and 200.

of the quantity $1-\rho$ is proportional to $\exp(-C_1 N_{\text{pod}})$, with $C_1 = 0.0422$.

In Section III-B, we show theoretically in Proposition 3.4 that the relative error

$$\frac{||U^{N_{\text{pod}}} - U||^2_{L^2_{\mathcal{F}}(0,T;H^1(\mathbb{R}^n))}}{||U||^2_{L^2_{\mathcal{F}}(0,T;W^r_{\alpha,\beta}(\mathbb{R}^n))}} \tag{27}$$

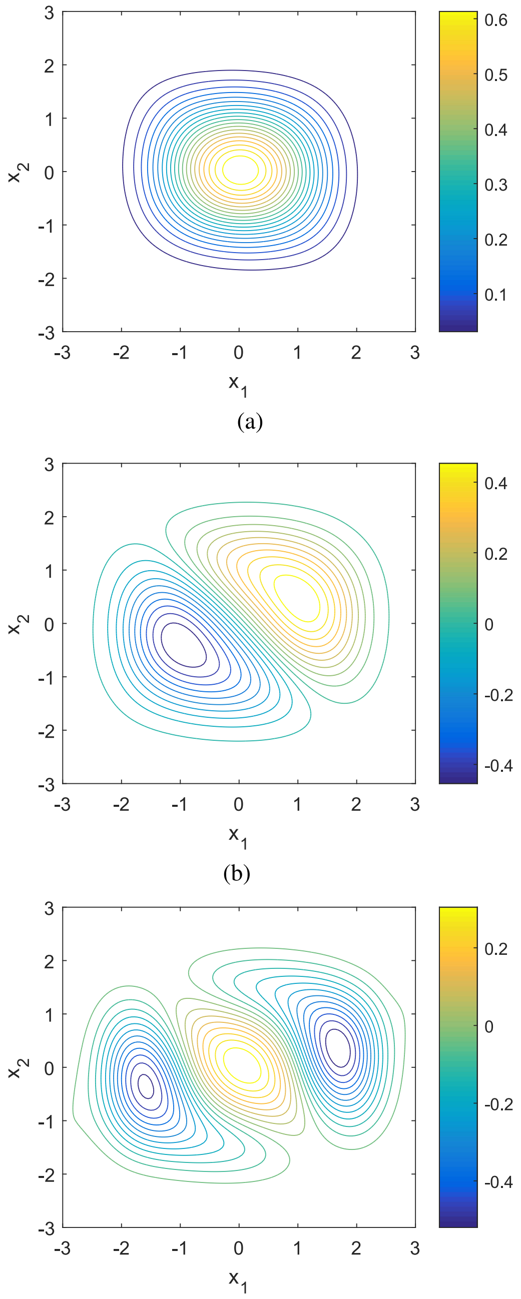
is controlled by $N_{\text{pod}}^{\frac{1-r}{n}}(1+T)$. In other words, if the reference solution is smooth enough, the relative error (27) can present exponential decay as $N_{\text{pod}} \to \infty$. Here, we generate $N_{\text{path}} = 300$ sample paths x_t and observation paths y_t. We record the relative L^2 error defined as

$$\text{err}(t) := \frac{||u_{\text{ref}} - u_{\text{pod}}||_{L^2}}{||u_{\text{ref}}||_{L^2}}$$

of the numerical solution obtained using fixed number of the basis ranging from 1 to 200 and for each sample paths. For fixed number of the basis, one averages the relative errors over all these 300 sample paths.

We use regression to fit the averaged relative L^2 errors over $N_{\text{path}} = 300$ sample paths between two methods and find the decay rate is proportional to $\exp(-C_2 N_{\text{pod}})$ with $C_2 = 0.0293$ and 0.0195 for Examples 1 and 2, respectively, where N_{pod} is the number of the basis. The relatively slow decay in the cubic sensor problem may imply that it has more complicated structure than the almost linear problem. This also explains why in Section IV-A different N_{pod} are chosen to guarantee $\mathcal{O}(10^{-2})$ relative error in two examples.

In Fig. 4, we plot the relative L^2 error evolution of one realization in the cubic sensor problem versus different number of the basis. One finds that at each time discretization the error decays monotonically as the number of the basis increases. More significant observation is that just increasing the number of the basis cannot improve resolutions, if the basis does not carry enough information after the training. This phenomena can be seen from Fig. 4 at time instant around $t = 3.7$ and 9.8, where the peaking of the errors are not relieved even after doubling the number of the basis.



(c)

Fig. 5. Profiles of the first three basis in the cubic sensor problem (26) from $N_{\text{mc}} = 500$ are displayed. (a) First basis. (b) Second basis. (c) Third basis.

