Automatica 156 (2023) 111176

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Solving nonlinear filtering problems with correlated noise based on Hermite–Galerkin spectral method☆

Zeju Sun [a], Stephen Shing-Toung Yau [a,b,*]

[a] Department of Mathematical Sciences, Tsinghua University, Beijing, 100084, PR China
[b] Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, Beijing, 101408, PR China

ARTICLE INFO

Article history:
Received 12 August 2021
Received in revised form 5 January 2023
Accepted 10 June 2023
Available online 16 July 2023

Keywords:
Nonlinear filtering (NLF)
Hermite–Galerkin spectral method (HGSM)
Correlated noise
Stochastic partial differential equation (SPDE)
Convergence analysis

ABSTRACT

Nonlinear filtering problem has important applications in various fields. One of the core issues in nonlinear filtering is to numerically solve the Duncan–Mortensen–Zakai (DMZ) equation, which is an evolution equation satisfied by the unnormalized conditional density of state process under noisy observations, in a real-time and memoryless manner. When the noise in observations is correlated to the state process, the DMZ equation we need to deal with is a second-order stochastic partial differential equation. In this paper, we will propose an algorithm to solve the DMZ equation in this case, based on Hermite–Galerkin spectral method. According to this method, the DMZ equation is converted into a system of linear stochastic differential equations generated by the observation process. The effects of different discretization schemes on this stochastic differential system will also be discussed. Moreover, rigorous convergence analysis of the algorithm is given under mild conditions. Numerical results show that the method proposed in this paper can provide an instantaneous and accurate estimation to the state process of the system.

1. Introduction

Nonlinear filtering (NLF) problem originated from the field of signal processing and has important applications in military, industrial and commercial areas. The fundamental problem in NLF is to give an instantaneous and accurate estimation to the state process, which is, in many cases, a diffusion process described by a system of stochastic differential equations (SDE), based on noisy observations (Bain & Crisan, 2009; Luo & Yau, 2013b).

One approach to solving NLF is based on particle filter or sequential Monte Carlo method, which is referred to Arulampalam, Maskell, Gordon, and Clapp (2002) and Liu and Chen (1998) and references therein. The principle of the particle filter is basically the Law of Large Numbers and therefore, the accuracy of the particle filter depends on the number of the particles exploited. With more particles engaged in, a more accurate estimation can be obtained, but in the meanwhile, the computation and storage cost will also increase. In 1960s, Duncan (1967), Mortensen (1996) and Zakai (1969) independently derived the so-called DMZ equation which is the evolution equation satisfied by the unnormalized probability density of state process conditioned on observations. A more general form of DMZ equations can be found in recent literature such as (Ceci & Colaneri, 2012). DMZ equations are stochastic partial differential equations (SPDE) whose solutions cannot be expressed in a closed form in general. Therefore, one of the core issues in NLF is numerically solving the DMZ equation in a real-time and memoryless manner.

The stochastic parts in an NLF system, as is shown later in (1), are the Wiener process generating the SDE satisfied by the state process, and the noise in the observations. If these two stochastic terms are independent (which is referred to as NLF with independent noise), then by introducing another scalar, the DMZ equation can be converted into a deterministic partial differential equation with stochastic coefficients, which is called the robust DMZ equation (Bain & Crisan, 2009; Baras, Blankenship, & Hopkins, 1983; Davis, 1980). Yau and Yau apply frozen coefficient method to solving robust DMZ equation in which the main computation cost comes from solving a Kolmogorov forward equation and can be computed off-line (Yau & Yau, 2000, 2008). Later on, based on this method, Yau and his collaborators develop a real-time nonlinear filtering algorithm, called Yau–Yau algorithm (Dong, Luo, & Yau, 2021; Luo & Yau, 2013a, 2013b; Wang, Luo, Yau, & Zhang, 2020).

If the Wiener process generating the state process and the noise in the observations are correlated (which is referred to as NLF with correlated noise), a robust DMZ equation cannot

☆ This work is supported by Tsinghua University Education Foundation fund (042202008). The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Tianshi Chen under the direction of Editor Alessandro Chiuso.

* Corresponding author at: Department of Mathematical Sciences, Tsinghua University, Beijing, 100084, PR China.
E-mail addresses: szj20@mails.tsinghua.edu.cn (Z. Sun), yau@uic.edu (S.S.-T. Yau).

https://doi.org/10.1016/j.automatica.2023.111176
0005-1098/

be obtained easily and we need to deal with the original DMZ equation directly. In 1991, Florchinger and Le Gland (1991) proposed a splitting-up method to solve DMZ equation, where the DMZ equation is decomposed into a deterministic PDE and a backward SDE. Other numerical methods of solving SPDEs can also be applied to solving DMZ equations (Ahmed & Radaideh, 1997; Budhiraja, Chen, & Lee, 2007; Cheng, Hou, & Zhang, 2013; Frey, Schmidt, & Xu, 2013).

Among all the numerical methods of solving PDEs and SPDEs, spectral methods construct a numerical solution by projecting the exact solution onto a finite dimensional function space spanned by an orthogonal system in square-integrable space. In comparison with other methods such as finite element, finite difference, which are more capable of describing the local structure of the solution, spectral methods can provide superior accuracy globally (Shen, Tang, & Wang, 2011) and are more suitable for the NLF problem, because the conditional expectation itself reflects global properties of the density function. The introduction of spectral methods in solving PDEs can date back to the 1970s, in the field of computational fluid dynamics (Gottlieb & Orszag, 1977). For a thorough comprehension of spectral methods, readers can refer to the monographs such as (Gottlieb & Orszag, 1977; Shen et al., 2011).

Spectral methods have also been proposed to solve PDEs appearing in NLF. As for the problem with independent noise, the Kolmogorov forward equations in the off-line part of Yau–Yau algorithm can be solved numerically by spectral method (Dong et al., 2021; Luo & Yau, 2013b); Frey et al. (2013) also use spectral methods to solve the DMZ equation in NLF with point process observations; as for the problem with correlated noise, Lototsky and collaborators applied chaos expansion to the solution of DMZ equations, separated the time variable, spatial variable and random variable completely, and proposed a numerical algorithm to solve SPDEs based on spectral methods (Lototsky, 2003; Lototsky, Mikulevicius, & Rozovskii, 1997).

In this paper, we will apply Hermite–Galerkin spectral method (HGSM) to solve the DMZ equation of NLF with correlated noise, in which the functions of the orthogonal system are chosen to be Hermite functions with suitable scaling and translating factors, as is studied in relevant researches (Funaro & Kavian, 1991; Luo & Yau, 2013c; Xiang & Wang, 2010). Hermite functions, as defined in (4), are supported on the whole space while vanishing exponentially at infinity. For NLF with mild coefficients, the conditional density function shares similar properties with Hermite functions: it also concentrates on a bounded domain and vanishes rapidly at infinity. Therefore, it is efficient to use Hermite functions to approximate the conditional density function in this case and we will also provide a rigorous convergence analysis of this numerical method.

Since the DMZ equations here are SPDEs, when we project the solution on the finite dimensional space spanned by Hermite functions, the linear combination coefficients, which are also referred to as Hermite–Fourier coefficients, satisfy another system of SDEs generated by the noisy observations. In the works of Lototsky and his collaborators (Lototsky, 2003; Lototsky et al., 1997), the time variable and the random variable are further separated based on the chaos expansion techniques and the problem of numerically solving the SPDE is converted into the problem of numerically solving a system of deterministic partial differential equations or a system of ordinary differential equations. In this paper, instead of further separating the time variable and random variable of the SDE system, we will apply more efficient time discretization schemes, such as Milstein scheme (Kloeden & Platen, 1992) and curved schemes (Armstrong & King, 2022), to solve the SDE system directly, and discuss the effects of these schemes on NLF problems.

When the spectral methods are applied in high dimensional problems, the number of the basis functions needed will increase rapidly with respect to the dimension and cause a great increment in the overall computational cost. This phenomenon is often referred to as *curse of dimensionality*. In this paper, we also introduce scaling and translating factors to the standard Hermite functions as in Luo and Yau (2013c), in order to improve the approximation efficiency. With proper scaling and translating factors, the HGSM proposed in this paper can successfully solve two-dimensional NLF problems. Together with other sparse mesh techniques, the HGSM also has the potential of being generalized to problems in medium high dimensions.

The structure of this paper is as follows. In Section 2, we introduce the basic concepts of NLF and numerical methods, such as DMZ equations and Hermite polynomials, in detail. In Section 3, we propose the HGSM to solve the DMZ equation introduced in Section 2 and discuss the convergence result of this method under mild conditions. In Section 4, we focus on numerical methods of solving SDEs. Some numerical results are provided in Section 5 and Section 6 is the conclusion.

2. Preliminaries

2.1. Nonlinear filtering problem with correlated noise

Consider the following nonlinear filtering system with correlated noise,

$$\begin{cases} dX_t = b(X_t)dt + \sigma(X_t)dW_t + \rho(X_t)dV_t \\ dY_t = h(X_t)dt + dV_t \end{cases}, \tag{1}$$

where $\{X_t : 0 \le t \le T\}$ and $\{Y_t : 0 \le t \le T\}$ are $\mathbb{R}^d$-valued stochastic processes, $\{W_t : 0 \le t \le T\}$ and $\{V_t : 0 \le t \le T\}$ are independent standard Brownian motions of appropriate dimension, and $T > 0$ is a fixed terminal time. The coefficients b, σ, ρ, h are globally Lipschitz continuous, vector or matrix valued functions defined on $\mathbb{R}^d$.

For the simplicity of notations, we define $a(x) = \sigma(x)\sigma(x)^\top$ and $c(x) = \rho(x)\rho(x)^\top$.

The unnormalized density function p_t satisfies the following DMZ equation (Bain & Crisan, 2009; Florchinger & Le Gland, 1991)

$$dp_t = L_0 p_t dt + \sum_{k=1}^{d} L_k p_t dY_t^k, \tag{2}$$

where

$$L_0(\star) = \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j}\left[\left(a^{ij} + c^{ij}\right)\star\right] - \sum_{i=1}^{d} \frac{\partial}{\partial x_i}\left(b^i \star\right),$$

and $L_i(\star) = h_i \star + \sum_{j=1}^{d} \frac{\partial}{\partial x_j}\left(\rho^{ij}\star\right)$.

If $\rho(x) \equiv 0$, then the filtering system (1) reduces to the noise-independent case considered in abundant literature and the DMZ equation is indeed in the standard form proposed in Duncan (1967), Mortensen (1996) and Zakai (1969).

In the noise-independent case, the DMZ equation, an SPDE, can be converted into a deterministic PDE with stochastic coefficients through invertible exponential transformation (Rozovsky, 1972). For the noise-correlated case we consider here, however, due to the existence of partial derivative with respect to the spatial variable, x, in the operators L_k, such kind of explicit exponential transformations do not exist. Therefore, we need to think of numerical methods to solve the SPDE (2) directly.

2.2. Hermite functions

Hermite functions form an orthogonal basis of $L^2(\mathbb{R}^d)$. For $d = 1$, univariate Hermite functions are constructed from Hermite

polynomials $\{h_n(x)\}_{n=0}^{\infty}$, which are orthogonal polynomials with respect to weight function $w(x) = e^{-x^2}$ and satisfy the following iterative formula:

$$\begin{aligned} &h_0(x) \equiv 1,\ h_1(x) = 2x, \\ &h_{n+1}(x) = 2xh_n(x) - 2nh_{n-1}(x),\ n \geq 1. \end{aligned} \tag{3}$$

The univariate Hermite functions $\left\{e^{-\frac{1}{2}x^2}h_n(x)\right\}_{n=0}^{\infty}$ can be used to approximate functions which concentrate on the neighborhood of the origin. In order to approximate more general functions with Hermite functions, we need to introduce scaling and translating factors α and β.

For $\alpha > 0$ and $\beta \in \mathbb{R}$ and $n \in \mathbb{N}$, the generalized univariate Hermite function $\mathcal{H}_n^{\alpha,\beta}$ is defined to be

$$\mathcal{H}_n^{\alpha,\beta}(x) = \left(\frac{\alpha}{2^n n!\sqrt{\pi}}\right)^{\frac{1}{2}} h_n(\alpha(x-\beta))e^{-\frac{1}{2}\alpha^2(x-\beta)^2}.$$

where $\left(\frac{\alpha}{2^n n!\sqrt{\pi}}\right)^{\frac{1}{2}}$ is the normalizing parameter, so that

$$\|\mathcal{H}_n^{\alpha,\beta}\|_{L^2(\mathbb{R})}^2 \triangleq \int_{\mathbb{R}} |\mathcal{H}_n^{\alpha,\beta}(x)|^2 dx = 1.$$

For fixed factors $\alpha > 0$ and $\beta \in \mathbb{R}$, the generalized univariate Hermite functions $\{\mathcal{H}_n^{\alpha,\beta}(x)\}_{n=0}^{\infty}$ defined above actually form an orthonormal basis of $L^2(\mathbb{R})$. This basis has the capability of approximating functions concentrating on the neighborhood of the translating factor β.

