ON THE CONVERGENCE ANALYSIS OF YAU-YAU NONLINEAR FILTERING ALGORITHM: FROM A PROBABILISTIC PERSPECTIVE*

ZEJU SUN†, XIUQIONG CHEN‡, AND STEPHEN S.-T. YAU§

Abstract. At the beginning of this century, a real time solution of the nonlinear filtering problem without memory was proposed in [25, 26] by the third author and his collaborator, and it is later on referred to as Yau-Yau algorithm. During the last two decades, a great many nonlinear filtering algorithms have been put forward and studied based on this framework. In this paper, we will generalize the results in the original works and conduct a novel convergence analysis of Yau-Yau algorithm from a probabilistic perspective. Instead of considering a particular trajectory, we estimate the expectation of the approximation error, and show that commonly-used statistics of the conditional distribution (such as conditional mean and covariance matrix) can be accurately approximated with arbitrary precision by Yau-Yau algorithm, for general nonlinear filtering systems with very liberal assumptions. This novel probabilistic version of convergence analysis is more compatible with the development of modern stochastic control theory, and will provide a more valuable theoretical guidance for practical implementations of Yau-Yau algorithm.

Key words. nonlinear filtering, DMZ equation, Yau-Yau algorithm, convergence analysis, stochastic partial differential equation

MSC codes. 60G35, 93E11, 60H15, 65M12

1. Introduction.

Filtering is an important subject in the field of modern control theory, and has wide applications in various scenarios such as signal processing [3][20], weather forecast [10][5], aerospace industrial [12][23] and so on. The core objective of filtering problem is pursuing accurate estimation or prediction to the state of a given stochastic dynamical system based on a series of noisy observations [13][2]. For practical implementations, it is also necessary that the estimation or prediction to the state can be computed in a recursive and real-time manner.

In the filtering problems we consider in this paper, the evolution of state processes, as well as the noisy observations, is governed by the following system of stochastic differential equations,

$$\begin{cases} dX_t = f(X_t)dt + g(X_t)dV_t, \quad X_0 = \xi, \\ dY_t = h(X_t)dt + dW_t, \quad Y_0 = 0, \end{cases} \qquad t \in [0,T], \tag{1.1}$$

in the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t=0}^T, P)$, where $T > 0$ is a fixed termination; $X = \{X_t : 0 \le t \le T\} \subset \mathbb{R}^d$ is the state process we would like to track; $Y = \{Y_t : 0 \le t \le T\} \subset \mathbb{R}^d$ is the noisy observation to the state process X; $\{V_t : 0 \le t \le T\}$ and $\{W_t : 0 \le t \le T\}$ are mutually independent, $\mathcal{F}_t$-adapted, d-dimensional standard Brownian motions; ξ is a $\mathbb{R}^d$-valued random variable with probability density function $\sigma_0(x)$, which is independent of V_t and W_t; $f : \mathbb{R}^d \to \mathbb{R}^d$, $g : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ and $h : \mathbb{R}^d \to \mathbb{R}^d$ are sufficiently smooth vector- or matrix- valued functions.

*Corresponding author: Stephen S.-T. Yau.

Funding: This work is supported by National Natural Science Foundation of China (NSFC) under Grant 123B2020 for Zeju Sun, by NSFC under Grant 42450242 and 12201631 for Xiuqiong Chen, and by NSFC under Grant 42450242, Tsinghua University Education Foundation fund for Stephen S.-T. Yau.

†Beijing Institute of Mathematical Sciences and Applications, Beijing, P. R. China (sunzeju@bimsa.cn).

‡School of Mathematics, Renmin University of China, Beijing, P .R. China (cxq0828@ruc.edu.cn).

§Department of Mathematical Sciences, Tsinghua University, Beijing, P. R. China. (yau@uic.edu).

Mathematically, for a given test function $\varphi : \mathbb{R}^d \to \mathbb{R}$, the *optimal* estimation of $\varphi(X_t)$ based on the historical observations up to time t is the conditional expectation $E[\varphi(X_t)|\mathcal{Y}_t]$, where $\mathcal{Y}_t := \sigma\{Y_s : 0 \leq s \leq t\}$ is the σ-algebra generated by historical observations. Such estimation is '*optimal*' in the sense that,

$$E[\varphi(X_t)|\mathcal{Y}_t] = \underset{U \text{ is } \mathcal{Y}_t\text{-measurable}}{\arg\min} E[(\varphi(X_t) - U)^2]. \tag{1.2}$$

Therefore, the main task of filtering problem can be specified into finding efficient algorithms to numerically compute the conditional expectation $E[\varphi(X_t)|\mathcal{Y}_t]$, or equivalently, the conditional probability distribution $P[X_t \in \cdot|\mathcal{Y}_t]$ or the conditional probability density function (if exists).

With some regularity assumptions on the coefficients f, g, h in the system (1.1), as well as the test function φ, the evolution of $E[\varphi(X_t)|\mathcal{Y}_t]$, (without considering a normalization constant), can be described by the well-known DMZ equation, which is named after three researchers: T. E. Duncan [8], R. E. Mortensen [18] and M. Zakai [27], who derived the equation independently in the late 1960s. Furthermore, if the conditional probability measure $P[X_t \in \cdot|\mathcal{Y}_t]$ is absolutely continuous with respect to the Lebesgue measure in $\mathbb{R}^d$, and the density function, as well as its derivatives, is square-integrable, then the unnormalized conditional probability density function $\sigma(t, x)$ will satisfies the DMZ equation in a duality form.

The dual DMZ equation satisfied by the unnormalized conditional probability density function $\sigma(t, x)$ is a second-order stochastic partial differential equation, and does not possess an explicit-form solution in general. In their works [25] and [26], the third author and his collaborator propose a two-stage algorithm framework to compute the solution of the dual DMZ equation numerically in a memoryless and real-time manner. Later on, this algorithm is referred to as Yau-Yau algorithm.

The basic idea of Yau-Yau algorithm is that the heavy computational burden of numerically solving a Kolmogorov-type partial differential equation (PDE) can be done off-line, and in the meanwhile, the on-line procedure only consists of the basic computations such as multiplication by the exponential function of the observations. In the framework of Yau-Yau algorithm, various kinds of methods to solving the Kolmogorov-type PDEs, such as spectral methods [17][7], proper orthogonal decomposition [24], tensor training [16], etc, are proposed and applied in specific examples of nonlinear filtering problems. Numerical results in the previous works mentioned above show that Yau-Yau algorithm can provide accurate and real-time estimations to the state process of very general nonlinear filtering problems in low and medium-high dimensional space.

In the original works [25] and [26], the convergence analysis of Yau-Yau algorithm is conducted pathwise. For *regular* paths of observations with some boundedness conditions, it is proved that the numerical solution provided by Yau-Yau algorithm will converge to the exact solution of DMZ equation both pointwisely and in L^2-sense, as the size of time-discretization step tends to zero, while in the works on the practical implementations of Yau-Yau algorithm, such as [17] and [7], the convergence analysis mainly focuses on the capability of numerical methods to approximate the solution of Kolmogorov-type PDEs arising in Yau-Yau algorithms.

In this paper, we will revisit the convergence analysis of Yau-Yau algorithm from a probabilistic perspective. Instead of considering the convergence results pathwise, we prove that the solution of Yau-Yau algorithm will converge to the exact solution of DMZ equation in expectation, and also, after the normalization procedure, the approximated solution to the filtering problem provided by Yau-Yau algorithm will converge to the conditional expectation $E[\varphi(X_t)|\mathcal{Y}_t]$.

The advantage of this probabilistic perspective is that for a theoretically rigorous convergence analysis, instead of regularity assumptions on observation paths, we only need to make assumptions on the coefficients f, g, h of the filtering system (1.1) and the test function φ, which are verifiable off-line in advance for practitioners. In the meanwhile, as shown in the main results of this paper in Section 3, those assumptions we need are in fact quite general, and it is straightforward to check that the most commonly used test functions $\varphi(x) = x_i$ and $\varphi(x) = x_i x_j$, $x = (x_1, \cdots, x_d)^\top \in \mathbb{R}^d$, corresponding to the conditional mean and conditional covariance matrix, as well as the linear Gaussian systems (with $f(x) = Fx$, $g(x) \equiv \Gamma$, and $h(x) = Hx$, $F, H, \Gamma \in \mathbb{R}^{d\times d}$), satisfy all the assumptions.

Moreover, to the best of the authors' knowledge, most of the theoretical analysis of PDE-based filtering algorithm mainly deals with convergence results with respect to the DMZ equation. In such probabilistic perspective we consider here in this paper, however, it is natural and convenient to make a step forward and discuss the approximation capability of Yau-Yau algorithm to the *normalized* conditional expectation and conditional probability distribution. In this way, we will provide a thorough convergence analysis of the Yau-Yau algorithm for filtering problems.

The organization of this paper is as follows. Section 2 serves as preliminaries, in which we will summarize some basic concepts of filtering problems and the main procedure of Yau-Yau algorithm. The main theorems in this paper will be stated in Section 3, together with a sketch of the proofs. In the next four sections, we will provide the detailed proofs of the lemmas and theorems. We first focus on the properties of the exact solution of DMZ equation in Section 4 and Section 5, and then deal with the approximated solutions given by Yau-Yau algorithm in Section 6 and Section 7. Finally, Section 8 is a conclusion.

2. Preliminaries.

In this section, we would like to briefly summarize the theory of nonlinear filtering, including the change-of-measure approach to deriving the DMZ equation, as well as the main idea and procedures of Yau-Yau algorithm.

In the change-of-measure approach to deriving the DMZ equation corresponding to the filtering system (1.1), we first introduce a series of reference probability measures $\{\tilde{P}_t : 0 \le t \le T\}$, absolutely continuous to the original probability measure P with Radon derivatives given by

$$Z_t \triangleq \left.\frac{d\tilde{P}_t}{dP}\right|_{\mathcal{F}_t} = \exp\left(- \int_0^t h(X_s)^\top dW_s - \frac{1}{2}\int_0^t |h(X_s)|^2 ds \right), \ t \in [0,T]. \tag{2.1}$$

According to Girsanov's theorem, as long as the process $\{Z_t : 0 \le t \le T\}$ defined in (2.1) is a martingale, then under the reference probability measure $\tilde{P}_T$, the observation process $\{Y_t : 0 \le t \le T\}$ is a standard Brownian motion which is independent of the state process X.

We also introduce the process $\{\tilde{Z}_t : 0 \le t \le T\}$, $\tilde{Z}_t = Z_t^{-1}$, to be the inverse of Z_t, which is also a Radon derivative and can be expressed by the stochastic integral with respect to Y as follows:

$$\tilde{Z}_t = Z_t^{-1} = \left.\frac{dP}{d\tilde{P}_t}\right|_{\mathcal{F}_t} = \exp\left(\int_0^t h(X_s)^\top dY_s - \frac{1}{2}\int_0^t |h(X_s)|^2 ds \right), \ t \in [0,T]. \tag{2.2}$$

Therefore, for any $\mathcal{F}_t$-measurable, integrable random variable $U \in \mathcal{F}_t$, its expectation with respect to measure P can be computed by $E[U] = \tilde{E}\left[\tilde{Z}_t U\right]$, where $\tilde{E}$ means the expectation is taken under the probability measure $\tilde{P}_T$.