2) Selection of the Training Solutions: The construction of the basis depends highly on the training set. How the training set affects the basis? Recall that we generate $N_{\text{mc}} = 500$ sample paths of the states and the observations, and the snapshots are $\mathcal{U} = \{u(\cdot, t_j, \omega_i)\}$, here $\omega_i \in \Omega, i = 1, \ldots, N_{\text{mc}}, t_j = j\Delta t, j = 1, \ldots, N_t (= \frac{T}{\Delta t})$ with $\Delta t = 0.2$. We also try to generate less sample paths such as $N_{\text{mc}} = 125$ or $N_{\text{mc}} = 250$, and find that the

Authorized licensed use limited to: Tsinghua University. Downloaded on April 16,2020 at 04:05:48 UTC from IEEE Xplore. Restrictions apply.

first few dominant basis from various $N_{\rm pod}$ are hard to distinguish by eyes from various $N_{\rm mc}$.

In Fig. 5, we show the first three dominant basis obtained with $N_{\rm mc} = 500$. The higher order of the basis is, the more local structures of the solutions have been captured. It would be interesting and challenging to generate the snapshots capable of capturing most of the variations of the solution space. This issue will be investigated in our on-going work, especially in the higher dimensional NLF problems.

V. Conclusion

The POD algorithm can be viewed as a compression in using offline and online algorithm developed in [26]. The beforehand numerical experiments or history data are necessary for the training purpose. The low-dimensional structures in the solution space of the FKE have been captured and used to build the basis by the method of snapshots in advance. Similar as the offline and online algorithm, in the offline stage, we not only construct the basis, but also compute the propagation of the basis according to the FKE operator. Then, in the online stage, we only need to compute the projection coefficients of the solutions on the basis and update the corresponding results with the new-coming observations. Since the basis functions in the POD algorithm are problem-dependent and more adaptive to the target solution space, the DOF in the POD algorithm is much smaller than other existing methods, which helps us alleviate the curse of dimensionality to a certain extent. Therefore, the POD algorithm enables us to solve the NLF problem in a real-time manner.

Under some generic assumptions as in [6], we provide the convergence analysis of our POD algorithm theoretically. Two 2-D NLF problems: almost linear problem and cubic sensor problem have been investigated numerically. The theoretical convergence rate has been verified numerically. It is shown numerically that our POD algorithm yields as good approximations as the reference solution obtained by FDM. But our algorithm can be much more efficient. We expect even better performance of efficiency in higher-dimensional NLF problems, which is one of our future topics.

Some further discussions on the POD algorithm, such as the choice of number of POD basis, the number of snapshots, etc, have been included. It seems that it is unnecessary to provide a huge amount of snapshots for training in our numerical experiments, but how to choose the parameter sample points in computing the solution snapshots remains a challenging open question, especially for systems with time-dependent and/or stochastic parameters, which is also raised in a recent review paper [5] and will be our future study.

References

[1] H. Abdi and L. Williams, "Principal component analysis," *Wiley Interdiscip. Rev. Comput. Statist.*, vol. 2, no. 4, pp. 433–459, 2010.

[2] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[3] A. Bain and D. Crisan, *Fundamentals of Stochastic Filtering* (Stochastic Modeling and Applied Probability, 60). Berlin, Germany: Springer, 2009.

[4] J. Baras, G. Blankenship, and W. Hopkins, "Existence, uniqueness, and asymptotic behavior of solutions to a class of Zakai equations with unbounded coefficients," *IEEE Trans. Autom. Control.*, vol. AC-28, no. 2, pp. 203–214, Feb. 1983.

[5] P. Benner, S. Gugercin, and K. Willcox., "A survey of projection-based model reduction methods for parametric dynamical systems," *SIAM Rev.*, vol. 57, no. 4, pp. 483–531, 2015.

[6] A. Bensoussan, R. Glowinski, and A. Rascanu, "Approximation of the Zakai equation by the splitting up method," *SIAM J. Control Optim.*, vol. 28, no. 6, pp. 1420–1431, 1990.

[7] G. Berkooz, P. Holmes, and J. L. Lumley, "The proper orthogonal decomposition in the analysis of turbulent flows," *Annu. Rev. Fluid Mech.*, vol. 25, no. 1, pp. 539–575, 1993.

[8] M. Cheng, T. Y. Hou, M. Yan, and Z. Zhang, "A data-driven stochastic method for elliptic PDEs with random coefficients," *SIAM/ASA J. Uncertain. Quant.*, vol. 1, pp. 452–493, 2013.

[9] M. Cheng, T. Y. Hou, and Z. Zhang, "A dynamically bi-orthogonal method for stochastic partial differential equations I: Derivation and algorithms," *J. Comput. Phys.*, vol. 242, pp. 843–868, 2013.

[10] M. Cheng, T. Y. Hou, and Z. Zhang, "A dynamically bi-orthogonal method for stochastic partial differential equations II: Adaptivity and generalizations," *J. Comput. Phys.*, vol. 242, pp. 753–776, 2013.