For the case when dimension $d > 1$, the Hermite functions can be constructed by tensor products of univariate ones. Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)$ be multi-factors, with $\alpha_i > 0$, $\beta_i \in \mathbb{R}$, $i = 1, \ldots, d$, and $\mathbf{n} = (n_1, \ldots, n_d)$ be a multi-index, with $n_i \in \mathbb{N}$, $i = 1, \ldots, d$. Then the d-dimensional Hermite function, $\mathbf{H}_{\mathbf{n}}^{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbf{x})$, is defined as

$$\mathbf{H}_{\mathbf{n}}^{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbf{x}) = \prod_{i=1}^{d} \mathcal{H}_{n_i}^{\alpha_i,\beta_i}(x_i). \tag{4}$$

According to the orthonormality of univariate Hermite functions, for fixed multi-factors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, the collection $\{\mathbf{H}_{\mathbf{n}}^{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbf{x}) : \mathbf{n} \in \mathbb{N}^d\}$ also forms an orthonormal basis of the Hilbert space $L^2(\mathbb{R}^d)$, which means that

$$\int_{\mathbb{R}^d} \mathbf{H}_{\mathbf{n}}^{\boldsymbol{\alpha},\boldsymbol{\beta}}(x)\mathbf{H}_{\mathbf{m}}^{\boldsymbol{\alpha},\boldsymbol{\beta}}(x)dx = \begin{cases} 1, & \mathbf{n} = \mathbf{m} \\ 0, & \text{otherwise} \end{cases},$$

and for any $u \in L^2(\mathbb{R}^d)$, we have

$$u(x) = \sum_{\mathbf{n}\in\mathbb{N}^d} \langle u, \mathbf{H}_{\mathbf{n}}^{\boldsymbol{\alpha},\boldsymbol{\beta}}\rangle \mathbf{H}_{\mathbf{n}}^{\boldsymbol{\alpha},\boldsymbol{\beta}}(x), \tag{5}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product in $L^2(\mathbb{R}^d)$ and the series on the right-hand side converge in L^2 sense.

If we do the summation only on a finite subset of $\mathbb{N}^d$ in (5), then the truncated series can be regarded as a good approximation to the original function $u(x)$. For example, let $\Omega_N = \{\mathbf{n} \in \mathbb{N}^d : 0 \leq n_i \leq N, 1 \leq i \leq d\}$ be the finite multi-index set, in which each component of a multi-index is no larger than N, then for N large enough, the projection

$$P_N u(x) = \sum_{\mathbf{n}\in\Omega_N} \langle u, \mathbf{H}_{\mathbf{n}}^{\boldsymbol{\alpha},\boldsymbol{\beta}}\rangle \mathbf{H}_{\mathbf{n}}^{\boldsymbol{\alpha},\boldsymbol{\beta}}(x) \tag{6}$$

is close to $u(x)$. Given additional regularity of $u(x)$, we can also give the rate of convergence, as is stated later in Theorem 2 in Section 3.

3. Hermite-Galerkin spectral method

The core procedure in HGSM is to obtain the Hermite–Galerkin approximation to the variational problem reformulated from the DMZ Eq. (2). To this end, we need first to write the differential operator L_0 in divergence form, which can be done if the coefficients are smooth enough.

$$L_0 u = \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial}{\partial x_i}\left[A^{ij}\frac{\partial u}{\partial x_j}\right] + \sum_{i=1}^{d} B^i \frac{\partial u}{\partial x_j} + Cu,$$

where $A^{ij} = \left(a^{ij} + c^{ij}\right)$, $B^i = \frac{1}{2}\sum_{j=1}^{d}\frac{\partial A^{ij}}{\partial x_j} - b^i$, $C = \frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2 A^{ij}}{\partial x_i \partial x_j} - \sum_{i=1}^{d}\frac{\partial b^i}{\partial x_i}$.

From now on, we use ω to represent randomness. For fixed $T > 0$, the variational problem will be considered in the Hilbert space $L_\omega^2([0,T], X)$, which is the space of all stochastic processes $u(t,\omega)$ in a separable Banach space X such that

$$E\int_0^T \|u(t,\omega)\|_X^2 dt < \infty. \tag{7}$$

Since both the Hermite functions and the conditional probability density functions vanish rapidly at infinity, in order to avoid tedious discussion on the properties of functions at infinity, we would like to focus on the behavior of the solution to DMZ equation in a bounded domain. Let $U \subset \mathbb{R}^d$ be a bounded open subset. $H^1(U)$ is defined to be the Sobolev space of square-integrable functions with partial derivatives also belongs to $L^2(U)$, i.e. $H^1(U) = \{u \in L^2(U) : \frac{\partial u}{\partial x_i} \in L^2(U),\ 1 \leq i \leq d\}$, equipped with the H^1-norm:

$$\|u\|_{H^1(U)} = \left(\|u\|_{L^2(U)}^2 + \sum_{i=1}^{d}\left\|\frac{\partial u}{\partial x_i}\right\|_{L^2(U)}^2\right)^{\frac{1}{2}},$$

where $\frac{\partial u}{\partial x_i}$, $1 \leq i \leq d$ are defined in weak sense. Moreover, $H_0^1(U)$ is defined to be the closure of the set of smooth functions with compact support, $C_0^\infty(U)$, under H^1-norm.

The generalized solution to the variational problem reformulated from the DMZ equation will be defined in the Hilbert space $L_\omega^2([0,T], H_0^1(U))$ as follows.

Definition 1. Let $U \subset \mathbb{R}^d$ be a bounded open subset. A function $p \in L_\omega^2([0,T], H_0^1(U))$ is called a generalized solution of Eq. (2) if, for every $y \in C_0^\infty(U)$, the following equality holds almost everywhere for $t \in [0,T]$ with respect to Lebesgue measure and almost surely with respect to the probability measure,

$$\begin{aligned} \langle p(t), y\rangle =& \langle \varphi, y\rangle + \int_0^t \mathscr{A}(p(s), y)ds \\ &+ \sum_{k=1}^{d}\int_0^t \langle L_k p(s), y\rangle dY_s^k, \end{aligned} \tag{8}$$

where

$$\begin{aligned} &\mathscr{A}(u,v) \\ &= -\frac{1}{2}\sum_{i,j=1}^{d}\left\langle A^{ij}\frac{\partial u}{\partial x_i}, \frac{\partial v}{\partial x_j}\right\rangle + \sum_{i=1}^{d}\left\langle B^i\frac{\partial u}{\partial x_i} + Cu, v\right\rangle \\ &= \int_U -\frac{1}{2}\sum_{i,j=1}^{d} A^{ij}\frac{\partial u}{\partial x_i}\frac{\partial v}{\partial x_j} + \left(\sum_{i=1}^{d} B^i\frac{\partial u}{\partial x_i} + Cu\right)v dx. \end{aligned}$$

Hermite–Galerkin approximation of the solution to the variational problem (8) is obtained by replacing the test function y to be a function in a finite dimensional function space spanned by Hermite functions.

Let us denote by $S^N_{\alpha,\beta}$ the finite dimensional space spanned by $\{\mathbf{H}^{\alpha,\beta}_{\mathbf{n}} : \mathbf{n} \in \Omega_N\}$,i.e. $S^N_{\alpha,\beta} = \text{Span}\{\mathbf{H}^{\alpha,\beta}_{\mathbf{n}} : \mathbf{n} \in \Omega_N\}$.

We called a stochastic process $p_N \in L^2_\omega([0,T], S^N_{\alpha,\beta})$ a Galerkin approximation of the solution to the variational problem (8), if it satisfies

$$\begin{aligned}\langle p_N(t), \mathbf{H}^{\alpha,\beta}_{\mathbf{n}}\rangle =& \langle \varphi, \mathbf{H}^{\alpha,\beta}_{\mathbf{n}}\rangle + \int_0^t \mathscr{A}(p_N(s), \mathbf{H}^{\alpha,\beta}_{\mathbf{n}})ds \\ &+ \sum_{k=1}^d \int_0^t \langle L_k p_N(s), \mathbf{H}^{\alpha,\beta}_{\mathbf{n}}\rangle dY^k_s,\end{aligned} \tag{9}$$

for all $\mathbf{n} \in \Omega_N$, and for $t \in [0,T]$ almost everywhere and almost surely.

Since for almost every $t \in [0,T]$, the Hermite–Galerkin approximation process $p_N(t,\cdot) \in S^N_{\alpha,\beta}$, there exist coefficients $\Psi(t) = (\psi_{\boldsymbol{n}}(t))_{\boldsymbol{n}\in\Omega_N}$, such that

$$p_N(t) = \sum_{\boldsymbol{n}\in\Omega_N} \psi_{\boldsymbol{n}}(t)\boldsymbol{H}^{\alpha,\beta}_{\boldsymbol{n}}(x). \tag{10}$$

The coefficients Ψ_t are often called *Hermite–Fourier coefficients*.

Taking Eq. (10) into (9) for each $\boldsymbol{n} \in \Omega_N$, we can obtain a system of SDEs for $\Psi(t)$, based on the orthonormality of Hermite functions.

$$d\Psi(t) = P\Psi(t)dt + \sum_{k=1}^d Q^{(k)}\Psi(t)dY^k(t), \tag{11}$$

where $P = (P_{\boldsymbol{n}_1,\boldsymbol{n}_2})_{\boldsymbol{n}_1,\boldsymbol{n}_2\in\Omega_N}$, $(Q^{(k)}_{\boldsymbol{n}_1,\boldsymbol{n}_2})_{\boldsymbol{n}_1,\boldsymbol{n}_2\in\Omega_N}$ are constant matrices with entries $P_{\boldsymbol{n}_1,\boldsymbol{n}_2} = \mathscr{A}(\boldsymbol{H}^{\alpha,\beta}_{\boldsymbol{n}_2}, \boldsymbol{H}^{\alpha,\beta}_{\boldsymbol{n}_1})$, $Q^{(k)}_{\boldsymbol{n}_1,\boldsymbol{n}_2} = \langle L_k\boldsymbol{H}^{\alpha,\beta}_{\boldsymbol{n}_2}, \boldsymbol{H}^{\alpha,\beta}_{\boldsymbol{n}_1}\rangle$ can be computed off-line.

By solving the stochastic differential system (11), we can obtain the coefficients $\Psi(t)$ and then compute the Hermite–Galerkin approximation $p_N(t,x)$ at a given time $t \in [0,T]$, which fulfills the core procedure in HGSM.

We postpone the discussion of methods for SDE systems to Section 4 and focus on the convergence analysis of HGSM here.

For the convenience of the discussion of convergence analysis, we need introduce the Sobolev-type spaces related to Hermite functions.

For the same multi-factors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ as in the Hermite functions (4), consider the operator

$$\boldsymbol{\mathcal{D}}^{\boldsymbol{k}}_{\boldsymbol{x}} = \prod_{j=1}^d \mathcal{D}^{k_j}_{x_j}, \tag{12}$$

where $\mathcal{D}_{x_j} = \frac{\partial}{\partial x_j} + \alpha_j^2(x_j - \beta_j)$, $1 \le j \le d$.

The connection between the operator $\boldsymbol{\mathcal{D}}^{\boldsymbol{k}}_{\boldsymbol{x}}$ and the Hermite functions $\boldsymbol{H}^{\alpha,\beta}_{\boldsymbol{n}}$ is that $\boldsymbol{\mathcal{D}}^{\boldsymbol{k}}_{\boldsymbol{x}}\boldsymbol{H}^{\alpha,\beta}_{\boldsymbol{n}} = \sqrt{\mu_{\boldsymbol{n},\boldsymbol{k}}}\boldsymbol{H}^{\alpha,\beta}_{\boldsymbol{n}-\boldsymbol{k}}$, with $\mu_{\boldsymbol{n},\boldsymbol{k}} = \prod_{j=1}^d \mu_{n_j,k_j}$, and

$$\mu_{n_j,k_j} = \begin{cases} \dfrac{2^{k_j}\alpha_j^{2k_j} n_j!}{(n_j-k_j)!}, & n_j \ge k_j > 0; \\ 1 \quad n_j \ge k_j,\ k_j = 0; \\ 0 \quad k_j > n_j \ge 0. \end{cases} \tag{13}$$

For further relationships between the operators $\boldsymbol{\mathcal{D}}^{\boldsymbol{k}}_{\boldsymbol{x}}$ and the Hermite functions, readers may refer to Luo and Yau (2013c).

Sobolev-type spaces $\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)$, $r \in \mathbb{N}$, are defined to be

$$\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U) = \left\{u \in L^2(U) : \boldsymbol{\mathcal{D}}^{\boldsymbol{k}}_{\boldsymbol{x}} \in L^2(U),\ \forall |\boldsymbol{k}| \le r\right\},$$

where $|\boldsymbol{k}| = \sum_{i=1}^d k_i$.

Similar to the standard Sobolev spaces, for $u \in \mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)$, the norm and semi-norm in $\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)$ are defined as follows:

$$\|u\|_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)} = \left(\sum_{|\boldsymbol{k}|\le r} \|\boldsymbol{\mathcal{D}}^{\boldsymbol{k}}_{\boldsymbol{x}}u\|^2_{L^2(U)}\right)^{\frac{1}{2}}$$

$$,|u|_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)} = \left(\sum_{j=1}^d \|\mathcal{D}^r_{x_j}u\|^2_{L^2(U)}\right)^{\frac{1}{2}},$$

and we use $\overline{\mathcal{W}}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)$ to denote the closure of $C^\infty_0(U)$ under the $\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}$-norm.