The following Kallianpur-Striebel formula allows us to express and calculate the solution of filtering problem, $E[\varphi(X_t)|\mathcal{Y}_t]$, by a ratio of conditional expectations under $\tilde{P}_T$:

$$E[\varphi(X_t)|\mathcal{Y}_t] = \frac{\tilde{E}\left[\tilde{Z}_t\varphi(X_t)|\mathcal{Y}_t\right]}{\tilde{E}\left[\tilde{Z}_t|\mathcal{Y}_t\right]},\ t \in [0,T]. \tag{2.3}$$

Since the denominator $\tilde{E}\left[\tilde{Z}_t|\mathcal{Y}_t\right]$ in (2.3) is independent of the test function φ, people often refer to the numerator, $\tilde{E}\left[\tilde{Z}_t\varphi(X_t)|\mathcal{Y}_t\right]$, as the unnormalized conditional expectation of $\varphi(X_t)$. The corresponding measure-valued stochastic process $\{\rho_t : 0 \le t \le T\}$ defined by $\rho_t(A) := \tilde{E}\left[\tilde{Z}_t 1_A|\mathcal{Y}_t\right], \forall\ A \in \mathcal{F}_t,\ t \in [0,T]$, is also referred to as unnormalized conditional probability measure, and we also denote the unnormalized conditional expectation by

$$\rho_t(\varphi) := \tilde{E}\left[\tilde{Z}_t\varphi(X_t)|\mathcal{Y}_t\right],\ \varphi \text{ is a test function},\ t \in [0,T], \tag{2.4}$$

With sufficient regularity assumptions on the coefficients f, g, h and test function φ, the evolution of $\rho_t(\varphi)$ is governed by the following well-known DMZ equation:

$$\rho_t(\varphi) = \rho_0(\varphi) + \int_0^t \rho_s(\mathcal{L}\varphi)ds + \sum_{j=1}^d \int_0^t \rho_s(h_j\varphi)dY_s^j,\ t \in [0,T]. \tag{2.5}$$

where

$$\mathcal{L} = \frac{1}{2}\sum_{i,j=1}^d a^{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^d f_i(x)\frac{\partial}{\partial x_i} \tag{2.6}$$

is a second-order elliptic operator with $a(x) := (a^{ij}(x))_{1\le i,j\le d} = g(x)g(x)^\top$.

If the stochastic measures ρ_t, $t \in [0,T]$, are almost surely absolutely continuous to the Lebesgue measure in $\mathbb{R}^d$, and the density functions (or the Radon derivatives) $\sigma(t,x)$, as well as the derivatives of $\sigma(t,x)$, are square-integrable, then $\sigma(t,x)$ is the solution to the following dual DMZ equation:

$$\begin{cases} d\sigma(t,x) = \mathcal{L}^*\sigma(t,x)dt + \sum_{j=1}^d h_j(x)\sigma(t,x)dY_t^j,\ t \in [0,T], \\ \sigma(0,x) = \sigma_0(x), \end{cases} \tag{2.7}$$

which is a second-order stochastic partial differential equation with

$$\mathcal{L}^*(\star) = \frac{1}{2}\sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j}(a^{ij}\star) - \sum_{i=1}^d \frac{\partial}{\partial x_i}(f_i\star). \tag{2.8}$$

the adjoint operator of $\mathcal{L}$.

In this case, the (normalized) conditional expectation $E\left[\varphi(X_t)|\mathcal{Y}_t\right]$ can be calculated by

$$E\left[\varphi(X_t)|\mathcal{Y}_t\right] = \frac{\tilde{E}\left[\tilde{Z}_t\varphi(X_t)|\mathcal{Y}_t\right]}{\tilde{E}\left[\tilde{Z}_t|\mathcal{Y}_t\right]} = \frac{\int_{\mathbb{R}^d}\varphi(x)\sigma(t,x)dx}{\int_{\mathbb{R}^d}\sigma(t,x)dx}. \tag{2.9}$$

Because the solution of (2.7) does not have a closed form for general nonlinear filtering systems, efficient numerical methods must be proposed, so that we can get a good approximation to the conditional expectation $E[\varphi(X_t)|\mathcal{Y}_t]$ through the equation (2.9).

At the beginning of this century, the third author and his collaborator proposed a two-stage algorithm to numerically solve the DMZ equation (2.7) in a memoryless and real-time manner, which is often referred to as Yau-Yau algorithm. Here, we would like to briefly introduce the basic idea and main procedure of this algorithm.

Firstly, if we consider the exponential transformation

$$w(t,x) := \exp\left(-h^\top(x)Y_t\right)\sigma(t,x),\ \ t\in[0,T], \tag{2.10}$$

then the function $w(t,x)$ satisfies the robust DMZ equation

$$\frac{\partial w}{\partial t} = \frac{1}{2}\sum_{i,j=1}^{d} a^{ij}(x)\frac{\partial^2 w}{\partial x_i \partial x_j} + \sum_{i=1}^{d} F_i(t,x)\frac{\partial w}{\partial x_i} + J(t,x)w(t,x), \tag{2.11}$$

where the stochastic differential terms in the original DMZ equation (2.7) are eliminated and

$$F_i(t,x) = \sum_{j=1}^{d}\left(\frac{\partial a^{ij}}{\partial x_j} + a^{ij}\sum_{k=1}^{d} Y_t^k \frac{\partial h_k}{\partial x_j}\right) - f_i(x), \quad i = 1,\cdots,d,$$

$$\begin{aligned} J(t,x) =& \frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2 a^{ij}}{\partial x_i\partial x_j} + \sum_{i,j,k=1}^{d} Y_t^k\frac{\partial h_k}{\partial x_j}\frac{\partial a^{ij}}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{d} a^{ij}\left(\sum_{k=1}^{d} Y_t^k\frac{\partial^2 h_k}{\partial x_i\partial x_j}\right.\\ &\left.+\sum_{k=1}^{d}\sum_{l=1}^{d} Y_t^k Y_t^l\frac{\partial h_k}{\partial x_i}\frac{\partial h_l}{\partial x_j}\right) - \sum_{i=1}^{d}\frac{\partial f_i}{\partial x_i} - \sum_{i,j=1}^{d} Y_t^j\frac{\partial h_j}{\partial x_i}f_i(x) - \frac{1}{2}|h|^2 \end{aligned} \tag{2.12}$$

are stochastic functions that depend on the specific value of observation Y_t at time t.

Instead of solving equation (2.7) directly, we will mainly focus on the robust DMZ equation (2.11), especially the corresponding initial-boundary value (IBV) problems in a closed ball $B_R := \{x\in\mathbb{R}^d : |x| \le R\}$, with a given radius $R > 0$.

$$\begin{cases} \dfrac{\partial u_R}{\partial t} = \dfrac{1}{2}\displaystyle\sum_{i,j=1}^{d} a^{ij}(x)\frac{\partial^2 u_R}{\partial x_i\partial x_j} + \sum_{i=1}^{d} F_i(t,x)\frac{\partial u_R}{\partial x_i} + J(t,x)u_R(t,x),\ \ t\in[0,T],\\ u_R(0,x) = \sigma_0(x)\cdot\mathscr{S}_R(x),\ \ x\in B_R,\ \ u_R(t,x) = 0,\ \ (t,x)\in[\tau_{k-1},\tau_k]\times\partial B_R. \end{cases} \tag{2.13}$$

where $\mathscr{S}_R(x)$ is a C^∞ function supported in B_R and satisfies $\mathscr{S}_R(x)\equiv 1$, for $|x| \le R - \frac{1}{R}$, $\mathscr{S}_R(x)\equiv 0$ for $|x|\ge R$, and $0\le\mathscr{S}_R(x)\le 1$ for $R-\frac{1}{R}\le|x|\le R$, so that the initial value is compatible with the boundary condition in (2.13). And from now on, we would like to drop the subscript, R, in the notation $u_R(t,x)$ for the simplicity of notations, and use $u(t,x)$ to denote the solution to the IBV problem (2.13).

Let $0=\tau_0<\tau_1<\cdots<\tau_K=T$ be a uniform partition of the time interval $[0,T]$, with $\tau_k-\tau_{k-1}=\delta=\frac{T}{K}$, $k=1,\cdots,K$. On each time interval $[\tau_{k-1},\tau_k]$, consider the

IBV problem of the following parabolic equation

$$\begin{cases} \dfrac{\partial u_k}{\partial t} = \dfrac{1}{2}\displaystyle\sum_{i,j=1}^{d} a^{ij}(x)\dfrac{\partial^2 u_k}{\partial x_i \partial x_j} + \sum_{i=1}^{d} F_i(\tau_{k-1},x)\dfrac{\partial u_k}{\partial x_i} + J(\tau_{k-1},x)u_k(t,x), \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (t,x) \in (\tau_{k-1},\tau_k] \times B_R, \\ u_k(\tau_{k-1},x) = u_{k-1}(\tau_{k-1},x),\ x \in B_R,\ u_k(t,x) = 0,\ (t,x) \in [\tau_{k-1},\tau_k] \times \partial B_R, \end{cases} \tag{2.14}$$

with the value of coefficients $F(t,x)$ and $J(t,x)$ frozen at the left point $t = \tau_{k-1}$ and initial value $u_0(\tau_0,x) := \sigma_0(x)$.

With another exponential transformation given by

$$\tilde{u}_k(t,x) = \exp\left(h^\top(x)Y_{\tau_{k-1}}\right)u_k(t,x),\ t \in [\tau_{k-1},\tau_k], \tag{2.15}$$

the newly-constructed function $\tilde{u}_k$ satisfies

$$\frac{\partial \tilde{u}_k}{\partial t} = \frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2}{\partial x_i \partial x_j}(a^{ij}\tilde{u}_k(t,x)) - \sum_{i=1}^{d}\frac{\partial}{\partial x_i}(f_i\tilde{u}_k(t,x)) - \frac{1}{2}|h|^2\tilde{u}_k(t,x), \tag{2.16}$$

and $\tilde{u}_k(\tau_{k-1},x) = \exp\left(h^\top(x)Y_{\tau_{k-1}}\right)u_k(\tau_{k-1},x)$. After the two exponential transformations (2.10) and (2.15), the function we would like to use to approximate the unnormalized conditional probability density function $\sigma(\tau_k,x)$ at time $t = \tau_k$ is given by

$$\sigma(\tau_k,x) \approx \exp\left(h^\top(x)(Y_{\tau_k} - Y_{\tau_{k-1}})\right)\tilde{u}_k(\tau_k,x) = \tilde{u}_{k+1}(\tau_k,x),\ k = 1,\cdots,K. \tag{2.17}$$

and the value for approximating the conditional expectation $E[\varphi(X_t)|\mathcal{Y}_t]$ is given by

$$E[\varphi(X_t)|\mathcal{Y}_t] \approx \frac{\int_{B_R}\varphi(x)\tilde{u}_{k+1}(\tau_k,x)dx}{\int_{B_R}\tilde{u}_{k+1}(\tau_k,x)dx},\ k = 1,\cdots,K. \tag{2.18}$$

The main idea of the Yau-Yau algorithm is that the problem of solving the DMZ equation satisfied by the unnormalized probability density function $\sigma(t,x)$ can be separated into two parts. The computationally expensive part of solving the (IBV) problems of parabolic equation (2.16) can be studied off-line, because it is a deterministic Kolmogorov-type PDE which is independent of observations. The framework of Yau-Yau algorithm is shown in Algorithm 2.1.

In the next section, we will give a mathematically rigorous interpretation of the approximation results (2.17) and (2.18) from a probabilistic perspective. In particular, we only need assumptions on the test function and the coefficients of the filtering systems to derive the convergence result. These assumptions are also easy to verify off-line before the observations come, and therefore, this convergence analysis will provide a guidance for practitioners to determine the parameters in the implementations of Yau-Yau algorithm for practical use.

3. Main Results. In this section, we would like to state the main result in this paper and also provide a sketch of the proof.