[11] T. Duncan, "Probability densities for diffusion processes with applications to nonlinear filtering theory and detection theory," Stanford Electronics Labs, Stanford Univ., Stanford, CA, USA, Tech. Rep. no. TR-7001-4, 1967.

[12] W. Fleming and S. Mitter, "Optimal control and nonlinear filtering for nondegenerate diffusion processes," *Stochastics*, vol. 8, no. 1, pp. 63–77, 1982.

[13] F. Gustafsson *et al.*, "Particle filters for positioning, navigation, and tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 425–437, Feb. 2002.

[14] I. Gyöngy and N. Krylov, "On the splitting-up method and stochastic partial differential equations," *Ann. Probab.*, vol. 31, no. 2, pp. 564–591, 2003.

[15] A Hannachi, I. Jolliffe, and D. Stephenson, "Empirical orthogonal functions and related techniques in atmospheric science: A review," *Int. J. Climatol.*, vol. 27, no. 9, pp. 1119–1152, 2007.

[16] P. Holmes, J. Lumley, and G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[17] T. Iliescu and Z. Wang, "Variational multiscale proper orthogonal decomposition: Convection-dominated convection-diffusion-reaction equations," *Math. Comp.*, vol. 82, no. 283, pp. 1357–1378, 2013.

[18] K. Itô, "Approximation of the Zakai equation for nonlinear filtering," *SIAM J. Control Optim.*, vol. 34, no. 2, pp. 620–634, 1996.

[19] K. Itô, *Diffusion Processes*. Hoboken, NJ, USA: Wiley, 1974.

[20] G. Kallianpur, *Stochastic Filtering Theory*, vol. 13. Berlin, Germany: Springer, 2013.

[21] K. Karhunen, Uber lineare methoden in der Wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fennicae. Ser. A. I. Math. Phys.*, vol. 37, pp. 1–79, 1947.

[22] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*. Berlin, Germany: Springer-Verlag, 1992.

[23] M. Loève, *Probability Theory. Vol. II (GTM. 46)*, 4th ed. Berlin, Germany: Springer-Verlag, 1978.

[24] S. Lototsky, R. Mikulevicius, and B. L. Rozovskii, "Nonlinear filtering revisited: A spectral approach," *SIAM J. Control Optim.*, vol. 35, no. 2, pp. 435–461, 1997.

[25] X. Luo and F. Wang, "Generalized Jacobi spectral method in solving nonlinear filtering problems," in *Proc. 57th IEEE Conf. Decis. Control*, Miami Beach, FL, USA, Dec. 17–19, 2018, pp. 7206–7212.

[26] X. Luo and S. S. T. Yau, "Complete real time solution of the general nonlinear filtering problem without memory," *IEEE Trans. Autom. Control.*, vol. 58, no. 10, pp. 2563–2578, Oct. 2013.

[27] X. Luo and S. S. T. Yau, "Hermite spectral method to 1-D forward Kolmogorov equation and its application to nonlinear filtering problems," *IEEE Trans. Autom. Control.*, vol. 58, no. 10, pp. 2495–2507, Oct. 2013.

[28] X. Luo and S. S. T. Yau, "Hermite spectral method with hyperbolic cross approximations to high-dimensional parabolic PDEs," *SIAM J. Numer. Anal.*, vol. 51, no. 6, pp. 3186–3212, 2013.

[29] R. Mortensen, "Optimal control of continuous-time stochastic systems," Berkeley Electron. Res. Lab, California Univ., Berkeley, CA, USA, Tech. Rep. no. ERL-66-1, 1966.

[30] N. Nagase, "Remarks on nonlinear stochastic partial differential equations: An application of the splitting-up method," *SIAM J. Control Optim.*, vol. 33, no. 6, pp. 1716–1730, 1995.

Authorized licensed use limited to: Tsinghua University. Downloaded on April 16,2020 at 04:05:48 UTC from IEEE Xplore. Restrictions apply.

[31] G. North, T. Bell, R. Cahalan, and F. Moeng, "Sampling errors in the estimation of empirical orthogonal functions," *Monthly Weather Rev.*, vol. 110, no. 7, pp. 699–706, 1982.

[32] E. Pardoux, "Stochastic partial differential equations and filtering of diffusion processes," *Stochastics*, vol. 3, pp. 127–167, 1980.

[33] C. Schwab and R. Todor, "Karhunen–Loève approximation of random fields by generalized fast multipole methods," *J. Comput. Phys.*, vol. 217, pp. 100–122, 2006.

[34] J. Shen and L.-L. Wang, "Sparse spectral approximations of high-dimensional problems based on hyperbolic cross," *SIAM J. Numer. Anal.*, vol. 48, no. 3, pp. 1087–1109, 2010.