Functions in the Sobolev-type spaces $u \in \overline{\mathcal{W}}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)$ can be approximated well by its projection on $S^N_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ provided that N is large enough. In fact, a more general result with respect to $U = \mathbb{R}^d$ also holds, as is stated in Theorem 2.

Theorem 2 (*Luo & Yau, 2013c*)**.** *For $r \in \mathbb{N}$, assume that the coefficient $N > 2(r-1)$ and the multi-factor $\boldsymbol{\alpha}$ satisfies*

$$\max_{1\le j\le d} |\alpha_j| / \min_{1\le j\le d} |\alpha_j| \le c_0,$$

for some $c_0 > 0$. Given $u \in \mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbb{R}^d)$, we have for any $0 \le l \le r$,

$$\begin{aligned}&\|P_N u - u\|_{\mathcal{W}^l_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbb{R}^d)} \\ &\quad\le C_{d,l}\sqrt{2c_0^{2r}+1}(2|\boldsymbol{\alpha}|^2_\infty)^{\frac{l-r}{2}} N^{\frac{l-r}{2}} |u|_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbb{R}^d)}\end{aligned} \tag{14}$$

where $|\boldsymbol{\alpha}|_\infty \triangleq \max_{1\le j\le d}|\alpha_j|$, and $C_{d,l}$ is some constant depending on d and l.

Remark 3. For practical use, the multi-factor $\boldsymbol{\alpha}$ can be chosen such that each component of $\boldsymbol{\alpha}$ are closed to each other. In that case, c_0 can be chosen to be very closed to 1, and the coefficients $\sqrt{2c_0^{2r}+1}$ in the right-hand side of (14) is not very large, even for a relatively large r. A detailed proof of Theorem 2 is demonstrated in Appendix A.

From now on, we further assume that $a(x)$ is uniformly elliptic, that is, there exists a constant $\delta > 0$ such that $\sum_{i,j=1}^d a^{ij}(x)\xi_i\xi_j \ge \delta|\xi|^2$ holds for all $\xi = (\xi_1,\dots,\xi_d) \in \mathbb{R}^d$ and $x \in U$.

Remark 4. Since U is a bounded open set, $\overline{U}$ is compact. Therefore, the uniformly elliptic condition can be replaced by: there exists a continuous function $\delta(x)$ defined on $\overline{U}$ such that $\delta(x) > 0$, $\forall x \in \overline{U}$ and $\sum_{i,j=1}^d a^{ij}(x)\xi_i\xi_j \ge \delta(x)|\xi|^2$ holds for all $x \in U$ and $\xi = (\xi_1,\dots,\xi_d) \in \mathbb{R}^d$

These assumptions guarantee the super-parabolicity of (2) as an SPDE mentioned in Rozovsky and Lototsky (2018) and thus, the Cauchy initial-value problem has a unique generalized solution. Moreover, the following theorem shows that the generalized solution also has the additional regularity corresponding to mild coefficients.

Theorem 5. *Let $U \subset \mathbb{R}^d$ be a bounded open subset. Suppose that the uniformly elliptic condition holds and for some positive integer r, the coefficients a, b, c, h, ρ, σ and their derivatives up to order r are uniformly bounded by a constant K; the initial value $\varphi \in \overline{\mathcal{W}}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)$.*

Then the generalized solution p of the Cauchy problem (2) belongs to $L^\infty_\omega([0,T], \overline{\mathcal{W}}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)) \cap L^2_\omega([0,T], \overline{\mathcal{W}}^{r+1}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U))$, and there exists a positive number M depending only on the coefficients and T, d, m, r, such that

$$\begin{aligned}E\sup_{t\le T}\|p(t)\|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)} + E\int_0^T \|p(t)\|^2_{\mathcal{W}^{r+1}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)}dt \\ \le M\|\varphi\|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)}.\end{aligned} \tag{15}$$

Proof. The bilinear form $\mathscr{A}(\cdot,\cdot)$ can be rewritten using the operators $\mathcal{D}_{x_i}$, $1 \le i \le d$, that is, for all $u, v \in \mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)$,

$$\begin{aligned}\mathscr{A}(u,v) =& -\frac{1}{2}\sum_{i,j=1}^{d}\langle A^{ij}\mathcal{D}_{x_i}u, \mathcal{D}_{x_j}v\rangle \\ &+ \sum_{i=1}^{d}\langle B^i\mathcal{D}_{x_i}u, v\rangle + \langle \widetilde{C}u, v\rangle,\end{aligned} \tag{16}$$

where

$$\begin{aligned}\widetilde{C}(x) =& C(x) - \frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial}{\partial x_j}\left(A^{ij}(x)\alpha_i^2(x_i-\beta_i)\right) \\ &+ \frac{1}{2}\sum_{i,j=1}^{d}A^{ij}(x)\alpha_i^2\alpha_j^2(x_i-\beta_i)(x_j-\beta_j) \\ &- \sum_{i=1}^{d}B^i(x)\alpha_i^2(x_i-\beta_i).\end{aligned}$$

Similarly, each L_i can also be rewritten using $\mathcal{D}_{x_j}$, $1 \le j \le d$.

$$L_iu = \sum_{j=1}^{d}\rho^{ij}(x)\mathcal{D}_{x_j}u + J_i(x)u, \tag{17}$$

where

$$J_i(x) = h_i(x) + \sum_{j=1}^{d}\left(\frac{\partial\rho^{ij}}{\partial x_j}(x) - \rho^{ij}(x)\alpha_j^2(x_j-\beta_j)\right).$$

Define the dual operator of $\mathcal{D}_{x_j}$ to be $\overline{D}_{x_j} = \frac{\partial}{\partial x_j} - \alpha_j^2(x_j-\beta_j)$, $1 \le j \le d$, and similarly $\overline{\mathcal{D}}^{\boldsymbol{k}}_{\boldsymbol{x}} = \prod_{j=1}^{d}\overline{\mathcal{D}}^{k_j}_{x_j}$.

The duality of $\mathcal{D}^{\boldsymbol{k}}_{\boldsymbol{x}}$ and $\overline{\mathcal{D}}^{\boldsymbol{k}}_{\boldsymbol{x}}$ is implied by $\langle u, \mathcal{D}^{\boldsymbol{k}}_{\boldsymbol{x}}v\rangle = (-1)^{|\boldsymbol{k}|}\langle\overline{\mathcal{D}}^{\boldsymbol{k}}_{\boldsymbol{x}}u, v\rangle$, for all $u, v \in \overline{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)$, $0 \le |\boldsymbol{k}| \le r$.

Since $\overline{\mathcal{W}}^{r+1}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)$ is the closure of $C_0^\infty(U)$ under the $\mathcal{W}^{r+1}_{\boldsymbol{\alpha},\boldsymbol{\beta}}$-norm, by direct computation and Cauchy–Schwarz inequality, we can obtain the coercivity condition, that is, there exist $\delta' > 0$ and $M' > 0$, such that $\forall u \in \overline{\mathcal{W}}^{r+1}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)$

$$\begin{aligned}&-\frac{1}{2}\sum_{i,j=1}^{d}\langle A^{ij}\mathcal{D}_{x_i}u, \mathcal{D}_{x_j}u\rangle_r + \sum_{i=1}^{d}\langle B^i\mathcal{D}_{x_i}u, u\rangle_r \\ &+ \langle\widetilde{C}u, u\rangle_r + \frac{1}{2}\sum_{i=1}^{d}\|L_iu\|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)} \\ &\le -\delta'\|u\|^2_{\mathcal{W}^{r+1}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)} + M'\|u\|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)},\end{aligned} \tag{18}$$

where $\langle\cdot,\cdot\rangle_r$ denotes the standard inner product on $\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)$.

With the coercivity condition (18), Theorem 5 holds according to the standard proof in the theory of coercive stochastic evolution systems, which can be found in monographs such as (Rozovsky & Lototsky, 2018). For readers' convenience, we put the calculation processes as well as the details of this proof in Appendix B.

Under mild conditions when the generalized solution, $p(t,x)$, is smooth enough, we can give an estimation of the difference between $p(t,x)$ and its Galerkin approximation, as is stated in the following theorem.

Theorem 6. *Let $U \subset \mathbb{R}^d$ be an arbitrary bounded open subset. Suppose that the conditions about the coefficients of the NLF system in Theorem 5 are satisfied and the generalized solution of Eq. (2) $p \in L^2_\omega([0,T], \overline{\mathcal{W}}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U))$ for some $r > 1$, then*

$$\begin{aligned}&\max_{0\le t\le T}E\|1_U(p_N(t)-p(t))\|^2 \\ &\le KN^{1-r}\max_{0\le t\le T}E|p(t)|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)},\end{aligned} \tag{19}$$

where 1_U is the indicator function on U, $\|\cdot\|$ denotes the norm in $L^2(\mathbb{R}^d)$ and K is a constant depending on T, r, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and the coefficients in the NLF system.

Proof. We further define the value of the generalized solution p and the coefficients to be zero outside the open subset U. Let $P_N : \mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbb{R}^d) \to S^N_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ be the projection operator defined in (6) and denote $q_N(t) = P_Np(t)$, then we have $\langle p(t) - q_N(t), \mathbf{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\mathbf{n}}\rangle = 0, \quad \forall\mathbf{n} \in \Omega_N$.

Therefore, by the definition of generalized solution,

$$\begin{aligned}&\langle q_N(t), \mathbf{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\mathbf{n}}\rangle = \langle p(t), \mathbf{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\mathbf{n}}\rangle = \langle\varphi, \mathbf{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\mathbf{n}}\rangle + \\ &\int_0^t\mathscr{A}(p(s), \mathbf{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\mathbf{n}})ds + \sum_{k=1}^{d}\int_0^t\langle L_kp(s), \mathbf{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\mathbf{n}}\rangle dY^k_s.\end{aligned} \tag{20}$$

Combining (9) and (20), we have

$$\begin{aligned}\langle p_N(t) - q_N(t), \mathbf{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\mathbf{n}}\rangle =& \int_0^t\mathscr{A}(p_N(s) - p(s), \mathbf{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\mathbf{n}})ds \\ &+ \sum_{k=1}^{d}\int_0^t\langle L_k(p_N(s) - p(s)), \mathbf{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\mathbf{n}}\rangle dY^k_s.\end{aligned}$$

Since $p_N(t) - q_N(t) \in S^N_{\boldsymbol{\alpha},\boldsymbol{\beta}}$,

$$\|p_N(t) - q_N(t)\|^2 = \sum_{\mathbf{n}\in\Omega_N}\langle p_N(t) - q_N(t), \mathbf{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\mathbf{n}}\rangle^2.$$

By Ito's formula,

$$\begin{aligned}&\langle p_N(t) - q_N(t), \mathbf{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\mathbf{n}}\rangle^2 = \langle p_N(0) - q_N(0), \mathbf{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\mathbf{n}}\rangle^2 \\ &+ 2\int_0^t\mathscr{A}(p_N(s) - p(s), \langle p_N(s) - q_N(s), \mathbf{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\mathbf{n}}\rangle\mathbf{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\mathbf{n}})ds \\ &+ 2\sum_{k=1}^{d}\int_0^t\langle L_k(p_N(s) - p(s)), \\ &\quad\langle p_N(s) - q_N(s), \mathbf{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\mathbf{n}}\rangle\mathbf{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\mathbf{n}}\rangle dY^k_s \\ &+ \sum_{k=1}^{d}\int_0^t\langle L_k(p_N(s) - p(s)), \mathbf{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\mathbf{n}}\rangle^2 ds.\end{aligned}$$

Therefore,

$$\begin{aligned}\|p_N(t) - q_N(t)\|^2 =& \|p_N(0) - q_N(0)\|^2 \\ &+ 2\int_0^t\mathscr{A}(p_N(s) - p(s), p_N(s) - q_N(s))ds \\ &+ 2\sum_{k=1}^{d}\int_0^t\langle L_k(p_N(s) - p(s)), p_N(s) - q_N(s)\rangle dY^k_s \\ &+ \sum_{k=1}^{d}\int_0^t\sum_{\mathbf{n}\in\Omega_N}\langle L_k(p_N(s) - p(s)), \mathbf{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\mathbf{n}}\rangle^2 ds.\end{aligned} \tag{21}$$

According to Bessel's inequality,

$$\sum_{\mathbf{n}\in\Omega_N}\langle L_k(p_N(s) - p(s)), \mathbf{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\mathbf{n}}\rangle^2 \le \|L_k(p_N(s) - p(s))\|^2.$$

Thus, taking expectations for both sides in (21), we obtain

$$\begin{aligned}&E\|p_N(t) - q_N(t)\|^2 \\ \le& E\|p_N(0) - q_N(0)\|^2 \\ &+ 2E\int_0^t\mathscr{A}(p_N(s) - p(s), p_N(s) - q_N(s))ds\end{aligned}$$

$$
\begin{aligned}
&+\sum_{k=1}^{d} E\int_0^t \|L_k(p_N(s)-p(s))\|^2 ds\\
=&E\|p_N(0)-q_N(0)\|^2\\
&+2\int_0^t E\mathscr{A}(p_N(s)-q_N(s),p_N(s)-q_N(s))ds\\
&+\sum_{k=1}^{d}\int_0^t E\|L_k(p_N(s)-q_N(s))\|^2 ds\\
&+2\int_0^t \mathscr{A}(q_N(s)-p(s),p_N(s)-q_N(s))ds\\
&+\sum_{k=1}^{d}\int_0^t \sum_{k=1}^{d}\|L_k(q_N(s)-p(s))\|^2 ds\\
&+\sum_{k=1}^{d}\int_0^t E\langle L_k(p_N(s)-q_N(s)),\\
&L_k(q_N(s)-p(s))\rangle ds, \qquad (22)
\end{aligned}
$$

where we use the fact that the stochastic integral terms in (21) are martingales and have expectations zero. For a general case where the stochastic integral terms are local martingales, similar results can be obtained through the standard localization process.