Firstly, besides the smoothness and regularity requirements which guarantee the existence of conditional expectation and the existence of the solution to the DMZ

Algorithm 2.1 The Two-Stage Framework of Yau-Yau Algorithm

```
Initialization: Input the terminal time T, the radius R of closed ball B_R, the number of time-discretization steps K, the initial distribution of state process σ_0(x), the test function φ(x), and the initial observation Y_0 = 0. Let δ = T/K be the time-discretization step size and {0 = τ_0 < τ_1 < ··· < τ_K = T} be a uniform partition of [0,T] with τ_k − τ_{k−1} = δ. Initialize ũ_1(0,x) = σ_0(x).
Off-Line Algorithm: Solve the IBV problem of Kolmogorov-type partial differential equation (2.16) in closed ball B_R, and determine or approximate the corresponding semi-group {S_t : t ∈ [0,T]}.
On-Line Algorithm:
for k = 1 to K do
    Obtain ũ_k(τ_k, x) from Off-Line Algorithm ũ_k(τ_k, x) = S_{τ_k − τ_{k−1}} ũ_k(τ_{k−1}, x).
    Renew the initial value of the partial differential equation satisfied by ũ_{k+1}(x,t),
    ũ_{k+1}(τ_k, x) = exp[h^⊤(x)(Y_{τ_k} − Y_{τ_{k−1}})] ũ_k(τ_k, x).
    Compute the approximated conditional expectation:
        ∫_{B_R} φ(x) ũ_{k+1}(τ_k, x) dx / ∫_{B_R} ũ_{k+1}(τ_k, x) dx.
end for
```

equation, let us further introduce four particular assumptions on the coefficients of the system, the initial distribution and the test function.

For the state equation in the filtering system (1.1), the drift term $f : \mathbb{R}^d \to \mathbb{R}^d$ is assumed to be Lipschitz, and the the diffusion term $g : \mathbb{R}^d \to \mathbb{R}^{d\times d}$, together with $a(x) = g(x)g(x)^\top$, is assumed to have bounded partial derivatives up to second order, i.e., $\exists\ L > 0,\ s.t.\ |f(x) - f(y)| \leq L|x - y|$, such that for all $x, y \in \mathbb{R}^d$ and $i, j, k, l = 1, \cdots, d$,

$$\textbf{(A1)}\text{: } |f(x) - f(y)| \leq L|x - y|,\ |a(x)| \leq L,\ \left|\frac{\partial a^{ij}(x)}{\partial x_k}\right| \leq L,\ \left|\frac{\partial^2 a^{ij}(x)}{\partial x_k \partial x_l}\right| \leq L.$$

Assumption **(A1)** guarantees that the state equation of (1.1) has a strong solution in $[0, T]$, and especially, the state equation for linear filter satisfies this assumption.

Also, in order to conduct energy estimations for the (stochastic) partial differential equations, we would like to assume that the diffusion term in the state equation is nondegenerate, in the sense that For each $x \in \mathbb{R}^d$, there exists a continuous function $\lambda(x) > 0$, such that

$$\textbf{(A2):}\quad \sum_{i,j=1}^{d} a^{ij}(x)\zeta_i\zeta_j \geq \lambda(x)|\zeta|^2,\ \forall\ \zeta = (\zeta_1, \cdots, \zeta_d)^\top \in \mathbb{R}^d.$$

For the initial distribution σ_0, we would like to assume that it is smooth enough and possesses finite high-order moments:

$$\textbf{(A3):}\ \int_{\mathbb{R}^d} |x|^{2n}\sigma_0(x)dx < \infty,\ \forall\ n \in \mathbb{N}.$$

Assumption **(A3)** is satisfied by commonly-used light-tailed distributions such as Gaussian distributions. In fact, in the following convergence analysis, we only require Assumption **(A3)** to hold for *sufficiently* large $n \geq 1$, rather than for all $n \in \mathbb{N}$.

Finally, it is assumed that the test function $\varphi : \mathbb{R}^d \to \mathbb{R}$ is at most polynomial growth:

(A4): $\exists\, L > 0,\ m \in \mathbb{N},\ s.t.\ |\varphi(x)| \leq L(1 + |x|^{2m}), \quad \forall\, x \in \mathbb{R}^d,$

which is satisfied by most of the commonly-used test functions, such as those correspond to the conditional mean and covariance matrix.

Based on the above assumptions **(A1)** to **(A4)**, the main result in this paper is stated as follows:

THEOREM 3.1. *Fix a terminal time $T > 0$ and the filtering system* (1.1) *with smooth coefficients. If Assumptions **(A1)** to **(A4)** hold, then for every $\epsilon > 0$, there exists $R > 0$, and $\delta > 0$, such that we can conduct the Yau-Yau algorithm in the closed ball $B_R = \{x \in \mathbb{R}^d : |x| \leq R\}$ and the uniform partition of $[0, T]$: $0 = \tau_0 < \tau_1 < \cdots < \tau_K = T$ with $\delta = \tau_k - \tau_{k-1}$ for $k = 1, \cdots, K$, and the numerical solution $\{\tilde{u}_{k+1}(\tau_k, x) : k = 1, \cdots, K\}$ approximates the exact solution of the filtering problems at each time step $t = \tau_k$ well in the sense of mathematical expectation, i.e.,*

$$E\left|E[\varphi(X_{\tau_k})|\mathcal{Y}_{\tau_k}] - \frac{\int_{B_R} \varphi(x)\tilde{u}_{k+1}(\tau_k, x)dx}{\int_{B_R} \tilde{u}_{k+1}(\tau_k, x)dx}\right| < \epsilon,\ \forall\ 1 \leq k \leq K. \tag{3.1}$$

Remark 3.2. In the statement of Theorem 3.1, for the sake of theoretical rigor, the radius R of the closed ball B_R, in which the Yau-Yau algorithm is conducted, is seemingly to be dependent on the terminal time T. This is because the system (1.1) with Assumptions **(A1)** to **(A4)** we consider here is the most general nonlinear filtering system, without stability conditions such as ergodicity or compactness of trajectories of the conditional expectations.

In the meanwhile, for general filtering systems without further stability assumptions, another implementation alternative is the "moving-window" technique, as introduced in [17]. The basic idea is to fix the radius R while translating the center x_0 of the closed ball from the original point, such that the major part of the conditional density will remain in the closed ball $B_R(x_0) := \{x \in \mathbb{R}^d : |x - x_0| \leq R\}$. A successful numerical implementation of Yau-Yau algorithm with a fixed radius R can be found in [7].

Here in this section, we provide a sketch of the proof of Theorem 3.1, in which the main idea of the proof is illustrated. The detailed proofs of those key estimations here will be given in order in the next four sections.

A Sketch of the Proof of Theorem 3.1. According to the properties of conditional expectations, the expectation of the approximation error of Yau-Yau algorithm can be estimated as follows:

$$\begin{aligned}
&E\left|E[\varphi(X_{\tau_k})|\mathcal{Y}_{\tau_k}] - \frac{\int_{B_R} \varphi(x)\tilde{u}_{k+1}(\tau_k, x)dx}{\int_{B_R} \tilde{u}_{k+1}(\tau_k, x)dx}\right| \\
&= \tilde{E}\left[\tilde{Z}_{\tau_k}\left|E[\varphi(X_{\tau_k})|\mathcal{Y}_{\tau_k}] - \frac{\int_{B_R} \varphi(x)\tilde{u}_{k+1}(\tau_k, x)dx}{\int_{B_R} \tilde{u}_{k+1}(\tau_k, x)dx}\right|\right] \\
&= \tilde{E}\left[\tilde{E}[\tilde{Z}_{\tau_k}|\mathcal{Y}_{\tau_k}]\left|\frac{\tilde{E}[\varphi(X_{\tau_k})\tilde{Z}_{\tau_k}|\mathcal{Y}_{\tau_k}]}{\tilde{E}[\tilde{Z}_{\tau_k}|\mathcal{Y}_{\tau_k}]} - \frac{\int_{B_R} \varphi(x)\tilde{u}_{k+1}(\tau_k, x)dx}{\int_{B_R} \tilde{u}_{k+1}(\tau_k, x)dx}\right|\right] \\
&\leq \tilde{E}\left[\tilde{E}[\tilde{Z}_{\tau_k}|\mathcal{Y}_{\tau_k}]\left|\frac{\tilde{E}[\varphi(X_{\tau_k})\tilde{Z}_{\tau_k}|\mathcal{Y}_{\tau_k}]}{\tilde{E}[\tilde{Z}_{\tau_k}|\mathcal{Y}_{\tau_k}]} - \frac{\int_{B_R} \varphi(x)\tilde{u}_{k+1}(\tau_k, x)dx}{\tilde{E}[\tilde{Z}_{\tau_k}|\mathcal{Y}_{\tau_k}]}\right|\right]
\end{aligned}$$

$$
\begin{aligned}
&+\tilde{E}\left[\tilde{E}[\tilde{Z}_{\tau_k}|\mathcal{Y}_{\tau_k}]\left|\frac{\int_{B_R}\varphi(x)\tilde{u}_{k+1}(\tau_k,x)dx}{\tilde{E}[\tilde{Z}_{\tau_k}|\mathcal{Y}_{\tau_k}]}-\frac{\int_{B_R}\varphi(x)\tilde{u}_{k+1}(\tau_k,x)dx}{\int_{B_R}\tilde{u}_{k+1}(\tau_k,x)dx}\right|\right]\\
&\le \tilde{E}\left[\left|\int_{\mathbb{R}^d}\varphi(x)\sigma(\tau_k,x)dx-\int_{B_R}\varphi(x)\tilde{u}_{k+1}(\tau_k,x)dx\right|\right]\\
&\quad+\tilde{E}\left[\frac{\int_{B_R}|\varphi(x)|\tilde{u}_{k+1}(\tau_k,x)dx}{\int_{B_R}\tilde{u}_{k+1}(\tau_k,x)dx}\left|\int_{\mathbb{R}^d}\sigma(\tau_k,x)dx-\int_{B_R}\tilde{u}_{k+1}(\tau_k,x)dx\right|\right]\\
&\triangleq I_1+I_2.
\end{aligned}
$$

where we use the fact that $\tilde{u}_{k+1}(\tau_k,x)$ is $\mathcal{Y}_{\tau_k}$-measurable and for integrable, $\mathcal{Y}_{\tau_k}$-measurable random variable V, $\tilde{E}[\tilde{Z}_{\tau_k}V]=\tilde{E}[\tilde{E}[\tilde{Z}_{\tau_k}|\mathcal{Y}_{\tau_k}]V]$.

Therefore, the remaining task for us is to estimate the two error terms I_1 and I_2, and to show that I_1 and I_2 can be arbitrarily small with sufficiently large $R>0$ and sufficiently small $\delta>0$.

Firstly, for the estimation of I_1, we would like to utilize an intermediate function $\sigma_R(t,x)$, $(t,x)\in[0,T]\times B_R$, which is the solution of IBV problem of the DMZ equation (2.7) and will be introduced in (4.3) in Section 4. And we have

$$
\begin{aligned}
I_1 \le&\tilde{E}\int_{|x|\ge R}|\varphi(x)|\sigma(\tau_k,x)dx+\tilde{E}\left[\left|\int_{B_R}\varphi(x)\sigma(\tau_k,x)dx-\int_{B_R}\varphi(x)\sigma_R(\tau_k,x)dx\right|\right]\\
&+\tilde{E}\left[\left|\int_{B_R}\varphi(x)\sigma_R(\tau_k,x)dx-\int_{B_R}\varphi(x)\tilde{u}_{k+1}(\tau_k,x)dx\right|\right]\\
\le&\tilde{E}\int_{|x|\ge R}|\varphi(x)|\sigma(\tau_k,x)dx+L(1+R^{2m})\tilde{E}\int_{B_R}|\sigma(\tau_k,x)-\sigma_R(\tau_k,x)|dx\\
&+L(1+R^{2m})\tilde{E}\int_{B_R}|\sigma_R(\tau_k,x)-\tilde{u}_{k+1}(\tau_k,x)|dx.
\end{aligned}\tag{3.2}
$$