[35] L. Sirovich, "Turbulence and the dynamics of coherent structures. I. Coherent structures," *Quart. Appl. Math.*, vol. 45, no. 3, pp. 561–571, 1987.

[36] G. Wang, Z. Wu, and J. Xiong, *An Introduction to Optimal Control of FBSDE With Incomplete Information.* Berlin, Germany: Springer, 2018.

[37] K. Willcox and J. Peraire, "Balanced model reduction via the proper orthogonal decomposition," *AIAA J.*, vol. 40, no. 11, pp. 2323–2330, 2002.

[38] J. Xiong, *An Introduction to Stochastic Filtering Theory*. London, U.K.: Oxford Univ. Press, 2008.

[39] S.-T. Yau and S. S.-T. Yau, "Real time solution of the nonlinear filtering problem without memory II," *SIAM J. Control Optim.*, vol. 47, no. 1, pp. 163–195, 2008.

[40] M. Yueh, W. Lin, and S. T. Yau, "An efficient numerical method for solving high-dimensional nonlinear filtering problems," *Commun. Inf. Syst.*, vol. 14, no. 4, pp. 243–262, 2014.

[41] M. Zakai, "On the optimal filtering of diffusion processes," *Probab. Theory Related Fields*, vol. 11, no. 3, pp. 230–243, 1969.

[42] S. Zhu, L. Dedè, and A. Quarteroni, "Isogeometric analysis and proper orthogonal decomposition for parabolic problems," *Numer. Math.*, vol. 135, no. 2, pp. 333–370, 2017.

Zhongjian Wang received the B.S. degree in mathematics from Tsinghua University, Beijing, China, in 2016. He is currently working toward the Ph.D. degree in applied and computational mathematics from Department of Mathematics, The University of Hong Kong, Hong Kong.

His research interests include numerical methods for stochastic differential equations and stochastic partial differential equations.

Xue Luo (SM'15) received the first Ph.D. degree in mathematics from East China Normal University (ECNU), Shanghai, China, in 2010 and the second Ph.D. degree in applied mathematics from University of Illinois at Chicago (UIC), Chicago, IL, USA, in 2013.

During her study as a Ph.D. candidate in ECNU, she visited the Department of Mathematics, University of Connecticut, Storrs, CT, USA, in 2008–2009 and the Department of Mathematics, Statistics and Computer Science, UIC, in 2009–2010, as a visiting scholar, respectively. After her graduation from UIC, she joined in Beihang University (BUAA), Beijing, China. She is currently an Associated Professor with the School of Mathematical Sciences, BUAA. His research interests include nonlinear filtering theory, numerical analysis of spectral methods, analysis of partial differential equations, sparse grid algorithm, and fluid mechanics.

Stephen S.-T. Yau (F'03) received the Ph.D. degree in mathematics from the State University of New York at Stony Brook, Stony Brook, NY, US, in 1976.

He was a member of Institute of Advanced Study at Princeton in 1976–1977 and 1981–1982, and a Benjamin Pierce Assistant Professor with the 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.

Dr. Yau was awarded Sloan Fellowship in 1980, Guggenheim Fellowship in 2000, IEEE Fellow Award in 2003, and AMS Fellow Award in 2013. In 2005, he was entitled the UIC Distinguished Professor. During 2005–2011, he became a Joint-Professor of Department of Electrical and Computer Engineering and MSCS, UIC. After his retirement in 2012, he joined Tsinghua University, Beijing, China, where he is a Full-Time Professor with the Department of Mathematical Sciences. His research interests include nonlinear filtering, bioinformatics, complex algebraic geometry, CR geometry, and singularities theory.

Dr. Yau is the Managing Editor and founder of *Journal of Algebraic Geometry* from 1991, and since 2000, the Editors-in-Chief and founder of *Communications in Information and Systems*. He was the General Chairman of IEEE International Conference on Control and Information, which was held in the Chinese University of Hong Kong in 1995.

Zhiwen Zhang received the B.S. and Ph.D. degree in mathematics from Tsinghua University, Beijing, China, in 2006 and 2011, respectively.

As a Ph.D. candidate with Tsinghua University, he visited the University of Wisconsin-Madison as a visiting student in 2008–2009. After his graduation, he was a Postdoctoral Scholar with the California Institute of Technology, Pasadena, CA, USA, from 2011 to 2015. He joined the Department of Mathematics, The University of Hong Kong as an Assistant Professor since 2015. His research interests include scientific computation. Research topics include uncertainty quantification (UQ), numerical methods for partial differential equations (PDEs) arising from quantum chemistry, wave propagation, multiscale porous media, nonlinear filtering, data assimilation, and stochastic fluid dynamics.

Authorized licensed use limited to: Tsinghua University. Downloaded on April 16,2020 at 04:05:48 UTC from IEEE Xplore. Restrictions apply.