Let us denote $\varrho_N(t)=p_N(t)-q_N(t)$, and consider the form of $\mathscr{A}$ and L_k in (16) and (17), respectively. With the uniformly elliptic property of $a(x)$ and the fact that all the coefficients vanish outside U, we have

$$
\begin{aligned}
&2\mathscr{A}(\varrho_N,\varrho_N)+\sum_{k=1}^{d}\|L_k\varrho_N\|^2\\
=&\int_U -\sum_{i,j=1}^{d} A^{ij}\mathcal{D}_{x_i}\varrho_N\mathcal{D}_{x_j}\varrho_N+2\left(\sum_{i=1}^{d}B^i\mathcal{D}_{x_i}\varrho_N+\widetilde{C}\varrho_N\right)\times\\
&\varrho_N dx+\sum_{k=1}^{d}\int_U\left(J_k\varrho_N+\sum_{i=1}^{d}\rho^{ki}\mathcal{D}_{x_i}\varrho_N\right)^2 dx\\
\le&-\delta|\varrho_N|^2_{\mathcal{W}^1_{\alpha,\beta}(U)}+2\int_U\left(\sum_{i=1}^{d}B^i\mathcal{D}_{x_i}\varrho_N+\widetilde{C}\varrho_N\right)\varrho_N dx\\
&+\sum_{k=1}^{d}\int_U\left(J_k^2\varrho_N^2+2J_k\varrho_N\left(\sum_{i=1}^{d}\rho^{ki}\mathcal{D}_{x_i}\varrho_N\right)\right)dx. \qquad (23)
\end{aligned}
$$

Other terms except $-\delta|\varrho_N|^2_{\mathcal{W}^1_{\alpha,\beta}(U)}$ can be dominated by $\epsilon|\varrho_N|^2_{\mathcal{W}^1_{\alpha,\beta}(U)}+K(\epsilon)\|\varrho_N 1_U\|^2$, where $\epsilon>0$ is an arbitrary positive number and $K(\epsilon)>0$ is a generic constant which may only depend on T, r, α, β, the coefficients in the system and ϵ.

In fact, for example, if we denote by K a generic constant that may only depend on T and the coefficients in the equation, then by the assumption of the theorem,

$$
\begin{aligned}
\int_U\left(\sum_{i=1}^{d}B^i\mathcal{D}_{x_i}\varrho_N\right)\varrho_N dx&\le K|\varrho_N|_{\mathcal{W}^1_{\alpha,\beta}(U)}\|\varrho_N 1_U\|\\
&\le\epsilon|\varrho_N|^2_{\mathcal{W}^1_{\alpha,\beta}(U)}+\frac{K}{2\epsilon}\|\varrho_N 1_U\|^2,
\end{aligned}
$$

where the first inequality holds according to the Cauchy–Schwarz inequality while the second holds because of the Young's inequality. Other terms on the right-hand side of (23) can be treated in a similar way.

Let us move back to the rest terms on the right hand side of (22).

$$
\begin{aligned}
&\mathscr{A}(q_N-p,p_N-q_N)\\
&\le K\|q_N-p\|_{\mathcal{W}^1_{\alpha,\beta}(U)}\|p_N-q_N\|_{\mathcal{W}^1_{\alpha,\beta}(U)}\\
&\le\epsilon\|\varrho_N\|^2_{\mathcal{W}^1_{\alpha,\beta}(U)}+K_\epsilon\|q_N-p\|^2_{\mathcal{W}^1_{\alpha,\beta}(U)}\\
&=\epsilon\|\varrho_N 1_U\|^2+\epsilon|\varrho_N|^2_{\mathcal{W}^1_{\alpha,\beta}(U)}+K_\epsilon\|q_N-p\|^2_{\mathcal{W}^1_{\alpha,\beta}(U)},
\end{aligned}
$$

$$
\sum_{k=1}^{d}\|L_k(q_N-p)\|^2\le K\|q_N-p\|^2_{\mathcal{W}^1_{\alpha,\beta}(U)},
$$

and

$$
\begin{aligned}
&\langle L_k(p_N-q_N),L_k(q_N-p)\rangle\\
&\le K\|\varrho_N\|_{\mathcal{W}^1_{\alpha,\beta}(U)}\|q_N-p\|^2_{\mathcal{W}^1_{\alpha,\beta}(U)}\\
&\le\epsilon\|\varrho_N\|^2_{\mathcal{W}^1_{\alpha,\beta}(U)}+K_\epsilon\|q_N-p\|^2_{\mathcal{W}^1_{\alpha,\beta}(U)}\\
&=\epsilon\|\varrho_N 1_U\|^2+\epsilon|\varrho_N|^2_{\mathcal{W}^1_{\alpha,\beta}(U)}+K_\epsilon\|q_N-p\|^2_{\mathcal{W}^1_{\alpha,\beta}(U)}.
\end{aligned}
$$

Therefore, from (22) and the fact that $\rho_N(0)=0$, we have

$$
\begin{aligned}
&E\|\varrho_N(t)1_U\|^2\le-\delta\int_0^t|\varrho_N|^2_{\mathcal{W}^1_{\alpha,\beta}(U)}ds+K(\epsilon)\|\varrho_N 1_U\|^2\\
&+\int_0^t\epsilon|\varrho_N|^2_{\mathcal{W}^1_{\alpha,\beta}(U)}+K(\epsilon)\|q_N-p\|^2_{\mathcal{W}^1_{\alpha,\beta}(U)}ds.
\end{aligned}
$$

Choose $\epsilon\in(0,\delta)$ and denote $\delta'=\delta-\epsilon$, we have

$$
\begin{aligned}
&E\|\varrho_N(t)1_U\|^2+\delta'\int_0^t|\varrho_N|^2_{\mathcal{W}^1_{\alpha,\beta}(U)}ds\\
&\le\int_0^t K(\epsilon)\|\varrho_N(s)1_U\|^2+K(\epsilon)\|q_N-p\|^2_{\mathcal{W}^1_{\alpha,\beta}(U)}ds.
\end{aligned}
$$

According to Gronwall's inequality, we have

$$
\begin{aligned}
E\|\varrho_N(t)1_U\|^2&\le K\int_0^t\|q_N-p\|^2_{\mathcal{W}^1_{\alpha,\beta}(U)}ds\\
&\le KT\max_{0\le t\le T}\|q_N(t)-p(t)\|^2_{\mathcal{W}^1_{\alpha,\beta}(U)}.
\end{aligned}
$$

Therefore, $E\|1_U(p_N(t)-p(t))\|^2\le E\|\varrho_N 1_U\|^2+E\|(q_N(t)-p(t))1_U\|^2$, and by Theorem 2,

$$
\begin{aligned}
&\sup_{0\le t\le T}E\|1_U(p_N(t)-p(t))\|^2\\
&\le K\max_{0\le t\le T}E\|q_N(t)-p(t)\|^2_{\mathcal{W}^1_{\alpha,\beta}(U)}\\
&\le KN^{1-r}\max_{0\le t\le T}E|p(t)|^2_{\mathcal{W}^r_{\alpha,\beta}(U)},
\end{aligned}
$$

where K is a generic constant depending on T, r, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and the coefficients in the system.

Remark 7. For practical implementations, one can use $\tilde{p}_N(t,x)\triangleq\max\{p_N(t,x),0\}$ to preserve the non-negativity, so that $\tilde{p}_N$ is actually a probability density function after normalization. In the meanwhile, since the target function $p(t,x)$ is itself an unnormalized probability density function and thus $p(t,x)\ge 0$ almost surely, we always have $|\tilde{p}_N(t,x)-p(t,x)|\le|p_N(t,x)-p(t,x)|$. Therefore, the convergence result (19) in Theorem 6 also holds true if p_N is replaced by $\tilde{p}_N$.

4. Numerical method for solving stochastic differential equations

Generally, the solution of the SDEs (11), which we need to solve in the procedure of Hermite–Galerkin approximation, does not have a closed form. Therefore, we need to further apply time discretization schemes and solve the SDE system numerically.

Let $P: 0=t_0<t_1<\cdots<t_n=T$ be a partition of the time period, where $t_i-t_{i-1}=\frac{T}{n}:=\Delta t$, $i=1,2,\ldots,n$. Our goal is to

obtain a good estimation of the value Ψ_t at time $t = t_i$, for each $1 \le i \le n$.

Here, we will introduce two kinds of time discretization schemes for the SDE system (11).

4.1. Euler scheme

For the given discretization P for time period $[0, T]$, an Euler approximation to SDE (11) is a continuous time stochastic process $\widetilde{\Psi} = \{\widetilde{\Psi}_t, 0 \le t \le T\}$ satisfying the iterative scheme

$$\widetilde{\Psi}_{t_l} = \widetilde{\Psi}_{t_{l-1}} + P\widetilde{\Psi}_{t_{l-1}}\Delta t + \sum_{k=1}^{d} Q^{(k)}\widetilde{\Psi}_{t_{l-1}}(Y^k_{t_l} - Y^k_{t_{l-1}}),$$

with initial value $\widetilde{\Psi}_0 = \Psi_0$, $l = 1, 2, \ldots, n$. Since we are more concerned about the values of $\widetilde{\Psi}$ at each time t_l, the values of $\widetilde{\Psi}_t$ with $t \in (t_{l-1}, t_l)$, for some l, can be simply determined by linear interpolation of $\widetilde{\Psi}_{t_{l-1}}$ and $\widetilde{\Psi}_{t_l}$, such that $\widetilde{\Psi}$ is a continuous stochastic process.

Theoretical analysis shows that Euler scheme is an order 0.5 scheme (Kloeden & Platen, 1992) and therefore, it may suffer from robustness problems for big time discretization step.

4.2. Curved scheme and milstein scheme

When the concise Euler scheme cannot meet our requirements for accuracy, we need to use more accurate numerical algorithms to solve SDE systems.

For one-dimensional cases, higher order time discretization schemes, such as Milstein scheme, can be obtained through stochastic Taylor expansion. As is introduced in Kloeden and Platen (1992), the stochastic Taylor expansion up to order 1.0 of the solution to the SDE (11) has the form

$$\begin{aligned}\Psi_{t_{i+1}} =& \Psi_{t_i} + P\Psi_{t_i}\Delta t + \sum_{j=1}^{d} Q^{(j)}\Psi_{t_i}(Y_{t_{i+1}} - Y_{t_i}) \\ &+ \sum_{j_1,j_2=1}^{d} Q^{(j_2)}Q^{(j_1)}\Psi_{t_i} I_{(j_1,j_2)} + R_3,\end{aligned} \tag{24}$$

where R_3 contains terms of order at least 1.5, and

$$I_{(j_1,j_2)} = \int_{t_i}^{t_{i+1}} \int_{t_i}^{s_1} dY^{j_1}_{s_2} dY^{j_2}_{s_1}$$

are multiple Ito integrals.

The iterative formula of Milstein scheme in one dimension is as follows,

$$\begin{aligned}\widetilde{\Psi}_{t_{i+1}} =& \widetilde{\Psi}_{t_i} + P\widetilde{\Psi}_{t_i}\Delta t + Q\widetilde{\Psi}_{t_i}(Y_{t_{i+1}} - Y_{t_i}) \\ &+ \frac{1}{2}Q^2\widetilde{\Psi}_{t_i}\left[(Y_{t_{i+1}} - Y_{t_i})^2 - \Delta t\right],\end{aligned}$$

where we use the fact that

$$\int_{t_i}^{t_{i+1}} \int_{t_i}^{s_1} dY_{s_2} dY_{s_1} = (Y_{t_{i+1}} - Y_{t_i})^2 - \Delta t.$$

For high dimensional cases, unfortunately, the multiple Ito integral $I_{(j_1,j_2)}$ cannot be expressed explicitly by the endpoint values $Y^{(j_1)}_{t_{i+1}}$, $Y^{(j_1)}_{t_i}$, $Y^{(j_2)}_{t_{i+1}}$ and $Y^{(j_2)}_{t_i}$ for different j_1 and j_2. We would like to use

$$\widetilde{I}_{(j_1,j_2)} = \frac{1}{2}(Y^{(j_1)}_{t_{i+1}} - Y^{(j_1)}_{t_i})(Y^{(j_2)}_{t_{i+1}} - Y^{(j_2)}_{t_i})$$

to estimate $I_{(j_1,j_2)}$ and obtain the following scheme

$$\begin{aligned}\widetilde{\Psi}_{t_{i+1}} =& \widetilde{\Psi}_{t_i} + P\widetilde{\Psi}_{t_i}\Delta t + \sum_{j=1}^{d} Q^{(j)}\widetilde{\Psi}_{t_i}(Y_{t_{i+1}} - Y_{t_i}) \\ &+ \sum_{j_1,j_2=1}^{d} Q^{(j_2)}Q^{(j_1)}\widetilde{\Psi}_{t_i}\widetilde{I}_{(j_1,j_2)}.\end{aligned} \tag{25}$$

In fact, scheme (25) can be regarded as a kind of curved scheme discussed in Armstrong and King (2022). Although the overall order of this scheme is still 0.5, numerical solutions provided by this scheme can better approximate a certain low dimensional manifold containing the exact solution.