For the estimation of I_2, since in the closed ball B_R, $|\varphi(x)|\le L(1+R^{2m})$,

$$
\frac{\int_{B_R}|\varphi(x)|\tilde{u}_{k+1}(\tau_k,x)dx}{\int_{B_R}\tilde{u}_{k+1}(\tau_k,x)dx}\le L(1+R^{2m})\frac{\int_{B_R}\tilde{u}_{k+1}(\tau_k,x)dx}{\int_{B_R}\tilde{u}_{k+1}(\tau_k,x)dx}=L(1+R^{m}).\tag{3.3}
$$

and thus,

$$
\begin{aligned}
I_2 &\le L(1+R^{2m})\tilde{E}\left[\left|\int_{\mathbb{R}^d}\sigma(\tau_k,x)dx-\int_{B_R}\tilde{u}_{k+1}(\tau_k,x)dx\right|\right]\\
&\le L(1+R^{2m})\left(\tilde{E}\int_{|x|\ge R}\sigma(\tau_k,x)dx+\tilde{E}\int_{B_R}|\sigma(\tau_k,x)-\sigma_R(\tau_k,x)|dx\right)\\
&\quad+L(1+R^{2m})\tilde{E}\int_{B_R}|\sigma_R(\tau_k,x)-\tilde{u}_{k+1}(\tau_k,x)|dx.
\end{aligned}\tag{3.4}
$$

Combining (3.2) and (3.4), we have

$$\begin{aligned} &E\left|E[\varphi(X_{\tau_k})|\mathcal{Y}_{\tau_k}] - \frac{\int_{B_R}\varphi(x)\tilde{u}_{k+1}(\tau_k,x)dx}{\int_{B_R}\tilde{u}_{k+1}(\tau_k,x)dx}\right| \leq I_1 + I_2 \\ &\leq \tilde{E}\int_{|x|\geq R}|\varphi(x)|\sigma(\tau_k,x)dx + L(1+R^m)\tilde{E}\int_{|x|\geq R}\sigma(\tau_k,x)dx \\ &+ 2L(1+R^{2m})\tilde{E}\int_{B_R}|\sigma(\tau_k,x)-\sigma_R(\tau_k,x)|dx \\ &+ 2L(1+R^{2m})\tilde{E}\int_{B_R}|\sigma_R(\tau_k,x)-\tilde{u}_{k+1}(\tau_k,x)|dx. \end{aligned} \tag{3.5}$$

According to Theorem 4.1 in Section 4, for every $n \in \mathbb{N}$, there exists $C_1 > 0$, which depends on d, m, n, L, T, such that

$$\begin{aligned} &\tilde{E}\int_{|x|\geq R}\sigma(\tau_k,x)dx \leq \frac{C_1}{1+R^{2n}}\int_{\mathbb{R}^d}|x|^{2n}\sigma_0(x)dx \\ &\tilde{E}\int_{|x|\geq R}|\varphi(x)|\sigma(\tau_k,x)dx \leq \frac{C_1}{1+R^{2n}}\int_{\mathbb{R}^d}|x|^{2(m+n)}\sigma_0(x)dx \end{aligned} \tag{3.6}$$

Therefore, for every $\epsilon > 0$, with Assumption **(A3)** for the initial distribution σ_0, as long as we take $n > m$, there exists $R_1 > 0$, such that

$$\begin{aligned} &\tilde{E}\int_{|x|\geq R_1}|\varphi(x)|\sigma(\tau_k,x)dx + L(1+R_1^m)\tilde{E}\int_{|x|\geq R_1}\sigma(\tau_k,x)dx \\ &\leq \frac{C_1(1+R_1^{2m})}{1+R_1^{2n}}\int_{\mathbb{R}^d}|x|^{2n}\sigma_0(x)dx + \frac{C_1}{1+R_1^{2n}}\int_{\mathbb{R}^d}|x|^{2(m+n)}\sigma_0(x)dx < \frac{\epsilon}{3} \end{aligned} \tag{3.7}$$

According to Theorem 5.1 in Section 5, there exists $C_2 > 0$, which depends on d, n, L, T, such that

$$\tilde{E}\int_{B_R}|\sigma(\tau_k,x)-\sigma_R(\tau_k,x)|dx \leq \frac{C_2}{1+R^{2n}}. \tag{3.8}$$

Therefore, as long as $n > m$, there exists $R_2 > 0$, such that

$$2L(1+R_2^{2m})\tilde{E}\int_{B_{R_2}}|\sigma(\tau_k,x)-\sigma_{R_2}(\tau_k,x)|dx \leq \frac{2C_2L(1+R_2^{2m})}{1+R_2^{2n}} < \frac{\epsilon}{3} \tag{3.9}$$

Let us choose $R = \max\{R_1, R_2\}$, and for this particular R, according to Theorem 7.1 in Section 7, there exists a time step $\delta > 0$, such that

$$\tilde{E}\int_{B_R}|\sigma_R(\tau_k,x)-\tilde{u}_{k+1}(\tau_k,x)|dx < \frac{\epsilon}{6L(1+R^{2m})}. \tag{3.10}$$

and thus,

$$2L(1+R^{2m})\tilde{E}\int_{B_R}|\sigma_R(\tau_k,x)-\tilde{u}_{k+1}(\tau_k,x)|dx \leq \frac{2L(1+R^{2m})\epsilon}{6L(1+R^{2m})} < \frac{\epsilon}{3}. \tag{3.11}$$

Take (3.7), (3.9) and (3.11) back to (3.5), and we obtain the desired result, that is, we have found $R > 0$ and $\delta > 0$, such that

$$E\left|E[\varphi(X_{\tau_k})|\mathcal{Y}_{\tau_k}] - \frac{\int_{B_R}\varphi(x)\tilde{u}_{k+1}(\tau_k,x)dx}{\int_{B_R}\tilde{u}_{k+1}(\tau_k,x)dx}\right| < \epsilon. \tag{3.12}$$

∎

4. Estimation of the density outside the ball B_R. In this section, we will provide an estimation of the value of the unnormalized conditional probability density $\sigma(t,x)$ outside a ball $B_R \subset \mathbb{R}^d$, with $R \gg 1$ large enough.

Especially, we will show that almost all the mass of $\sigma(t,x)$ is contained in the closed ball B_R, and the estimations (3.6) in the proof of Theorem 3.1 in Section 3 holds with Assumptions **(A1)** to **(A4)**.

THEOREM 4.1. *With Assumptions **(A1)** to **(A4)**, there exists a constant $C>0$ which only depends on T, L, d, m and n, such that*

$$\sup_{0\le t\le T} \tilde{E}\int_{|x|\ge R} \sigma(t,x)dx \le \frac{C}{1+R^{2n}}\int_{\mathbb{R}^d}|x|^{2n}\sigma_0(x)dx, \tag{4.1}$$

and

$$\sup_{0\le t\le T} \tilde{E}\int_{|x|\ge R} |\varphi(x)|\sigma(t,x)dx \le \frac{C}{1+R^{2n}}\int_{\mathbb{R}^d}|x|^{2(m+n)}\sigma_0(x)dx \tag{4.2}$$

holds for all $R>0$.

Proof of Theorem 4.1. We first consider the following IBV problem on the ball B_R:

$$\begin{cases} d\sigma_R(t,x) = \mathcal{L}\sigma_R(t,x)dt + \displaystyle\sum_{j=1}^d h_j(x)\sigma_R(t,x)dY_t^j, \ (t,x)\in(0,T]\times B_R; \\ \sigma_R(0,x) = \sigma_{0,R}(x) \triangleq \sigma_0(x)\cdot\mathscr{S}_R(x), \ x\in B_R; \ \sigma_R(t,x)=0, \ (t,x)\in[0,T]\times\partial B_R. \end{cases} \tag{4.3}$$

where $\mathscr{S}_R(x)$ is the C^∞ function, such that the initial value is compatible with the boundary conditions.

Let $\psi(x)=\log\left(1+|x|^{2n}\right)$ and define $\Phi(t)=\int_{B_R} e^{\psi(x)}\sigma_R(t,x)dx$. Then, according to the IBV problem (4.3) satisfied by the function $\sigma_R(t,x)$, we have

$$\begin{aligned} d\Phi(t) = &\left[\frac{1}{2}\sum_{i,j=1}^d \int_{B_R} \frac{\partial^2}{\partial x_i\partial x_j}\left[\left(a^{ij}(x)\right)\sigma_R(t,x)\right]e^{\psi(x)}dx \right.\\ &\left. -\sum_{i=1}^d\int_{B_R}\frac{\partial}{\partial x_i}\left(f_i(x)\sigma_R(t,x)\right)e^{\psi(x)}dx\right]dt + \sum_{j=1}^d\left[\int_{B_R} e^{\psi(x)}h_j(x)\sigma_R(t,x)dx\right]dY_t^j \\ \triangleq &\left[I_1(t)-I_2(t)\right]dt + \sum_{j=1}^d I_{3,j}(t)dY_t^j. \end{aligned} \tag{4.4}$$

By the Gauss-Green formula, we have

$$\begin{aligned} I_1(t) = &\frac{1}{2}\sum_{i,j=1}^d\left[\int_{B_R} e^{\psi(x)}\left(\frac{\partial\psi(x)}{\partial x_i}\frac{\partial\psi(x)}{\partial x_j}+\frac{\partial^2\psi(x)}{\partial x_i\partial x_j}\right)a^{ij}(x)\sigma_R(t,x)dx\right.\\ &\left.-\int_{B_R}\frac{\partial}{\partial x_j}\left(e^{\psi(x)}\frac{\partial\psi(x)}{\partial x_i}a^{ij}\sigma_R\right)dx + \int_{B_R}\frac{\partial}{\partial x_i}\left(e^{\psi(x)}\frac{\partial}{\partial x_j}\left[a^{ij}\sigma_R\right]\right)dx\right] \\ = &\frac{1}{2}\sum_{i,j=1}^d\int_{B_R} e^{\psi(x)}\left(\frac{\partial\psi}{\partial x_i}\frac{\partial\psi}{\partial x_j}+\frac{\partial^2\psi}{\partial x_i\partial x_j}\right)a^{ij}\sigma_R dx - \int_{\partial B_R}(\vec{\mathfrak{M}}_1-\vec{\mathfrak{M}}_2)\cdot\vec{\mathfrak{n}}dS, \end{aligned} \tag{4.5}$$

where $\vec{\mathfrak{n}}$ is the unit outward normal vector of ∂B_R, dS denotes the measure on ∂B_R,

$$\vec{\mathfrak{M}}_i(t,x) = \left(\mathfrak{M}_{i,1}(t,x), \cdots, \mathfrak{M}_{i,d}(t,x)\right), \quad i = 1, 2, \tag{4.6}$$

and

$$\mathfrak{M}_{1,j} = \frac{1}{2}\sum_{i=1}^{d} e^{\psi(x)} \frac{\partial \psi}{\partial x_i} a^{ij}\sigma_R, \; \mathfrak{M}_{2,i} = \frac{1}{2}\sum_{j=1}^{d} e^{\psi(x)} \frac{\partial}{\partial x_j}\left[a^{ij}\sigma_R\right], \quad i,j = 1,\cdots,d. \tag{4.7}$$

Since $\sigma_R(t,x) \equiv 0$, $\forall (t,x) \in [0,T]\times\partial B_R$, $\mathfrak{M}_{1,j}(t,x) \equiv 0$ on $[0,T]\times\partial B_R$ and $\int_{\partial B_R} \vec{\mathfrak{M}}_1(t,x)\cdot\vec{\mathfrak{n}}dS = 0$. Moreover, we have

$$\nabla\sigma_R = \left(\frac{\partial \sigma_R}{\partial x_1}, \cdots, \frac{\partial \sigma_R}{\partial x_d}\right) = -c\,\vec{\mathfrak{n}}, \quad \text{on } \partial B_R, \tag{4.8}$$

where $c(x) > 0$ is a continuous function on ∂B_R, because $\sigma_R \geq 0$ and $\sigma_R|_{\partial B_R} \equiv 0$.