Numerical results in the next section show that in the application of solving NLF problems, in comparison with the classical Euler scheme, Milstein schemes (24) and curved schemes (25) are more robust with respect to the time discretization step Δt, and provide more accurate estimations to the state process.

5. Numerical results

In this section, we will discuss the strategies of choosing scaling factor $\boldsymbol{\alpha}$ and translating factor $\boldsymbol{\beta}$, and provide two numerical results of solving nonlinear filtering problems with correlated noise using HGSM.

5.1. Determination of scaling and translating factors

When approximating functions with the Hermite basis, the optimal choice of scaling and translating factors is still an open problem. Generally speaking, a proper choice of scaling factor $\boldsymbol{\alpha}$ can promote the efficiency of approximating functions concentrating on the neighborhood of the origin, while the translating factor $\boldsymbol{\beta}$ largely decides the efficiency of approximating functions *away* from the origin. For NLF problems, the state process may not stay near the origin and therefore, the main part of the conditional density can also be away from the origin. Thus, the translating factor β is more important and we will focus on the choice of $\boldsymbol{\beta}$ in this section. For guidance of choosing scaling factor $\boldsymbol{\alpha}$, readers can refer to Luo and Yau (2013b).

If the state process wanders near the origin, then the conditional density concentrates on the neighborhood of the origin, and we can fix the translating factor $\boldsymbol{\beta} = 0$. If the state process tends to deviate from the origin, then the translating factor $\boldsymbol{\beta}$ need to be adjusted adaptively. The self-adjusting procedure is summarized as follows.

Fix a threshold value $\beta_0 > 0$ and an initial translating factor $\boldsymbol{\beta}$. If the $k(1 \le k \le d)$th component of the estimation of the state process given by the HGSM continues to exceed the threshold to the one side, i.e., several consecutive estimates $\widehat{x}^k \ge \beta_k + \beta_0$ (or $\widehat{x}^k \le \beta_k - \beta_0$), then the major part of the conditional density tends to move right (left) on the kth component, and we set the kth component of the new translating factor $\boldsymbol{\beta}'$ to be $\beta'_k = \beta_k + \beta_0$, $(\beta'_k = \beta_k - \beta_0)$.

Before the adjustment, the conditional density $p(t)$ is approximated by

$$p_N(t) = \sum_{\boldsymbol{n}\in\Omega_N} \psi_{\boldsymbol{n}}(t)\boldsymbol{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\boldsymbol{n}}(x).$$

After the adjustment, $p(t)$ is approximated by $\overline{p}_N(t) = \sum_{\boldsymbol{n}\in\Omega_N} \overline{\psi}_{\boldsymbol{n}}(t)\boldsymbol{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}'}_{\boldsymbol{n}}(x)$, with $\overline{\Psi} = T\Psi$, $T_{\boldsymbol{n_1}}, \boldsymbol{n_2} = \langle \boldsymbol{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\boldsymbol{n_2}}, \boldsymbol{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}'}_{\boldsymbol{n_1}} \rangle$.

Coefficients in the SDE (11) are also reset using the basis with respect to $\boldsymbol{\beta}'$: $P_{\boldsymbol{n_1},\boldsymbol{n_2}} = \mathscr{A}(\boldsymbol{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}'}_{\boldsymbol{n_2}}, \boldsymbol{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}'}_{\boldsymbol{n_1}})$, $Q^{(k)}_{\boldsymbol{n_1},\boldsymbol{n_2}} = \langle L_k\boldsymbol{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}'}_{\boldsymbol{n_2}}, \boldsymbol{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}'}_{\boldsymbol{n_1}} \rangle$, and we have move the HGSM to the new basis $\{\boldsymbol{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}'}_{\boldsymbol{n}}\}$.

The whole procedure of the NLF algorithm proposed in this paper is summarized in Algorithm 1.

Algorithm 1 NLF Algorithm Based on HGSM

Step 1: (Off-Line) Choose suitable parameters $\boldsymbol{\alpha}$ and N. Fix $\boldsymbol{\beta} = 0$ or set the threshold β_0. Conduct the Hermite–Galerkin approximation procedure and obtain the system of SDEs (11) satisfied by the Hermite–Fourier coefficients Ψ_t.
Step 2: (On-Line) When new observations come, use proper time discretization schemes to numerically solve the SDE system (11) and obtain the approximated Hermite–Fourier coefficients, $\widetilde{\Psi}_t$, at each time step. If the condition of changing translating factor is satisfied, then conduct the factor adjusting procedure.
Step 3: (On-Line) Compute the unnormalized conditional density function at each time step based on $\widetilde{\Psi}_t$ and obtain the conditional expectation of the state process.

5.2. One-dimensional cases

We first consider the following one-dimensional nonlinear filtering problem

$$\begin{cases} dx_t = dw_t + dv_t, \\ dy_t = x_t(1 + 0.25\sin(x_t))dt + dv_t, \end{cases} \tag{26}$$

where $v = \{v_t : 0 \le t \le T\}$ and $w = \{w_t : 0 \le t \le T\}$ are mutually independent standard 1-dimensional Brownian motion.

We would like to use this example to test the effect of different time discretization schemes on the HGSM. Therefore, we truncate the state process x_t and restrict $x_t \in [-3, 3]$, so that we can fix the translating factor $\beta \equiv 0$. The terminal time T is set to be 50. Other parameters in Hermite–Galerkin approximation procedure are chosen to be $\alpha = 1$ and $N = 10$. Both Euler and Milstein schemes are used to numerically solving the system of SDEs, respectively.

Hereafter, the rooted mean square errors (RMSEs) are used to assess the performance of each method and the RMSE at time t_i is calculated by

$$RMSE_i = \sqrt{\frac{1}{i+1}\sum_{j=0}^{i} |x_i - \hat{x}_i|^2}, \tag{27}$$

and the subscript is omitted when we consider the overall RMSE at the terminal time T.

Typical performances for HGSM with time discretization step $\Delta t = 0.01$ s and $\Delta t = 0.05$ s are shown in Figs. 1 and 2. The performances of the resampling particle filters with 400 particles are illustrated as a benchmark. Since the purpose of filtering is to give a precise estimation to the state process, and at each time, the expectations of the state process is equal to those of the conditional expectations given the observations, we also draw the trajectories of the state process in Figs. 1 and 2, as a reference. The RMSEs of the methods are shown in Tables 1 and 2.

When the time discretization step is $\Delta t = 0.01$ s, HGSM with both Euler and Milstein scheme can obtain a good estimation of the state process. However, if the Δt increases to 0.05 s, only HGSM with Milstein scheme can properly track the state process. In fact, in the 100 numerical experiments with time step $\Delta t = 0.05$ s, some simulations of the Euler scheme blow up, and therefore, the average RMSE in those experiments of the Euler scheme cannot be calculated and is reported to be 'NaN' in Table 2. Such results imply that for big time discretization steps, Milstein scheme is more suitable for solving the SDE systems arising in the HGSM, when solving nonlinear filtering problems.

In order to show that the numerical solution to the DMZ equation based on HGSM with Milstein scheme is a good approximation to the real solution, which is the conditional probability density of the state process given the observations, the conditional probability density function obtained by HGSM and the distribution of the particles at 9 time points uniformly distributed in the time period are shown in Fig. 3. At these time points, the shapes of the conditional probability density functions obtained by the HGSM is very closed to the outlines of the histograms of the particles, especially when most of the particles are inside the interval $[-3, 3]$. Since the distribution of the particles can well reflect the properties of the conditional probability density, the solution to the NLF based on HGSM approximates the conditional probability density well.

Table 1
The one-dimensional example: rooted mean square error (RMSE) and CPU time of two time discretization schemes with time step $\Delta t = 0.01$ s.

Scheme	Milstein scheme	Euler scheme
RMSE[a]	0.6334	0.5485
CPU time[b]	0.0570 s	0.0578 s

[a]Rooted mean square error (RMSE) of a method is estimated by the average RMSE in 100 experiments.
[b]The CPU time of a method is the average time cost in 100 experiments.

Table 2
The one-dimensional example: rooted mean square error (RMSE) and CPU time of two time discretization schemes with time step $\Delta t = 0.05$ s.

Scheme	Milstein scheme	Euler scheme
RMSE[a]	0.7569	NaN[c]
CPU time[b]	0.0102 s	0.0105 s

[a]Rooted mean square error (RMSE) of a method is estimated by the average RMSE in 100 experiments.
[b]The CPU time of a method is the average time cost in 100 experiments.
[c]The average RMSE is reported to be 'NaN', because some of the simulations blow up.

5.3. Two-dimensional case

Now, we consider the following nonlinear filtering problem with state and observation processes both in dimension two.

$$\begin{cases} dx_t^{(1)} = x_t^{(2)}dt + 0.1 * dw_t^{(1)} + dv_t^{(1)}, \\ dx_t^{(2)} = -x_t^{(1)}dt + 0.1 * dw_t^{(2)} + dv_t^{(2)}, \\ dy_t^{(1)} = x_t^{(1)}(1 + \cos x_t^{(2)})dt + dv_t^{(1)}, \\ dy_t^{(2)} = x_t^{(2)}(1 + \cos x_t^{(1)})dt + dv_t^{(2)}, \end{cases} \tag{28}$$

where $v^{(1)} = \{v_t^{(1)} : 0 \le t \le T\}$, $v^{(2)} = \{v_t^{(2)} : 0 \le t \le T\}$, $w^{(1)} = \{w_t^{(1)} : 0 \le t \le T\}$ and $w^{(2)} = \{w_t^{(2)} : 0 \le t \le T\}$ are mutually independent standard 1-dimensional Brownian motions. The state process simulates a uniform circular motion under white noise disturbances. Since the deterministic part of the state process is at the critical point of stability, simulation results show that the effect of the stochastic part is gradually enlarging the radius of the circular motion. Therefore, adjusting the transforming factor $\boldsymbol{\beta}$ during the filtering process is necessary.

Here we set the time discretization steps to be $\Delta t = 0.01$ s and use curved scheme (25) to solve the SDEs in the HGSM. The terminal time T is also set to be 50. We choose $N = 9$ and thus the dimension of the finite dimensional subspace $S^N_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ is $|\Omega_N| = (N+1)^2 = 100$. The scalar factor in Hermite–Galerkin approximation procedure is set to be $\boldsymbol{\alpha} = (\alpha_1, \alpha_2) = (1, 1)$ and the transforming factor takes value in $\boldsymbol{\beta} = (\beta_1, \beta_2) = (2k_1, 2k_2)$, with $k_1, k_2 \in \mathbb{Z}$, because Hermite functions with scalar factor $\alpha =$

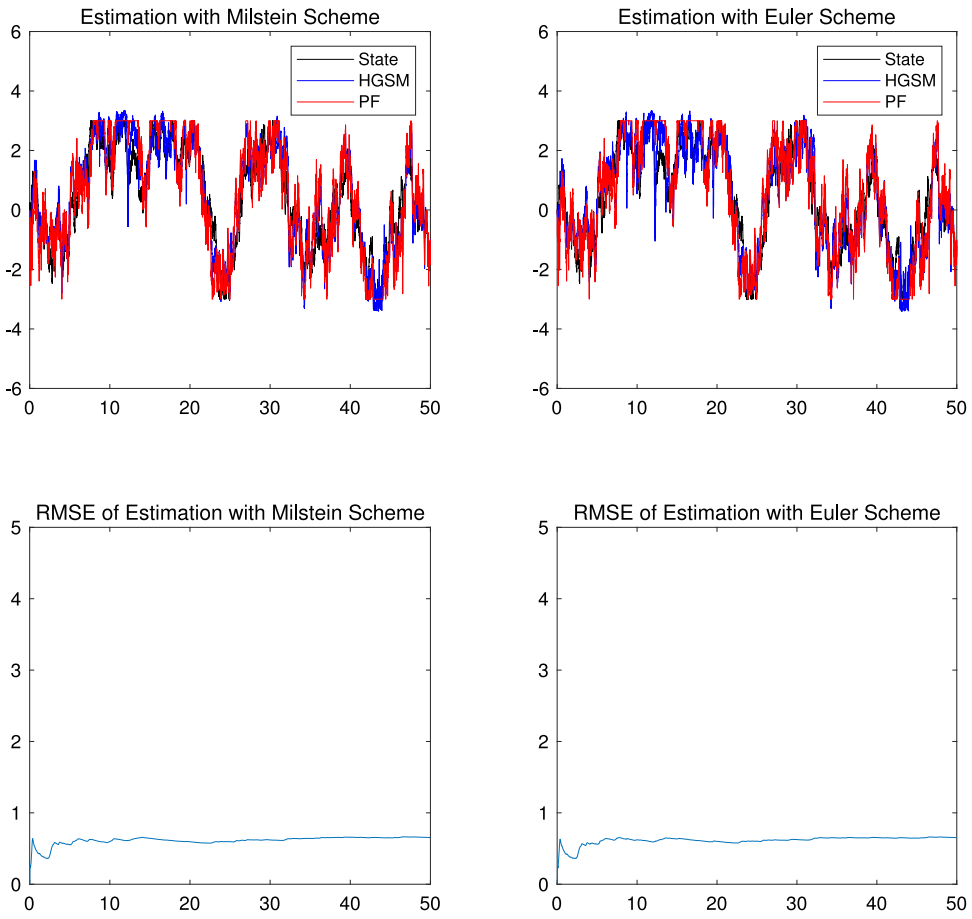

Fig. 1. The one-dimensional example: a typical performance for HGSM using Milstein scheme and Euler scheme with time discretization step $\Delta t = 0.01$ s.

1 and translating factor β can efficiently approximate probability density functions with expectation value $\mu \in [\beta - 2, \beta + 2]$.

The transforming factor $\boldsymbol{\beta}$ will change as long as one component of 20 estimations in a row given by HGSM are outside the interval $[\beta_i - 2, \beta_i + 2]$, and fall to the same side. For example, if at time t_0, 20 estimations of $x^{(1)}$, $\{x^{(1)}_{t_0-k} : k = 0, 1, \ldots, 9\}$ are larger than $\beta_1 + 2$, then the transforming factor will be reset to $\boldsymbol{\beta} = (\beta_1 + 2, \beta_2)$ at time t_0.

Since the parameter matrices in curved scheme can be calculated beforehand, the main on-line computation cost comes from multiplications of sparse matrices of size $(N + 1)^2$ with only $O(N)$ non-zero entries each row. Therefore, at each time step, the computation cost is $O(N^3)$. The change of translating factor β will bring additional computations, because Hermite–Fourier coefficients $\tilde{\Psi}_t$ need to be transferred to adapt to the new basis. However, the introduction of adaptive translating factor $\boldsymbol{\beta}$ significantly improves the approximation capability of Hermite basis. Accurate estimation to the state process can be obtained with smaller N. In the meanwhile, the transferring procedure also only consists of the multiplication of matrices and vectors and the overall computation cost is still $O(N^3)$.

The result of the resampling particle filter with particle number $N_{PF} = 400$ is again used as a benchmark in this example and the trajectories of the state process are also shown as a reference. A typical performance of HGSM and particle filter is shown in Fig. 4. Green vertical lines in Fig. 4 demonstrate the time when the corresponding component of the translating factor changes. The rooted mean square errors of the estimation to the state processes and the computation time of HGSM as well as particle filter are shown in Table 3. We also consider the approximation capability of the HGSM on higher moments of the conditional probability distributions. For this particular trajectory, the RMSEs of estimating the first, second and third moments of the process using HGSM and resampling particle filter evolves as is shown in Fig. 5.

Table 3
The two-dimensional example: rooted mean square error (RMSE) and CPU time of HGSM and the particle filter for tracking the state process in the two-dimensional example ($\Delta t = 0.01$ s).

Method	HGSM	Particle filter with $N_{PF} = 400$
RMSE[a]	1.5546	1.6510
CPU time[b]	3.3421 s	38.0886 s

[a]Rooted mean square error (RMSE) of a method is estimated by the average RMSE in 100 experiments.
[b]The CPU time of a method is the average time cost in 100 experiments.

Results show that the conditional mean and also the second and third moment of the conditional probability distributions are well approximated by the HGSM in this example. In terms of the RMSEs, the HGSM performs almost on a par with the resampling particle filter in this particular trajectory (Fig. 5), and the average RMSE of HGSM in the 100 experiments is a little smaller than that of the resampling particle filter (Table 3). In the meanwhile, the average computation time it takes is only 3.3421s, which is only about $\frac{1}{10}$ of the computation time of resampling particle filter with 400 particles. Therefore, in this example, the HGSM method is more capable in giving instantaneous estimations to the state process.

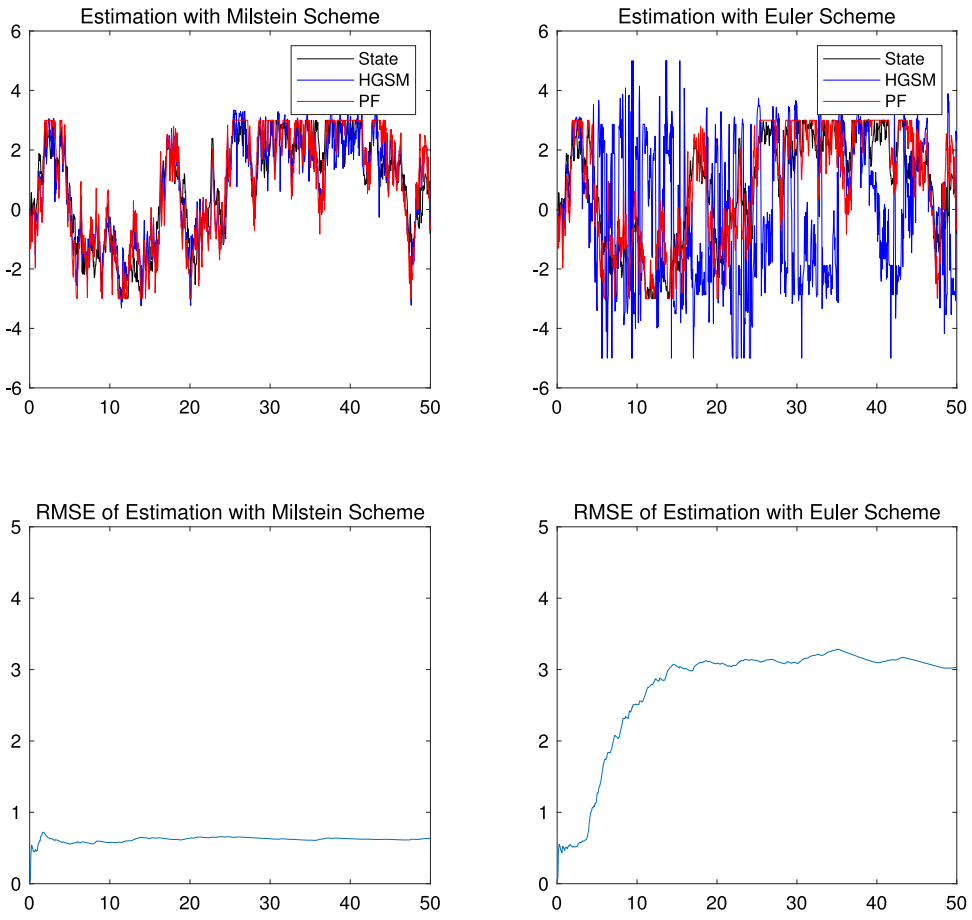

Fig. 2. The one-dimensional example: a typical performance for HGSM using Milstein scheme and Euler scheme with time discretization step $\Delta t = 0.05$ s.

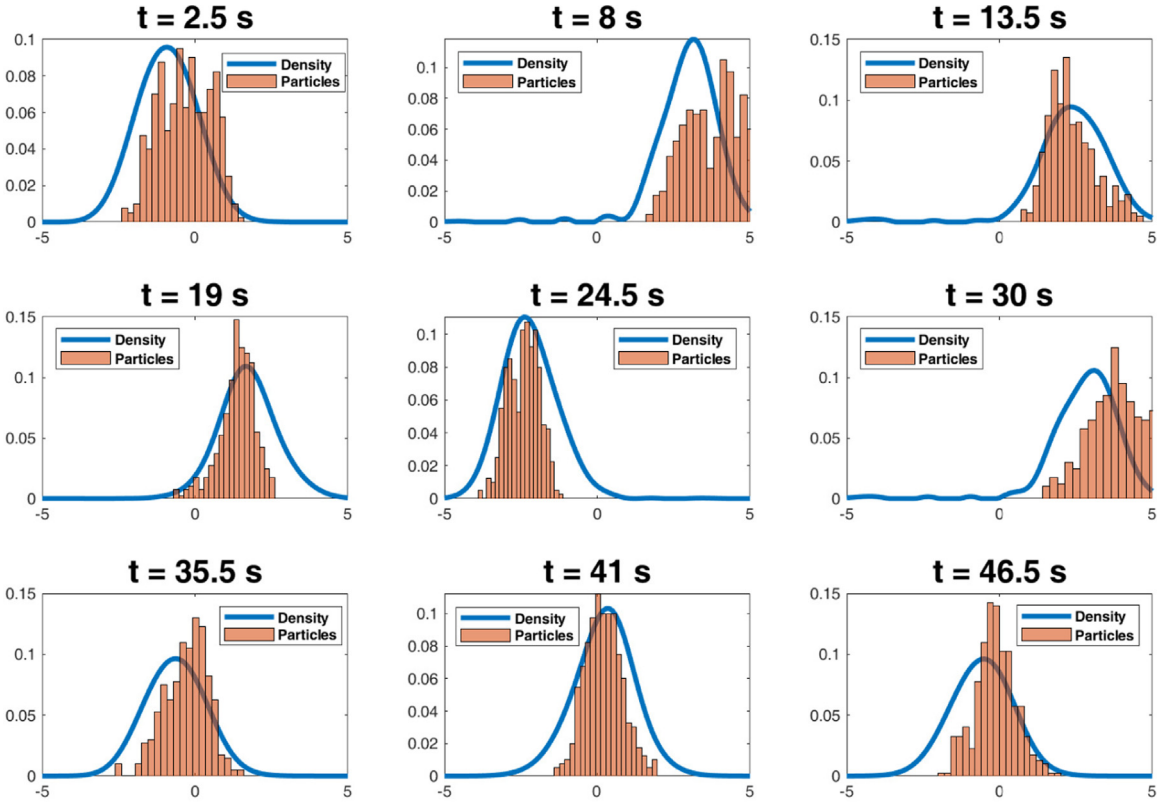

Fig. 3. The one-dimensional example: conditional probability density function obtained by HGSM with Milstein scheme and the distribution of the particles.

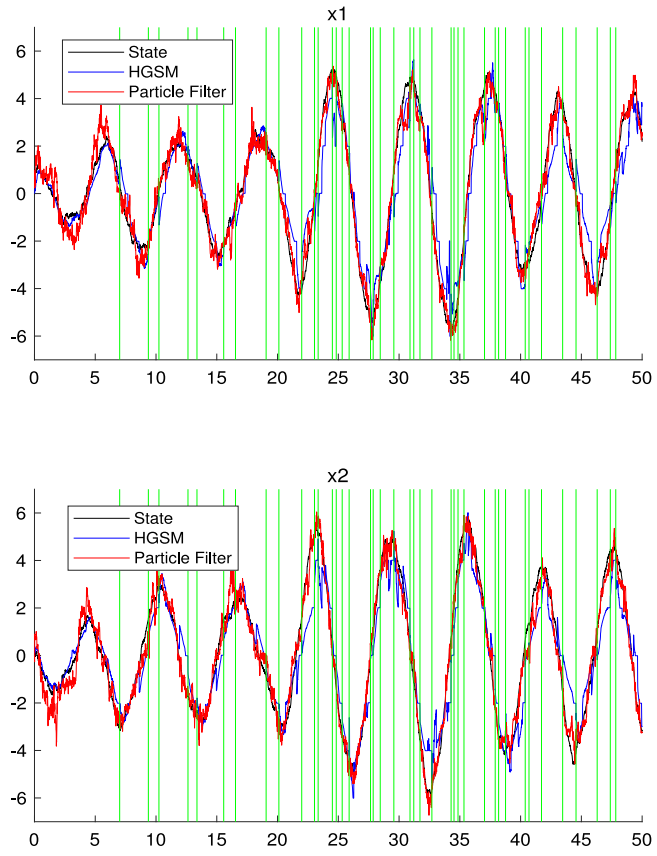

Fig. 4. The two-dimensional example: a typical tracking performance with time discretization step $\Delta t = 0.01$ s for HGSM and particle filter with $N = 400$ particles. Green vertical lines demonstrate the time when translating factor $\boldsymbol{\beta}$ changes.

6. Conclusions

In this paper, we use HGSM to solve the DMZ equation which occurs in the NLF with correlated noise. The convergence result as well as the convergence rate of the algorithm on bounded open sets are studied under mild conditions. After projecting the exact solution onto the finite dimensional space spanned by Hermite functions, we obtain a system of SDEs generated by the observation process. Accurate time discretization schemes such as the Milstein scheme and curved schemes can be applied to solving the SDE system. Numerical results show that the algorithm with these schemes can provide a robust and accurate estimation to the state process, while the computation cost can be sharply reduced in comparison with the resampling particle filter, in order to get a similar mean square error.

With the dimension of the NLF increasing, the cardinality of the set Ω_N is $(N+1)^d$, which increases exponentially. The computational cost of the algorithm will thus also increase rapidly for high dimensional cases, which is well-known as the *curse of dimensionality*. Numerical results show that with proper scaling and translating factors, a relatively small N can provide an accurate estimation to the state process. Besides, for a particular N, another possible approach to deal with this problem is reducing the number of basis functions by choosing a proper subset of multi-indices in Ω_N, which is also discussed in relevant works such as (Luo & Yau, 2013b; Wang et al., 2020). Combining the above two techniques together, we think that the algorithm proposed in this paper also has the potential of dealing with NLF in medium high dimensions.

Generally speaking, the algorithm of numerically solving the DMZ equation based on HGSM performs well in NLF with correlated noise in low dimension cases, and has the potential of being generalized to NLF in medium high dimensions. Theoretical and numerical results both imply that this algorithm is especially suitable for NLF systems in which the state process is contained in a bounded domain with high probability.