Therefore,

$$\begin{aligned} \vec{\mathfrak{M}}_2(t,x)\cdot\vec{\mathfrak{n}} =& -e^{\psi(x)}\frac{1}{2}\sum_{i,j=1}^{d} a^{ij}(x)\frac{\partial\sigma_R}{\partial x_j}\frac{\partial\sigma_R}{\partial x_i} - e^{\psi(x)}\sigma_R\frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial}{\partial x_j}a^{ij}(x)\frac{\partial\sigma_R}{\partial x_i} \\ =& -e^{\psi(x)}\frac{1}{2}\sum_{i,j=1}^{d} a^{ij}(x)\frac{\partial\sigma_R}{\partial x_j}\frac{\partial\sigma_R}{\partial x_i} \leq 0, \quad \text{on } \partial B_R, \end{aligned} \tag{4.9}$$

where the last inequality holds because $a(x) = g(x)g(x)^\top$ is positive semi-definite. Thus,

$$I_1(t) \leq \frac{1}{2}\sum_{i,j=1}^{d}\int_{B_R} e^{\psi(x)}\left(\frac{\partial\psi(x)}{\partial x_i}\frac{\partial\psi(x)}{\partial x_j} + \frac{\partial^2\psi(x)}{\partial x_i\partial x_j}\right)a^{ij}(x)\sigma_R(t,x)dx. \tag{4.10}$$

Similarly, $I_2(t) = -\sum_{i=1}^{d}\int_{B_R} f_i(x)\sigma_R(t,x)e^{\psi(x)}\frac{\partial\psi(x)}{\partial x_i}dx$. Therefore,

$$d\Phi(t) \leq \left(\int_{B_R}\mathfrak{F}(x)e^{\psi(x)}\sigma_R(t,x)dx\right)dt + \sum_{j=1}^{d}\left(\int_{B_R} h_j(x)e^{\psi(x)}u(t,x)dx\right)dY_t^j, \tag{4.11}$$

where

$$\mathfrak{F}(x) = \frac{1}{2}\sum_{i,j=1}^{d}\left(\frac{\partial\psi(x)}{\partial x_i}\frac{\partial\psi(x)}{\partial x_j} + \frac{\partial^2\psi(x)}{\partial x_i\partial x_j}\right)a^{ij}(x) + \sum_{i=1}^{d} f_i(x)\frac{\partial\psi(x)}{\partial x_i}. \tag{4.12}$$

Since $\psi(x) = \log\left(1+|x|^{2n}\right)$, then

$$\frac{\partial\psi}{\partial x_i} = \frac{2n|x|^{2n-2}x_i}{1+|x|^{2n}}, \; \frac{\partial^2\psi}{\partial x_i\partial x_j} = \frac{4nx_ix_j|x|^{2n-4}\left((n-1)-|x|^{2n}\right)}{\left(1+|x|^{2n}\right)^2} + \frac{2n|x|^{2n-2}\delta_{ij}}{1+|x|^{2n}}, \tag{4.13}$$

where δ_{ij} is the Kronecker's symbol, with $\delta_{ij} = 1$, if $i = j$, and $\delta_{ij} = 0$ otherwise.

Notice that

$$\left|\frac{\partial\psi}{\partial x_i}\right| \leq 2n, \; \left|\frac{\partial^2\psi}{\partial x_i\partial x_j}\right| \leq 4n^2 + 2n. \tag{4.14}$$

With the assumption that $|a^{ij}(x)| \le L$, we have

$$|\mathfrak{F}(x)| \le d^2(4n^2+n)L + 2n\sum_{i=1}^{d}\frac{|f_i(x)x_i|\cdot|x|^{2n-2}}{1+|x|^{2n}},\ \forall x\in\mathbb{R}^d. \tag{4.15}$$

Because $f(x)$ is Lipschitz continuous according to (**A1**), $|f(x)| \le L|x| + |f(0)|$, $\forall x \in \mathbb{R}^d$. Therefore,

$$\begin{aligned}|\mathfrak{F}(x)| &\le d^2(4n^2+n)L + \frac{2n|x|^{2n-2}}{1+|x|^{2n}}\sum_{i=1}^{d}\frac{|f_i(x)|^2+|x_i|^2}{2}\\ &\le d^2(4n^2+n)L + \frac{n|x|^{2n}+n|x|^{2n-2}|f(x)|^2}{1+|x|^{2n}}\\ &\le d^2(4n^2+n)L + n(L^2+1) + n|f(0)|^2 + 2nL|f(0)| \triangleq M(n,d,L),\ \forall x\in\mathbb{R}^d.\end{aligned} \tag{4.16}$$

Take expectation with respect to the reference probability measure $\tilde{P}$, we obtain $\frac{d}{dt}\tilde{E}\Phi(t) \le M(n,d,L)\tilde{E}\Phi(t)$, where we use the fact that Y_t is a Brownian motion with respect to $\tilde{P}$.

According to the Gronwall's inequality, we have

$$\sup_{0\le t\le T}\tilde{E}\int_{B_R}\left(1+|x|^{2n}\right)\sigma_R(t,x)dx \le e^{M(n,d,L)T}\int_{B_R}\left(1+|x|^{2n}\right)\sigma_0(x)dx. \tag{4.17}$$

Let R tends to infinity, and we have

$$\sup_{0\le t\le T}\tilde{E}\int_{\mathbb{R}^d}\left(1+|x|^{2n}\right)\sigma(t,x)dx \le e^{M(n,d,L)T}\int_{\mathbb{R}^d}\left(1+|x|^{2n}\right)\sigma_0(x)dx. \tag{4.18}$$

Therefore,

$$\begin{aligned}&\sup_{0\le t\le T}\tilde{E}\int_{|x|\ge R}\sigma(t,x)dx \le \frac{1}{1+R^{2n}}\sup_{0\le t\le T}\tilde{E}\int_{|x|\ge R}\left(1+|x|^{2n}\right)\sigma(t,x)dx\\ &\le \frac{1}{1+R^{2n}}\sup_{0\le t\le T}\tilde{E}\int_{\mathbb{R}^d}\left(1+|x|^{2n}\right)\sigma(t,x)dx \le \frac{e^{M(n,d,L)T}}{1+R^{2n}}\int_{\mathbb{R}^d}\left(1+|x|^{2n}\right)\sigma_0(x)dx.\end{aligned} \tag{4.19}$$

Moreover, with condition **(A4)**,

$$\begin{aligned}\sup_{0\le t\le T}\tilde{E}\int_{|x|\ge R}|\varphi(x)|\sigma(t,x)dx &\le \sup_{0\le t\le T}\tilde{E}\int_{|x|\ge R}\left(1+|x|^{2m}\right)\sigma(t,x)dx\\ &\le \frac{1}{1+R^{2n}}\sup_{0\le t\le T}\tilde{E}\int_{|x|\ge R}\left(1+|x|^{2n}\right)\left(1+|x|^{2m}\right)\sigma(t,x)dx\\ &\le \frac{2}{1+R^{2n}}\sup_{0\le t\le T}\tilde{E}\int_{\mathbb{R}^d}\left(1+|x|^{2(m+n)}\right)\sigma(t,x)dx\\ &\le \frac{2e^{M(n+m,d,L)T}}{1+R^{2n}}\int_{\mathbb{R}^d}\left(1+|x|^{2(m+n)}\right)\sigma_0(x)dx.\end{aligned} \tag{4.20}$$

□

Remark 4.2. According to the proof, the constant C in the statement of Theorem 4.1 may grow exponentially with respect to the terminal time T, which is not very satisfactory. In fact, numerical experiments in [6, 7, 16] show that the estimation error of different nonlinear filters based on the Yau-Yau algorithm can remain small in a relatively long period of time, and will not accumulate or even explode with respect to the time T. Therefore, an important future direction is to establish a tighter upper bound on the estimation error, which has only a mild dependence on, or is even independent of, the terminal time T.

5. Approximation of $\sigma(t,x)$ by the IBV problem in B_R. With the estimation in Theorem 4.1, because almost all the density of $\sigma(t,x)$ is contained in the closed ball B_R for R large enough, it is natural to think about approximating $\sigma(t,x)$ by the solution, $\sigma_R(t,x)$, to the corresponding initial-boundary value (IBV) problem (4.3) of DMZ equation in the ball B_R.

It will be rigorously proved in this section that, for R large enough, $\sigma(t,x)$ can be approximated well by $\sigma_R(t,x)$ defined in (4.3), and in particular, the estimation (3.8) holds in the proof of Theorem 3.1 in Section 3.

The main result in this section is stated as follows:

THEOREM 5.1. *With Assumptions **(A1)** to **(A4)**, there exists a constant $C>0$ which only depends on T, n, d and L, such that*

$$\sup_{0\le t\le T}\tilde{E}\int_{B_{\sqrt{R}}}|\sigma(t,x)-\sigma_R(t,x)|dx\le\frac{C}{1+R^n} \tag{5.1}$$

holds for all $R>0$, where $\sigma_R(t,x)$ is the solution of the IBV problem (4.3).

Proof of Theorem 5.1. For each $R>0$, consider the auxiliary function

$$\phi(x)=\log\left(1+R^n\left(1-\left(1-\frac{|x|^{2n}}{R^{2n}}\right)^2\right)\right),\quad x\in B_R, \tag{5.2}$$

and

$$\psi(x)=e^{-\phi(x)}-e^{-\phi(R)},\quad x\in B_R. \tag{5.3}$$

Define $v(t,x)=\sigma(t,x)-\sigma_R(t,x)$, $(t,x)\in[0,T]\times B_R$. Then, according to the maximum principle for SPDEs (cf. [4], for example), we have $v(t,x)\ge 0$, for all $(t,x)\in[0,T]\times B_R$ and a.s. $\tilde{P}$. Let Φ_t be the stochastic process defined by $\Phi_t=\int_{B_R}\psi(x)v(t,x)dx$. Since $v(t,x)$ is the solution to the SPDE

$$dv(t,x)=\left[\frac{1}{2}\sum_{i,j=1}^d\frac{\partial^2}{\partial x_i\partial x_j}(a^{ij}v)-\sum_{i=1}^d\frac{\partial}{\partial x_i}(f_iv)\right]dt+\sum_{j=1}^d h_jv\,dY_t^j, \tag{5.4}$$

the $\mathbb{R}$-valued stochastic process Φ_t satisfies

(5.5)

$$d\Phi_t=\frac{1}{2}\left(\sum_{i,j=1}^d\int_{B_R}\psi\frac{\partial^2(a^{ij}v)}{\partial x_i\partial x_j}dx-\sum_{i=1}^d\int_{B_R}\psi\frac{\partial(f_iv)}{\partial x_i}dx\right)dt+\sum_{j=1}^d\left(\int_{B_R}h_j\psi v\,dx\right)dY_t^j.$$

According to the Gauss-Green formula, we have

$$d\Phi_t=\frac{1}{2}\left(\sum_{i,j=1}^d\int_{B_R}a^{ij}\frac{\partial^2\psi}{\partial x_i\partial x_j}v\,dx\right)dt+\left(\sum_{i=1}^d\int_{B_R}\frac{\partial\psi}{\partial x_i}f_iv\,dx\right)dt$$

$$+\sum_{j=1}^{d}\left(\int_{B_R} h_j\psi v dx\right)dY_t^j+\left[-\int_{\partial B_R}\vec{\mathfrak{M}}_1\cdot\vec{\mathfrak{n}}dS+\int_{\partial B_R}\vec{\mathfrak{M}}_2\cdot\vec{\mathfrak{n}}dS\right]dt,$$

where, as in the proof of Theorem 4.1,

$$\vec{\mathfrak{M}}_i(t,x)=\left(\mathfrak{M}_{i,1}(t,x),\cdots,\mathfrak{M}_{i,d}(t,x)\right),\quad i=1,2, \tag{5.6}$$

$$\mathfrak{M}_{1,j}=\frac{1}{2}\sum_{i=1}^{d}\frac{\partial\psi}{\partial x_i}a^{ij}v,\ \mathfrak{M}_{2,i}(t,x)=\psi\left(\frac{1}{2}\sum_{j=1}^{d}\frac{\partial}{\partial x_j}(a^{ij}v)-f_iv\right),\ i,j=1,\cdots,d, \tag{5.7}$$

$\vec{\mathfrak{n}}$ denotes the outward normal vector of the boundary ∂B_R and dS denotes the measure on ∂B_R.

Notice that $\psi|_{\partial B_R}\equiv 0$ and $\frac{\partial\psi}{\partial x_j}=-e^{-\phi(x)}\frac{\partial\phi}{\partial x_j}$. Moreover,

$$\frac{\partial\phi}{\partial x_i}=\frac{2R^n\left(1-\frac{|x|^{2n}}{R^{2n}}\right)\frac{2n|x|^{2n-2}x_i}{R^{2n}}}{1+R^n\left(1-\left(1-\frac{|x|^{2n}}{R^{2n}}\right)^2\right)}, \tag{5.8}$$

and therefore, $\left.\frac{\partial\phi}{\partial x_i}\right|_{\partial B_R}=0=\left.\frac{\partial\psi}{\partial x_i}\right|_{\partial B_R}$, $i=1,\cdots,d$. Hence,

$$\begin{aligned} d\Phi_t =&\frac{1}{2}\left(\int_{B_R}e^{-\phi}v\sum_{i,j=1}^{d}a^{ij}\left(-\frac{\partial^2\phi}{\partial x_i\partial x_j}+\frac{\partial\phi}{\partial x_i}\frac{\partial\phi}{\partial x_j}\right)dx\right)dt\\ &-\left(\int_{B_R}\sum_{i=1}^{d}e^{-\phi}vf_i\frac{\partial\phi}{\partial x_i}dx\right)dt+\sum_{j=1}^{d}\left(\int_{B_R}h_j\psi v dx\right)dY_t^j \end{aligned} \tag{5.9}$$

Take expectation with respect to the probability measure $\tilde{P}$, and we have

$$\begin{aligned} \frac{d\tilde{E}\Phi_t}{dt}=&\tilde{E}\left(\frac{1}{2}\int_{B_R}\psi v\sum_{i,j=1}^{d}a^{ij}\left(-\frac{\partial^2\phi}{\partial x_i\partial x_j}+\frac{\partial\phi}{\partial x_i}\frac{\partial\phi}{\partial x_j}\right)dx-\int_{B_R}\sum_{i=1}^{d}\psi vf_i\frac{\partial\phi}{\partial x_i}dx\right)\\ &+e^{-\phi(R)}\tilde{E}\left(\frac{1}{2}\int_{B_R}v\sum_{i,j=1}^{d}a^{ij}\left(-\frac{\partial^2\phi}{\partial x_i\partial x_j}+\frac{\partial\phi}{\partial x_i}\frac{\partial\phi}{\partial x_j}\right)dx-\int_{B_R}\sum_{i=1}^{d}vf_i\frac{\partial\phi}{\partial x_i}dx\right) \end{aligned} \tag{5.10}$$

For $x\in B_R$, $|x_i|\le|x|\le R$, $i=1,\cdots,d$, and together with the Lipschitz conditions for $f(x)$,

$$\left|\frac{\partial\phi}{\partial x_i}(x)\right|\le\frac{4nR^n|x|^{2n-2}|x_i|}{R^{2n}\left(1+\frac{|x|^{2n}}{R^n}\right)}\le 4n,\ \forall\ x\in B_R; \tag{5.11}$$

$$\left|f_i(x)\frac{\partial\phi}{\partial x_i}\right|\le(|f(0)|+L|x|)\left|\frac{\partial\phi}{\partial x_i}\right|\le 4n(|f(0)|+L),\ \forall\ x\in B_R. \tag{5.12}$$

Also, according to direct computations,

(5.13)
$$\begin{aligned}\frac{\partial^2\phi}{\partial x_i\partial x_j}=&\frac{8n|x|^{2n-4}x_ix_j(R^{2n}(n-1)-(2n-1)|x|^{2n})(R^{3n}+2R^{2n}|x|^{2n}-|x|^{4n})}{(R^{3n}+2R^{2n}|x|^{2n}-|x|^{4n})^2}\\&-\frac{16n^2|x|^{2n-2}x_ix_j(R^{2n}|x|^{2n-2}-|x|^{4n-2})(R^{2n}-|x|^{2n})}{(R^{3n}+2R^{2n}|x|^{2n}-|x|^{4n})^2}\\&+\frac{4n\delta_{ij}|x|^{2n-2}(R^{2n}-|x|^{2n})}{R^{3n}+2R^{2n}|x|^{2n}-|x|^{4n}}.\end{aligned}$$

where δ_{ij} is the Kronecker's symbol. Thus,

$$\left|\frac{\partial^2\phi}{\partial x_i\partial x_j}\right|\le 8n(3n-2)+16n^2+4n,\ \forall x\in B_R. \tag{5.14}$$

We would like to remark that the estimation in (5.14) is quite rough. Each term on the right-hand side of (5.14) corresponds to one term on the right-hand side of (5.13), and the purpose is just to show the second-order derivatives are also bounded by a constant independent of R.

Notice that $e^{-\phi(R)}=\frac{1}{1+R^n}$. Together with the bounded condition for $a^{ij}(x)$, we have

$$\frac{d\tilde{E}\Phi(t)}{dt}\le C_1\tilde{E}\Phi(t)+\frac{C_1}{1+R^n}\tilde{E}\int_{B_R}v(t,x)dx \tag{5.15}$$

where $C_1>0$ is a constant which depends on n,d,L, but does not depend on R.

According to Theorem 4.1, the integral

$$\tilde{E}\int_{B_R}v(t,x)dx\le\tilde{E}\int_{B_R}\sigma(t,x)dx\le\tilde{E}\int_{\mathbb{R}^d}\sigma(t,x)dx, \tag{5.16}$$

which is also bounded by a constant independent of R, thus,

$$\frac{d\tilde{E}\Phi(t)}{dt}\le C_1\tilde{E}\Phi(t)+\frac{C_2}{1+R^n}. \tag{5.17}$$

where $C_2>0$ is a constant which depends on T,n,d,L.

By Gronwall's inequality,

$$\tilde{E}\Phi(t)\le\frac{C_3}{1+R^n} \tag{5.18}$$

where $C_3>0$ is a constant which depends on T, n, d and L.

On the other hand, for $R>5$ and for all $x\in B_{\sqrt{R}}$, i.e., $|x|\le\sqrt{R}$,

(5.19)
$$\phi(x)\in\left[0,\log\left(3-\frac{1}{R^n}\right)\right],\quad \psi(x)\ge\psi(\sqrt{R})=\frac{R^n}{3R^n-1}-\frac{1}{1+R^n}\ge\frac{1}{3}-\frac{1}{6}=\frac{1}{6}.$$

Then,

$$\tilde{E}\Phi(t)=\tilde{E}\int_{B_R}\psi(x)v(t,x)dx\ge\tilde{E}\int_{B_{\sqrt{R}}}\psi(x)v(t,x)dx\ge\frac{1}{6}\tilde{E}\int_{B_{\sqrt{R}}}v(t,x)dx. \tag{5.20}$$

Combining (5.18) and (5.20), we obtain that, for all $R\gg 1$,

$$\tilde{E}\int_{B_{\sqrt{R}}}|\sigma(t,x)-\sigma_R(t,x)|dx=\tilde{E}\int_{B_{\sqrt{R}}}v(t,x)dx\le\frac{6C_3}{1+R^n}. \tag{5.21}$$

□

6. Regularity of the Approximated Function $u_k(t,x)$. In this section, we will discuss the regularity of $u_k(t,x)$, $t \in [0,T]$, which is the solution of a series of coefficient-frozen equations (2.14).

The main purpose of this section is to show that under mild conditions, the recursively defined functions $u_k(t,x)$ will not explode in the finite time interval $[0,T]$, even if the time-discretization step $\delta \to 0$, in the sense that the L^2-norm of $u_k(\tau_k,x)$ ($k=1,\cdots,K$) is square integrable with respect to the probability measure $\tilde{P}$, and the expectations, $\tilde{E}\int_{B_R}|u_k(\tau_k,x)|^2dx$, are uniformly bounded for $k=1,\cdots,K$.

As shown in the next section, this following theorem is an essential intermediate result for the convergence analysis of this time-discretization scheme.

THEOREM 6.1. *Let $\{u_k(t,x) : \tau_{k-1} \le t \le \tau_k\}_{k=1}^K$ be the solution to the IBV problem of the coefficients-frozen equation* (2.14)*. Then, with Assumptions **(A1)** to **(A4)**, the L^2-norm of $u_k(\tau_k,x)$ is square-integrable with respect to the probability measure $\tilde{P}$, and we have*

$$\tilde{E}\int_{B_R}|u_k(\tau_k,x)|^2dx \le C < \infty,\ \forall\ k=1,\cdots,K, \tag{6.1}$$

where $C>0$ is a constant that depends on d, T, R, L, but is uniform in $k=1,\cdots,K$.

In the proof of Theorem 6.1, we will consider another exponential transformation given by $\sigma_k(t,x)=\exp(h^\top(x)Y_{\tau_{k-1}})u_k(t,x)$, $t\in[\tau_{k-1},\tau_k]$, $k=1,\cdots,K$. Direct computation implies that $\sigma_k(t,x)$ is the solution of

$$\begin{cases} \dfrac{\partial\sigma_k}{\partial t} = \dfrac{1}{2}\displaystyle\sum_{i,j=1}^d \dfrac{\partial^2}{\partial x_i\partial x_j}(a^{ij}\sigma_k) - \sum_{i=1}^d \dfrac{\partial}{\partial x_i}(f_i\sigma_k) - \dfrac{1}{2}|h|^2\sigma_k,\ (t,x)\in[\tau_{k-1},\tau_k]\times B_R, \\ \sigma_k(\tau_{k-1},x) = \exp\left(h^\top(x)Y_{\tau_{k-1}}\right)u_{k-1}(\tau_{k-1},x),\ x\in B_R \\ \sigma_k(t,x)=0,\ (t,x)\in[\tau_{k-1},\tau_k]\times\partial B_R, \end{cases} \tag{6.2}$$

and recursively, we can rewrite the initial value in (6.2) by

$$\sigma_k(\tau_{k-1},x) = \exp\left(h^\top(x)(Y_{\tau_{k-1}}-Y_{\tau_{k-2}})\right)\sigma_{k-1}(\tau_{k-1},x),\ k=2,\cdots,K. \tag{6.3}$$

Under the reference probability measure $\tilde{P}$, $\{Y_t : 0\le t\le T\}$ is a Brownian motion and $Y_{\tau_k}-Y_{\tau_{k-1}}\sim\mathcal{N}(0,\delta I_d)$, $k=1,\cdots,K$, with $I_d\in\mathbb{R}^{d\times d}$ the d-dimensional identity matrix. We would like to study the regularity of $\sigma_k(t,x)$ first, utilizing the Markov property of Y, and then derive the regularity results for $u_k(t,x)$.

For the sake of discussing the regularity of $\sigma_k(t,x)$ in a recursive manner, we need the following lemma which describes the relationship between $\sigma_k(\tau_{k-1},x)$ and $\sigma_{k-1}(\tau_{k-1},x)$ from (6.3).

LEMMA 6.2. *For $k=2,\cdots,K$, let $\sigma_k(t,x)$, $t\in[\tau_{k-1},\tau_k]$ be the solution of* (6.2)*. The end-point values $\sigma_k(\tau_{k-1},x)$ and $\sigma_{k-1}(\tau_{k-1},x)$ satisfy* (6.3)*. Let us denote by $L^4(B_R)$ the space of quartic-integrable functions in B_R. Assume that $\sigma_{k-1}(\tau_{k-1},\cdot)\in L^4(B_R)$, and the L^4-norm, $\|\sigma_{k-1}(\tau_{k-1},\cdot)\|_{L^4}$, is quartic integrable with respect to $\tilde{P}$, i.e., $\tilde{E}\int_{B_R}\sigma_{k-1}^4(\tau_{k-1},x)dx<\infty$, then $\sigma_k(\tau_{k-1},\cdot)\in L^4(B_R)$, its L^4-norm is quartic integrable with respect to $\tilde{P}$, and for sufficiently small time-discretization step size $\delta=\tau_k-\tau_{k-1}$, we have*

$$\tilde{E}\int_{B_R}\sigma_k^4(\tau_{k-1},x)dx \le (1+C\delta)\tilde{E}\int_{B_R}\sigma_{k-1}^4(\tau_{k-1},x)dx \tag{6.4}$$

where C is a constant that depends on d and R.