Appendix A. Proof of Theorem 2

Proof. For all $u \in \mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbb{R}^d)$, $r \in \mathbb{N}$, we have the expression $u(x) = \sum_{\boldsymbol{n}\in\mathbb{N}^d} \widehat{u}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\boldsymbol{n}} \boldsymbol{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\boldsymbol{n}}(x)$ and $P_N u(x) - u(x) = -\sum_{\boldsymbol{n}\in\Omega_N^c} \widehat{u}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\boldsymbol{n}} \boldsymbol{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\boldsymbol{n}}(x)$, where $\Omega_N^c = \{\boldsymbol{n} \in \mathbb{N}^d : n_j \geq N, \text{ for some } 1 \leq j \leq d\}$ and $\widehat{u}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\boldsymbol{n}} = \langle u, \boldsymbol{H}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\boldsymbol{n}}\rangle$ are the Hermite–Fourier coefficients.

According to the Parseval's equality, for $l \leq r$, we have

$$|P_N u - u|^2_{\mathcal{W}^l_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbb{R}^d)} = \sum_{j=1}^d \sum_{\boldsymbol{n}\in\Omega_N^c} \mu_{n_j,l} |\widehat{u}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\boldsymbol{n}}|^2,$$

where $\mu_{n_j,l}$ are defined in (13).

For all $1 \leq j \leq d$, denote by $\Lambda_N^{1,j} = \{\boldsymbol{n} \in \Omega_N^c, n_j > N\}$ and $\Lambda_N^{2,j} = \{\boldsymbol{n} \in \Omega_N^c, n_j \leq N\}$, then

$$\begin{aligned}&\sum_{\boldsymbol{n}\in\Omega_N^c} \mu_{n_j,l} |\widehat{u}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\boldsymbol{n}}|^2 \\ &= \sum_{\boldsymbol{n}\in\Lambda_N^{1,j}} \mu_{n_j,l} |\widehat{u}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\boldsymbol{n}}|^2 + \sum_{\boldsymbol{n}\in\Lambda_N^{2,j}} \mu_{n_j,l} |\widehat{u}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\boldsymbol{n}}|^2 \triangleq I_1 + I_2\end{aligned}$$

For I_1, we have the following estimation:

$$I_1 \leq \max_{\boldsymbol{n}\in\Lambda_N^{1,j}} \left\{\frac{\mu_{n_j,l}}{\mu_{n_j,r}}\right\} \sum_{\boldsymbol{n}\in\Lambda_N^{1,j}} \mu_{n_j,r} |\widehat{u}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\boldsymbol{n}}|^2$$

Notice that

$$\sum_{\boldsymbol{n}\in\Lambda_N^{1,j}} \mu_{n_j,r} |\widehat{u}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\boldsymbol{n}}|^2 \leq \sum_{\boldsymbol{n}\in\mathbb{N}^d} \mu_{n_j,r} |\widehat{u}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\boldsymbol{n}}|^2 = |u|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbb{R}^d)}$$

and

$$\max_{\boldsymbol{n}\in\Lambda_N^{1,j}} \left\{\frac{\mu_{n_j,l}}{\mu_{n_j,r}}\right\} \leq (2\alpha_j^2)^{l-r} N^{l-r}.$$

Therefore, $I_1 \leq (2\alpha_j^2)^{l-r} N^{l-r} |u|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbb{R}^d)}$.

Now, we come to the estimation of I_2. For $\boldsymbol{n} \in \Lambda_N^{2,j}$, there exists some $k_j \neq j$ such that $n_k > N \geq n_j$, then

$$I_2 \leq \max_{\boldsymbol{n}\in\Lambda_N^{2,j}} \left\{\frac{\mu_{n_j,l}}{\mu_{n_{k_j},r}}\right\} \sum_{\boldsymbol{n}\in\Lambda_N^{2,j}} \mu_{n_{k_j},r} |\widehat{u}^{\boldsymbol{\alpha},\boldsymbol{\beta}}_{\boldsymbol{n}}|^2$$

and

$$\max_{\boldsymbol{n}\in\Lambda_N^{2,j}} \left\{\frac{\mu_{n_j,l}}{\mu_{n_{k_j},r}}\right\} \leq 2^{l-r} \frac{\alpha_j^{2l}}{\alpha_{k_j}^{2r}} \frac{N^l}{(N-r+1)^r}.$$

When $N > 2(r-1)$,

$$\max_{\boldsymbol{n}\in\Lambda_N^{2,j}} \left\{\frac{\mu_{n_j,l}}{\mu_{n_{k_j},r}}\right\} \leq \frac{(2\alpha_j^2)^l}{\alpha_{k_j}^{2r}} N^{l-r}$$

Therefore,

$$\begin{aligned}&|P_N u - u|^2_{\mathcal{W}^l_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbb{R}^d)} \\ &\quad\leq \sum_{j=1}^d \left((2\alpha_j^2)^{l-r} + \frac{(2\alpha_j^2)^l}{\alpha_{k_j}^{2r}}\right) N^{l-r} |u|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbb{R}^d)}\end{aligned}$$

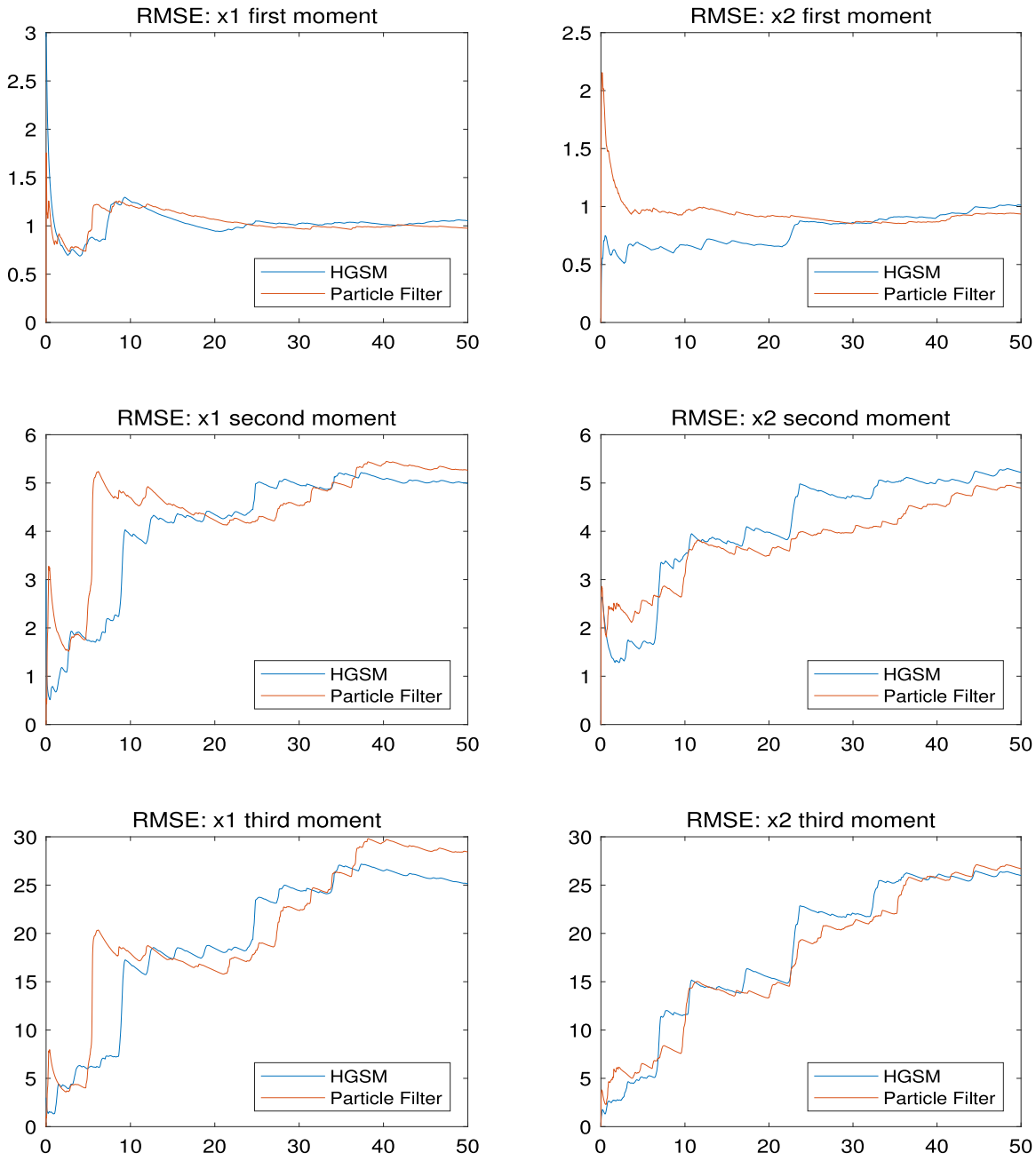

Fig. 5. The two-dimensional example: the evolution of the RMSE for the estimation of the first, second and third moments using the HGSM and the resampling particle filter.

When $\frac{\max_{1\le j\le d}|\alpha_j|}{\min_{1\le j\le d}|\alpha_j|}\le c_0$,

$$|P_N u-u|^2_{\mathcal{W}^l_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbb{R}^d)}\le d(2c_0^{2r}+1)(2|\boldsymbol{\alpha}|^2_\infty)^{l-r}N^{l-r}|u|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbb{R}^d)}$$

Similar estimations can be done for other terms in $\|P_N u-u\|_{\mathcal{W}^l_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbb{R}^d)}$ and therefore, we have

$$\|P_N u-u\|_{\mathcal{W}^l_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbb{R}^d)}\le C_{d,l}\sqrt{2c_0^{2r}+1}(2|\boldsymbol{\alpha}|^2_\infty)^{\frac{l-r}{2}}N^{\frac{l-r}{2}}|u|_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbb{R}^d)}$$

where $C_{d,l}$ is a constant that only depends on d and l.

Appendix B. Proof of Theorem 5

For any test function $v\in C_0^\infty(U)$, according to the duality of $\boldsymbol{\mathcal{D}}^{\boldsymbol{k}}_{\boldsymbol{x}}$ and $\overline{\boldsymbol{\mathcal{D}}}^{\boldsymbol{k}}_{\boldsymbol{x}}$, we have

$$\begin{aligned}&-\frac{1}{2}\sum_{i,j=1}^d\langle A^{ij}\mathcal{D}_{x_i}v,\mathcal{D}_{x_j}v\rangle_r+\sum_{i=1}^d\langle B^i\mathcal{D}_{x_i}v,v\rangle_r+\langle\widetilde{C}v,v\rangle_r\\&=-\frac{1}{2}\sum_{i,j=1}^d\sum_{|\boldsymbol{k}|=r}C^{\boldsymbol{k}}\left\langle\boldsymbol{\mathcal{D}}^{\boldsymbol{k}}_{\boldsymbol{x}}\left(\mathcal{D}_{x_j}v\right),A^{ij}\boldsymbol{\mathcal{D}}^{\boldsymbol{k}}_{\boldsymbol{x}}\left(\mathcal{D}_{x_i}v\right)\right\rangle\\&-\frac{1}{2}\sum_{i,j=1}^d\sum_{|\boldsymbol{k}|=r}\sum_{\substack{\boldsymbol{l}+\boldsymbol{m}=\boldsymbol{k}\\|\boldsymbol{l}|\ge1}}C^{\boldsymbol{k}}\left\langle\boldsymbol{\mathcal{D}}^{\boldsymbol{l}}_{\boldsymbol{x}}A^{ij}\boldsymbol{\mathcal{D}}^{\boldsymbol{m}}_{\boldsymbol{x}}\left(\mathcal{D}_{x_i}v\right),\boldsymbol{\mathcal{D}}^{\boldsymbol{k}}_{\boldsymbol{x}}\left(\mathcal{D}_{x_j}v\right)\right\rangle\end{aligned}$$

$$+\sum_{i=1}^{d}\sum_{|\boldsymbol{k}|=r} C^{\boldsymbol{k}}\left\langle \mathcal{D}_{\boldsymbol{x}}^{\boldsymbol{k}} v, B^i \mathcal{D}_{\boldsymbol{x}}^{\boldsymbol{k}}\left(\mathcal{D}_{x_i} v\right)\right\rangle + U_1(v),$$

where $C^{\boldsymbol{k}} = \frac{|\boldsymbol{k}|!}{k_1!\cdots k_d!}$, $\forall |\boldsymbol{k}| \le r$, $\langle\cdot,\cdot\rangle$ denotes the inner product in $L^2(U)$ and $|U_1(v)| \le K\|v\|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)}$ for some constant $K > 0$.