Proof of Lemma 6.2. According to the expression (6.3) and the definition of σ_{k-1} on $[\tau_{k-2},\tau_{k-1}]$, because of the Markov property of Y, $\exp\left(h^\top(x)(Y_{\tau_{k-1}}-Y_{\tau_{k-2}})\right)$ is independent of $\sigma_{k-1}(\tau_{k-1},x)$.

Because the observation function h is assumed to be smooth enough, and B_R is a bounded domain in $\mathbb{R}^d$, there exists a constant M, which may depend on R, such that the maximum of the absolute value of h, together with its partial derivatives up to order m, is bounded above by M.

Therefore, by Fubini's theorem,

$$\begin{aligned}\tilde{E}\int_{B_R}\sigma_k^4(\tau_{k-1},x)dx &= \tilde{E}\int_{B_R}\exp\left(4h^\top(x)(Y_{\tau_{k-1}}-Y_{\tau_{k-2}})\right)\sigma_{k-1}^4(\tau_{k-1},x)dx \\ &= \int_{B_R}\tilde{E}\exp\left(4h^\top(x)(Y_{\tau_{k-1}}-Y_{\tau_{k-2}})\right)\tilde{E}\sigma_{k-1}^4(\tau_{k-1},x)dx.\end{aligned} \tag{6.5}$$

Next, let us estimate the expectations of functions of normal random variable $\xi := Y_{\tau_{k-1}}-Y_{\tau_{k-2}}$ arising in the above expressions, for small time-discretization step δ.

In fact, because $\xi\sim\mathcal{N}(0,\delta I_d)$, we have

$$\tilde{E}\exp\left(4h(x)^\top\xi\right)=\prod_{j=1}^d\tilde{E}e^{4h_j(x)\xi_j}=\prod_{j=1}^d\left(\int_{\mathbb{R}}\frac{1}{\sqrt{2\pi\delta}}e^{4h_j(x)z}e^{-\frac{z^2}{2\delta}}dz\right) \tag{6.6}$$

In the bounded domain B_R,

$$\int_{\mathbb{R}}\frac{1}{\sqrt{2\pi\delta}}e^{4h_j(x)z}e^{-\frac{z^2}{2\delta}}dz=\int_{\mathbb{R}}\frac{1}{\sqrt{2\pi}}e^{4h_j(x)\sqrt{\delta}z}e^{-\frac{z^2}{2}}dz=e^{8h_j^2(x)\delta}\le e^{8M^2\delta}. \tag{6.7}$$

Therefore, for $\delta\ll 1$ (for example $\delta\le\frac{1}{16M^2d}$), $\tilde{E}\exp\left(4h(x)^\top\xi\right)\le e^{8dM^2\delta}\le 1+16dM^2\delta$. Thus,

$$\tilde{E}\int_{B_R}\sigma_k^4(\tau_{k-1},x)dx\le(1+16dM^2\delta)\tilde{E}\int_{B_R}\sigma_{k-1}^4(\tau_{k-1},x)dx. \tag{6.8}$$

□

Now, we are ready to give the proof of Theorem 6.1.

Proof of Theorem 6.1. The idea of this proof is to study the regularity of $\sigma_k(t,x)$, recursively, and then obtain the regularity of $u_k(t,x)$ based on the relationship (6.3).

In fact, according to the Cauchy-Schwartz inequality,

$$\begin{aligned}&\tilde{E}\int_{B_R}|u_k(\tau_k,x)|^2dx=\tilde{E}\int_{B_R}\exp\left(-2h^\top(x)Y_{\tau_{k-1}}\right)\sigma_k^2(\tau_k,x)dx\\ &\le\left(\tilde{E}\exp\left(4M\sum_{j=1}^d|Y_{\tau_{k-1},j}|\right)\right)^{\frac12}\left(\tilde{E}\left(\int_{B_R}\sigma_k^2(\tau_k,x)dx\right)^2\right)^{\frac12}\\ &\le C_1\left(\tilde{E}\exp\left(4M\sum_{j=1}^d|Y_{\tau_{k-1},j}|\right)\right)^{\frac12}\left(\tilde{E}\int_{B_R}\sigma_k^4(\tau_k,x)dx\right)^{\frac12}\end{aligned}$$

with $C_1>0$, a constant depending only on R.

Under the reference probability measure $\tilde{P}$, $\{Y_t:0\le t\le T\}$ is a standard d-dimensional Brownian motion. Therefore, the expectation $\tilde{E}\exp(4M\sum_{j=1}^d|Y_{\tau_{k-1},j}|)$ is bounded.

Hence, it remains to show that there exists a constant $C_2 > 0$, such that,

$$\tilde{E}\int_{B_R}\sigma_k^4(\tau_k,x)dx \le C_2 < \infty, \tag{6.9}$$

holds uniformly for $k = 1, \cdots, K$.

In the time interval $[\tau_{k-1}, \tau_k]$, $\sigma_k(t,x)$ is the solution to (6.2). According to the regularity results of parabolic partial differential equations, we have

$$\int_{B_R}\sigma_k^4(\tau_k,x)dx \le e^{C_4\delta}\int_{B_R}\sigma_k^4(\tau_{k-1},x)dx,\ \forall\ k=1,\cdots,K. \tag{6.10}$$

where C_4 is a constant which depends on the coefficients of the filtering system. The techniques in the proof of (6.10) is standard, and the proof of a counterpart, in which L^2-norm (instead of L^4-norm) is considered, can be found in the textbook [9]. We also provide a detailed proof in the Appendix, for the readers' convenience and in order to keep this paper self-contained.

Thus, with the result in Lemma 6.2, there exists $C_5, C_6 > 0$, such that for small enough δ,

$$\begin{aligned}\tilde{E}\int_{B_R}\sigma_k^4(\tau_k,x)dx &\le e^{C_4\delta}\tilde{E}\int_{B_R}\sigma_k^4(\tau_{k-1},x)dx \le e^{C_4\delta}(1+C_5\delta)\tilde{E}\int_{B_R}\sigma_{k-1}^4(\tau_{k-1},x)dx \\ &\le (1+C_6\delta)\tilde{E}\int_{B_R}\sigma_{k-1}^4(\tau_{k-1},x)dx.\end{aligned} \tag{6.11}$$

Inductively, we have

$$\tilde{E}\int_{B_R}\sigma_k^4(\tau_k,x)dx \le (1+C_6\delta)^{\frac{T}{\delta}}\int_{B_R}\sigma_0^4(x)dx \le e^{C_6T}\int_{B_R}\sigma_0^4(x)dx. \tag{6.12}$$

Thus, we have proved the boundedness of $\tilde{E}\int_{B_R}\sigma_k^4(\tau_k,x)dx$, and also, the result of Theorem 6.1 holds. ∎

7. Convergence Analysis of the Time Discretization Scheme.

This section serves to show that the solution $u_k(t,x)$ of the coefficient-frozen equations (2.14) can approximate the solution $u(t,x)$ of the original robust DMZ equation (2.13) well, if the time-discretization step size δ is small enough.

Also, we will show in this section that, after the exponential transformation $\exp(h^\top(x)Y_{\tau_k})$, the L^1-norm of the difference between the unnormalized densities $\sigma_R(\tau_k,x)$ (defined by (4.3)) and $\tilde{u}_{k+1}(\tau_k,x)$ (defined by (2.16)) still converges to zero, as $\delta \to 0$. In particular, the estimation (3.10) holds in the proof of Theorem 3.1 in Section 3.

THEOREM 7.1. *Fix $R > 0$. With Assumptions **(A1)** to **(A4)**, we can use the solution $u_k(t,x)$ of equation (2.14) to approximate the solution $u(t,x)$ of equation (2.13). In particular, for every $\epsilon > 0$, there exists a constant $\delta > 0$, such that*

$$\tilde{E}\int_{B_R}|\sigma_R(\tau_k,x) - \tilde{u}_{k+1}(\tau_k,x)|dx = \tilde{E}\int_{B_R}e^{h^\top(x)Y_{\tau_k}}\left|u(\tau_k,x) - u_k(\tau_k,x)\right|dx < \epsilon, \tag{7.1}$$

holds for every $k = 1, \cdots, K$.

Proof of Theorem 7.1. Since f is globally Lipschitz, $h \in C^2(B_R)$, and B_R is a bounded domain, there exists a constant $M_0 > 0$, such that the absolute value of each component in $f(x)$ and $h(x)$, as well as there first and second order derivatives, are dominated by M in the ball B_R, i.e.,

$$\max_{x\in B_R}\left\{\max_{1\le i\le d}|f_i|,\max_{1\le i\le d}|h_i|,\max_{1\le i,j\le d}\left|\frac{\partial f_i}{\partial x_j}\right|,\right. \\ \left.\max_{1\le i,j\le d}\left|\frac{\partial h_i}{\partial x_j}\right|,\max_{1\le i,j,k\le d}\left|\frac{\partial^2 f_i}{\partial x_j\partial x_k}\right|,\max_{1\le i,j,k\le d}\left|\frac{\partial^2 h_i}{\partial x_j\partial x_k}\right|\right\}\le M_0. \tag{7.2}$$

Let $B_{R,t}^+=\{x\in B_R: u(t,x)-u_k(t,x)\ge 0\}$. According to the technical Lemma 4.1 in [26], we have

$$\frac{d}{dt}\int_{B_{R,t}^+}(u(t,x)-u_k(t,x))dx=\int_{B_{R,t}^+}\frac{\partial}{\partial t}(u(t,x)-u_k(t,x))dx, \tag{7.3}$$

for almost all $t\in[0,T]$.

Then, according to equations (2.13) and (2.14) satisfied by $u(t,x)$ and $u_k(t,x)$ in $[\tau_{k-1},\tau_k]$,

$$\begin{aligned}
&\frac{d}{dt}\int_{B_{R,t}^+}(u(t,x)-u_k(t,x))dx=\int_{B_{R,t}^+}\frac{\partial}{\partial t}(u(t,x)-u_k(t,x))dx\\
&=\frac12\int_{B_{R,t}^+}\sum_{i,j=1}^d a^{ij}(x)\frac{\partial^2}{\partial x_i\partial x_j}(u-u_k)dx+\int_{B_{R,t}^+}\sum_{i=1}^d F_i(\tau_{k-1},x)\frac{\partial}{\partial x_i}(u-u_k)dx\\
&\quad+\int_{B_{R,t}^+}J(\tau_{k-1},x)(u(t,x)-u_k(t,x))dx\\
&\quad+\int_{B_{R,t}^+}\sum_{i=1}^d(F_i(t,x)-F_i(\tau_{k-1},x))\frac{\partial u}{\partial x_i}dx+\int_{B_{R,t}^+}(J(t,x)-J(\tau_{k-1},x))u(t,x)dx.
\end{aligned} \tag{7.4}$$