Similarly,

$$\begin{aligned}
\|L_i v\|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)} &= \left\| \sum_{j=1}^{d} \rho^{ij}\mathcal{D}_{x_j} v + J_i v \right\|_r^2 \\
&= \sum_{|\boldsymbol{k}|=r} C^{\boldsymbol{k}} \left\langle \sum_{j=1}^{d} \rho^{ij} \mathcal{D}_{\boldsymbol{x}}^{\boldsymbol{k}}\left(\mathcal{D}_{x_j} v\right), \sum_{j=1}^{d} \rho^{ij} \mathcal{D}_{\boldsymbol{x}}^{\boldsymbol{k}}\left(\mathcal{D}_{x_j} v\right) \right\rangle \\
&+ \sum_{|\boldsymbol{k}|=r} \sum_{\substack{\boldsymbol{l}+\boldsymbol{m}=\boldsymbol{k} \\ |\boldsymbol{l}|\ge 1}} C^{\boldsymbol{k}} \left\langle \sum_{j=1}^{d} \mathcal{D}_{\boldsymbol{x}}^{\boldsymbol{l}}(\rho^{ij}) \mathcal{D}_{\boldsymbol{x}}^{\boldsymbol{m}}\left(\mathcal{D}_{x_j} v\right), \right. \\
&\left. \sum_{j=1}^{d} \rho^{ij} \mathcal{D}_{\boldsymbol{x}}^{\boldsymbol{k}}\left(\mathcal{D}_{x_j} v\right) \right\rangle \\
&+ 2\sum_{|\boldsymbol{k}|=r} C^{\boldsymbol{k}} \left\langle \sum_{j=1}^{d} \rho^{ij} \mathcal{D}_{\boldsymbol{x}^{\boldsymbol{k}}}\left(\mathcal{D}_{x_j} v\right), J_i \mathcal{D}_{\boldsymbol{x}}^{\boldsymbol{k}} v \right\rangle + U_2(v),
\end{aligned}$$

where $|U_2(v)| \le K'\|v\|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)}$, for some $K' > 0$.

According to the uniformly elliptic condition of $a(x)$ and the fact that $A(x) = a(x) + c(x) = a(x) + \rho(x)\rho(x)^\top$, we have

$$\begin{aligned}
&-\frac{1}{2}\sum_{i,j=1}^{d}\langle A^{ij}\mathcal{D}_{x_i} v, \mathcal{D}_{x_j} v\rangle_r + \sum_{i=1}^{d}\langle B^i \mathcal{D}_{x_i} v, v\rangle_r \\
&\qquad + \langle \widetilde{C} v, v\rangle_r + \frac{1}{2}\sum_{i=1}^{d}\|L_i v\|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)} \\
&\le -\frac{\delta}{2}\sum_{|\boldsymbol{k}|=r} C^{\alpha}\sum_{j=1}^{d}\left\|\mathcal{D}_{\boldsymbol{x}}^{\boldsymbol{k}}\left(\mathcal{D}_{x_j} v\right)\right\|^2 + V(v) + U_1(v) + U_2(v),
\end{aligned}$$

with $|V(v)|^2 \le K''\|v\|_{\mathcal{W}^{r+1}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)}\|v\|_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)}$, for some constant $K'' > 0$.

By Cauchy–Schwarz inequality, for arbitrarily chosen $\epsilon > 0$, $|V(v)| \le \epsilon\|v\|^2_{\mathcal{W}^{r+1}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)} + K(\epsilon)\|v\|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)}$.

Therefore,

$$\begin{aligned}
&-\frac{1}{2}\sum_{i,j=1}^{d}\langle A^{ij}\mathcal{D}_{x_i} v, \mathcal{D}_{x_j} v\rangle_r + \sum_{i=1}^{d}\langle B^i \mathcal{D}_{x_i} v, v\rangle_r \\
&+ \langle \widetilde{C} v, v\rangle_r + \frac{1}{2}\sum_{i=1}^{d}\|L_i v\|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)} \\
&\le -\delta'\|v\|^2_{\mathcal{W}^{r+1}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)} + M'\|v\|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)} \qquad \text{(B.1)}
\end{aligned}$$

holds for all $v \in C_0^\infty(U)$. Since $\overline{\mathcal{W}}^{r+1}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)$ is the closure of $C_0^\infty(U)$ under the $\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}$-norm, (B.1) holds for all $u \in \overline{\mathcal{W}}^{r+1}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)$.

Given the coercivity condition (B.1), the estimation (15) can be derived using standard Galerkin approach.

Let $\{d_i\}_{i=1}^\infty$ be a orthonormal basis of $\overline{\mathcal{W}}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)$, and assume that $\{d_i\}_{i=1}^\infty \subset \overline{\mathcal{W}}^{r+1}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)$.

For each $n \in \mathbb{N}$, define $p_n(t,x) = \sum_{i=1}^n p_n^i(t) d_i(x)$, with $p_n^i(t)$ given by the following SDEs:

$$\begin{aligned}
p_n^i(t) = \langle \varphi, d_i\rangle_r &+ \int_0^t \langle L_0 p_n(s), d_i\rangle_r ds \\
&+ \sum_{l=1}^{d}\int_0^t \langle L_l p_n(s), d_i\rangle_r dY_s^l.
\end{aligned}$$

By Ito's formula, we have

$$\begin{aligned}
|p_n^i(t)|^2 =& |\langle \varphi, d_i\rangle_r|^2 + 2\int_0^t \langle L_0 p_n(s), p_n^i(s) d_i\rangle_r ds \\
&+ 2\sum_{l=1}^{d}\int_0^t \langle L_l p_n(s), p_n^i(s) d_i\rangle_r dY_s^l \\
&+ \sum_{l=1}^{d}\int_0^t |\langle L_l p_n(s), d_i\rangle_r|^2 ds.
\end{aligned}$$

Since $\{d_i\}_{i=1}^\infty$ is an orthonormal basis of $\overline{\mathcal{W}}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)$,

$$\begin{aligned}
&E\|p_n(t)\|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)} = E\sum_{i=1}^{n}|p_n^i(t)|^2 \\
&\le \|\varphi\|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)} + 2E\int_0^t \langle L_0 p_n(s), p_n(s)\rangle_r ds \\
&\quad + \sum_{l=1}^{d} E\int_0^t \|L_l p_n(s)\|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)} ds \\
&\le \|\varphi\|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)} - E\int_0^t 2\delta'\|p_n(s)\|^2_{\mathcal{W}^{r+1}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)} ds \\
&\quad + 2M'\int_0^t \|p_n(s)\|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)} ds.
\end{aligned}$$

According to Gronwall's inequality and Burkholder–Davis–Gundy inequality, we have

$$\begin{aligned}
E\sup_{t\le T}\|p_n(t)\|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)} + E\int_0^T \|p_n(t)\|^2_{\mathcal{W}^{r+1}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)} dt \\
\le M\|\varphi\|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)},
\end{aligned}$$

for all $n \in \mathbb{N}$, where $M > 0$ is a constant that is independent to n.

Therefore, there exists a subsequence of $\{p_n(t)\}$, which converges to an element $p \in L^\infty_\omega([0,T], \overline{\mathcal{W}}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)) \cap L^2_\omega([0,T], \overline{\mathcal{W}}^{r+1}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U))$, which is the generalized solution of (2) and inequality (15)

$$\begin{aligned}
E\sup_{t\le T}\|p(t)\|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)} + E\int_0^T \|p(t)\|^2_{\mathcal{W}^{r+1}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)} dt \\
\le M\|\varphi\|^2_{\mathcal{W}^r_{\boldsymbol{\alpha},\boldsymbol{\beta}}(U)}
\end{aligned}$$

holds. Now, we have finished the proof of Theorem 5.

References

Ahmed, N. U., & Radaideh, S. M. (1997). A powerful numerical technique solving Zakai equation for nonlinear filtering. *Dynamics and Control*, *7*, 293–308.

Armstrong, J., & King, T. (2022). Curved schemes for stochastic differential equations on, or near, manifolds. *Proceedings of the Royal Society A*, *478*, 20210785.

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

Bain, A., & Crisan, D. (2009). *Fundamentals of stochastic filtering* (first ed.). New York: Springer-Verlag.

Baras, J. S., Blankenship, G. L., & Hopkins, W. E., Jr. (1983). Existence, uniqueness, and asymptotic behaviour of solutions to a class of Zakai equations with unbounded coefficients. *IEEE Transactions on Automatic Control*, *28*(2), 203–214.

Budhiraja, A., Chen, L., & Lee, C. (2007). A survey of numerical methods for nonlinear filtering problems. *Physica D*, *230*, 27–36.

Ceci, C., & Colaneri, K. (2012). Nonlinear filtering for jump diffusion observations. *Advances in Applied Probability*, *44*(3), 678–701.

Cheng, M., Hou, T. Y., & Zhang, Z. (2013). A dynamically bi-orthogonal method for time-dependent stochastic partial differential equations: Derivation and algorithms. *Journal of Computational Physics*, *242*, 843–868.

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

Dong, W., Luo, X., & Yau, S. S.-T. (2021). Solving nonlinear filtering problems in real time by Legendre Galerkin spectral method. *IEEE Transactions on Automatic Control*, *66*(4), 1559–1572.

Duncan, T. E. (1967). *Probability densities for diffusion processes with applications to nonlinear filtering theory* (Ph.D. Dissertation), Stanford, CA, USA: Stanford University.

Florchinger, P., & Le Gland, F. (1991). Time-discretization of the Zakai equation for diffusion processes observed in correlated noise. *Stochastics and Stochastics Reports*, *35*, 233–256.

Frey, R., Schmidt, T., & Xu, L. (2013). On Galerkin approximations for the Zakai equation with diffusive and point process observations. *SIAM Journal on Numerical Analysis*, *51*(4), 2036–2062.

Funaro, D., & Kavian, O. (1991). Approximation of some diffusion evolution equations in unbounded domains by Hermite functions. *Mathematics of Computation*, *57*(196), 597–619.

Gottlieb, D., & Orszag, S. (1977). *Numerical analysis of spectral methods: Theory and applications* (first ed.). Philadelphia: SIAM-CBMS.

Kloeden, P. E., & Platen, E. (1992). *Numerical solution of stochastic differential equations* (first ed.). Berlin, Heidelberg: Springer.

Liu, J. S., & Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, *93*(443), 1032–1044.

Lototsky, S. V. (2003). Nonlinear filtering of diffusion processes in correlated noise: Analysis by separation of variables. *Applied Mathematics and Optimization*, *47*(2), 167–194.

Lototsky, S., Mikulevicius, R., & Rozovskii, B. L. (1997). Nonlinear filtering revisited: A spectral approach. *SIAM Journal on Control and Optimization*, *35*(2), 435–461.

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

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

Luo, X., & Yau, S. S.-T. (2013c). Hermite spectral method with hyperbolic cross approximations to high-dimensional parabolic PDEs. *SIAM Journal on Numerical Analysis*, *51*(6), 3186–3212.

Mortensen, R. E. (1996). *Optimal control of continuous time stochastic systems* (Ph.D. dissertation), CA, USA: University of California at Berkeley.

Rozovsky, B. L. (1972). Stochastic partial differential equations arising in nonlinear filtering problems. *Uspekhi Matematicheskikh Nauk*, *27*, 213–214.

Rozovsky, B. L., & Lototsky, S. V. (2018). *Stochastic evolution systems* (second ed.). Cham: Springer.

Shen, J., Tang, T., & Wang, L.-L. (2011). *Spectral methods* (first ed.). Berlin, Heidelberg: Springer-Verlag.

Wang, Z., Luo, X., Yau, S. S.-T., & Zhang, Z. (2020). Proper orthogonal decomposition method to nonlinear filtering problems in medium-high dimension. *IEEE Transactions on Automatic Control*, *65*(4), 1613–1624.

Xiang, X.-M., & Wang, Z. Q. (2010). Generalized Hermite spectral method and its applications to problems in unbounded domains. *SIAM Journal on Numerical Analysis*, *48*(4), 1231–1253.

Yau, S.-T., & Yau, S. S.-T. (2000). Real time solution of nonlinear filtering problem without memory I. *Mathematical Research Letters*, *7*, 671–693.

Yau, S.-T., & Yau, S. S.-T. (2008). Real time solution of the nonlinear filtering problem without memory II. *SIAM Journal on Control and Optimization*, *47*(1), 163–195.

Zakai, M. (1969). On the optimal filtering of diffusion processes. *Zeitschrift FÜ*, *11*(3), 230–243.

Zeju Sun received the B.S. degree from Department of Mathematical Sciences, Tsinghua University, Beijing, China, in 2020. He is currently pursuing the Ph.D. degree in applied mathematics with the department of Mathematical Sciences, Tsinghua University, Beijing, China.

His research interests include control theory and nonlinear filtering.

Stephen Shing-Toung Yau received the Ph.D. degree in mathematics from the State University of New York at Stony Brook, Stony Brook, NY, USA, in 1976. He was a Member of the Institute of Advanced Study at Princeton, Princeton, NJ, USA, from 1976 to 1977 and from 1981 to 1982, and a Benjamin Pierce Assistant Professor with Harvard University, Cambridge, MA, USA, from 1977 to 1980. After that, he joined the Department of Mathematics, Statistics and Computer Science (MSCS), University of Illinois at Chicago (UIC), Chicago, IL, USA, and served for more than 30 years. During 2005–2011, he became a Joint Professor with the Department of Electrical and Computer Engineering, 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, Cauchy–Riemann (CR) geometry, and singularities theory. Dr. Yau was a recipient of the Sloan Fellowship in 1980, the Guggenheim Fellowship in 2000, the AMS Fellow Award in 2013, and the AAIA Fellow in 2023. He was the General Chairperson of the IEEE International Conference on Control and Information, which was held in The Chinese University of Hong Kong, Hong Kong, in 1995. He 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. In 2005, he was entitled the UIC Distinguished Professor.