Because $u(t,x)=u_k(t,x)\equiv 0$ on the boundary ∂B_R, and $\partial B_{R,t}^+\subset\partial B_R\cup\{x\in B_R: u(t,x)-u_k(t,x)=0\}$, we have $(u-u_k)|_{\partial B_{R,t}^+}=0$ and $\nabla(u-u_k)|_{\partial B_{R,t}^+}=-c(x)\vec{\mathfrak{n}}$ with $\vec{\mathfrak{n}}$ the outward normal vector of $B_{R,t}^+$ and $c(x)\ge 0$ on $\partial B_{R,t}^+$. Thus, the first three terms on the right-hand side of (7.4) can be estimated by

$$\begin{aligned}
&\frac12\int_{B_{R,t}^+}\sum_{i,j=1}^d a^{ij}\frac{\partial^2}{\partial x_i\partial x_j}(u-u_k)dx+\int_{B_{R,t}^+}\sum_{i=1}^d F_i(\tau_{k-1},x)\frac{\partial}{\partial x_i}(u-u_k)dx\\
&\quad+\int_{B_{R,t}^+}J(\tau_{k-1},x)(u(t,x)-u_k(t,x))dx\\
&\quad=\frac12\int_{B_{R,t}^+}\sum_{i,j=1}^d\frac{\partial^2 a^{ij}}{\partial x_i\partial x_j}(u-u_k)dx-\int_{B_{R,t}^+}\sum_{i=1}^d\frac{\partial F_i(\tau_{k-1},x)}{\partial x_i}(u-u_k)dx\\
&\quad+\int_{B_{R,t}^+}J(\tau_{k-1},x)(u-u_k)dx-\frac12\int_{\partial B_{R,t}^+}\sum_{i,j=1}^d a^{ij}\frac{\partial}{\partial x_i}(u-u_k)\frac{\partial}{\partial x_j}(u-u_k)dS\\
&\qquad-\frac12\int_{\partial B_{R,t}^+}(u-u_k)\sum_{i,j=1}^d\frac{\partial a^{ij}}{\partial x_i}\vec{\mathfrak{n}}_j dS+\int_{\partial B_{R,t}^+}(u-u_k)\sum_{i=1}^d F_i(\tau_{k-1},x)\vec{\mathfrak{n}}_i dS\\
&\quad\le\left(\frac{d^2L}{2}+C(d,L,M_0)\left(1+\sum_{j=1}^d|Y_{\tau_{k-1},j}|\right)^2\right)\int_{B_{R,t}^+}(u-u_k)dx,
\end{aligned}$$

$$\leq C_1\left(1+\sum_{j=1}^{d}|Y_{\tau_{k-1},j}|\right)^2\int_{B^+_{R,t}}(u-u_k)dx$$

where we use the fact that $a(x)=g(x)g(x)^\top$ is positive semi-definite and the definition of $F_i(t,x)$ and $J(t,x)$ in (2.12); $C(d,L,M_0)$ and C_1 are constants which depend only on d,L,M_0; and dS denotes the measure on $\partial B^+_{R,t}$.

Also, by the definition of $F_i(t,x)$ and $J(t,x)$ in (2.12), we have the following estimation of the differences

$$\begin{aligned}&|F(t,x)-F(\tau_{k-1},x)|\leq C_2|Y_t-Y_{\tau_{k-1}}|,\\ &|J(t,x)-J(\tau_{k-1},x)|\leq C_3\left(1+\sum_{j=1}^{d}(|Y_{t,j}|+|Y_{\tau_{k-1},j}|)\right)|Y_t-Y_{\tau_k}|,\ \forall\ x\in B_R,\end{aligned}\tag{7.5}$$

where C_2 and C_3 are constants which only depends on $d,L.M_0$.

Hence,

$$\begin{aligned}&\frac{d}{dt}\int_{B^+_{R,t}}(u(t,x)-u_k(t,x))dx\leq C_1\left(1+\sum_{j=1}^{d}|Y_{\tau_{k-1},j}|\right)^2\int_{B^+_{R,t}}(u-u_k)dx\\ &+C_2|Y_t-Y_{\tau_{k-1}}|\int_{B_R}|\nabla u|dx+C_3\left(1+\sum_{j=1}^{d}(|Y_{t,j}|+|Y_{\tau_{k-1},j}|)\right)|Y_t-Y_{\tau_k}|\int_{B_R}|u|dx\end{aligned}\tag{7.6}$$

holds for almost all $t\in[\tau_{k-1},\tau_k]$ and almost surely, where C_1, C_2 and C_3 are constants which depend on the coefficients of the system.

Under the reference probability distribution $\tilde{P}$, the observation process $\{Y_t:0\leq t\leq T\}$ is a standard d-dimensional Brownian motion, and therefore, $\sum_{j=1}^{d}(Y_{t,j}-Y_{\tau_{k-1},j})\sim N(0,d(t-\tau_{k-1}))$.

Let $\Omega_{M_1}=\left\{\omega:\sup\limits_{0\leq t\leq T}\sum_{j=1}^{d}|Y_{t,j}(\omega)|\leq M_1\right\}$ be the event which represents the observation process Y_t is not severely abnormal, and $1_A(\cdot)$ is the indicator function of the set A.

For a fixed $M_1>0$, let us first take the expectation with respect to $\tilde{P}$ on the event Ω_{1,M_1} for both sides of (7.6), and we have

$$\begin{aligned}&\frac{d}{dt}\tilde{E}\left[1_{\Omega_{M_1}}\int_{B^+_{R,t}}(u-u_k)dx\right]\leq C_1\tilde{E}\left[1_{\Omega_{M_1}}\left(1+\sum_{j=1}^{d}|Y_{\tau_{k-1},j}|\right)^2\int_{B^+_{R,t}}(u-u_k)dx\right]\\ &\quad+C_2\tilde{E}\left[1_{\Omega_{M_1}}|Y_t-Y_{\tau_{k-1}}|\int_{B_R}|\nabla u|dx\right]\\ &\quad+C_3\tilde{E}\left[1_{\Omega_{M_1}}\left(1+\sum_{j=1}^{d}(|Y_{t,j}|+|Y_{\tau_{k-1},j}|)\right)|Y_t-Y_{\tau_k}|\int_{B_R}|u|dx\right]\\ &\leq C_1(1+M_1)^2\tilde{E}\left[1_{\Omega_{M_1}}\int_{B^+_{R,t}}(u-u_k)dx\right]+C_2\tilde{E}\left[1_{\Omega_{M_1}}|Y_t-Y_{\tau_{k-1}}|\int_{B_R}|\nabla u|dx\right]\\ &\quad+C_3(1+2M_1)\tilde{E}\left[1_{\Omega_{M_1}}|Y_t-Y_{\tau_{k-1}}|\int_{B_R}|u|dx\right]\\ &\leq C_1(1+M_1)^2\tilde{E}\left[1_{\Omega_{M_1}}\int_{B^+_{R,t}}(u-u_k)dx\right]\end{aligned}$$

$$+ C_2\left(\tilde{E}|Y_t - Y_{\tau_{k-1}}|^2\right)^{\frac{1}{2}} \left(\tilde{E}\left[1_{\Omega_{M_1}} \left(\int_{B_R} |\nabla u| dx\right)^2\right]\right)^{\frac{1}{2}}$$

$$+ C_3(1+2M_1)\left(\tilde{E}|Y_t - Y_{\tau_{k-1}}|^2\right)^{\frac{1}{2}} \left(\tilde{E}\left[1_{\Omega_{M_1}} \left(\int_{B_R} |u| dx\right)^2\right]\right)^{\frac{1}{2}}$$

$$= C_1(1+M_1)^2\tilde{E}\left[1_{\Omega_{M_1}} \int_{B^+_{R,t}} (u(t,x) - u_k(t,x))dx\right]$$

$$+ C_2 d^{\frac{1}{2}}(t-\tau_{k-1})^{\frac{1}{2}} \left(\tilde{E}\left[1_{\Omega_{M_1}} \left(\int_{B_R} |\nabla u| dx\right)^2\right]\right)^{\frac{1}{2}}$$

$$+ C_3(1+2M_1) d^{\frac{1}{2}}(t-\tau_{k-1})^{\frac{1}{2}} \left(\tilde{E}\left[1_{\Omega_{M_1}} \left(\int_{B_R} |u| dx\right)^2\right]\right)^{\frac{1}{2}}.$$

Here, the second inequality holds because of the property of the event Ω_{M_1}, the third inequality holds according to the Cauchy-Schwartz inequality and the last equality holds because Y_t is a normal distributed random vector.

On the event Ω_{M_1}, the observation process $\{Y_t : 0 \leq t \leq T\}$ is bounded. Therefore, according to the regularity results of parabolic partical differential equations (cf. [9], Section 7.1, Theorem 6), the integrals $\int_{B_R} |\nabla u| dx$ and $\int_{B_R} |u| dx$ are also bounded for almost every $t \in [0,T]$, as long as $f \in C^1(B_R)$ and $h \in C^2(B_R)$. Thus,

(7.7)
$$\frac{d}{dt}\tilde{E}\left[1_{\Omega_{M_1}} \int_{B^+_{R,t}} (u-u_k)dx\right] \leq C_4\tilde{E}\left[1_{\Omega_{M_1}} \int_{B^+_{R,t}} (u-u_k)dx\right] + C_5(t-\tau_{k-1})^{\frac{1}{2}},$$

where $C_4, C_5 > 0$ are constants which depend on d, L, M_0, M_1, T.

Similarly, we also have the estimation for the integral on the set $B^-_{R,t} = \{x \in B_R : u(t,x) - u_k(t,x) \leq 0\}$:

(7.8)
$$\frac{d}{dt}\tilde{E}\left[1_{\Omega_{M_1}} \int_{B^-_{R,t}} (u-u_k)dx\right] \leq C_4\tilde{E}\left[1_{\Omega_{M_1}} \int_{B^-_{R,t}} (u-u_k)dx\right] + C_5(t-\tau_{k-1})^{\frac{1}{2}},$$

and thus

(7.9)
$$\frac{d}{dt}\tilde{E}\left[1_{\Omega_{M_1}} \int_{B_R} |u-u_k|dx\right] \leq C_4\tilde{E}\left[1_{\Omega_{M_1}} \int_{B_R} |u-u_k|dx\right] + 2C_5(t-\tau_{k-1})^{\frac{1}{2}}.$$

Therefore,

(7.10)
$$\tilde{E}\left[1_{\Omega_{M_1}} \int_{B_R} |u(t,x) - u_k(t,x)|dx\right] \leq e^{C_4(t-\tau_{k-1})}\left(\tilde{E}\left[1_{\Omega_{M_1}} \int_{B_R} |u(\tau_{k-1},x) - u_k(\tau_{k-1},x)|dx\right] + \frac{4}{3}C_5(t-\tau_{k-1})^{\frac{3}{2}}\right).$$

Notice that $u_k(\tau_{k-1},x) \equiv u_{k-1}(\tau_{k-1},x)$ by definition. Inductively, we have

$$\begin{aligned}
&\tilde{E}\left[1_{\Omega_{M_1}} \int_{B_R} |u(\tau_k,x) - u_k(\tau_k,x)| dx\right] \\
&\le e^{C_4\delta}\left(\tilde{E}\left[1_{\Omega_{M_1}} \int_{B_R} |u(\tau_{k-1},x) - u_{k-1}(\tau_{k-1},x)| dx\right] + \frac{4}{3}C_5\delta^{\frac{3}{2}}\right) \qquad (7.11)\\
&\le \frac{4}{3}C_5\delta^{\frac{3}{2}} \sum_{i=1}^{k} e^{C_4(i-1)\delta} \le \frac{4}{3}C_5\delta^{\frac{3}{2}} k e^{C_4 k\delta} \le C_6\delta^{\frac{1}{2}}.
\end{aligned}$$

where C_6 is a constant which depends on d, L, M_0, M_1, T.

Also, for the value we are concerned with in (7.1),

$$\tilde{E}\left[1_{\Omega_{M_1}} \int_{B_R} e^{h^\top(x) Y_{\tau_k}} |u(\tau_k,x) - u_k(\tau_k,x)| dx\right] \le C_6 e^{M_0 M_1} \delta^{\frac{1}{2}} \qquad (7.12)$$

On $\Omega_{M_1}^c = \{\omega : \sup_{0\le t\le T} \sum_{j=1}^d |Y_{t,j}(\omega)| > M_1\}$, let $\overline{Y}_T \triangleq \sup_{0\le t\le T}\sum_{j=1}^d |Y_{t,j}|$, then,

(7.13)

$$\begin{aligned}
&\tilde{E}\left[1_{\Omega_{M_1}^c} \int_{B_R} e^{h^\top(x) Y_{\tau_k}} |u(\tau_k,x) - u_k(\tau_k,x)| dx\right] \\
&\le \tilde{E}\left[1_{\Omega_{1,M_1}^c} \frac{\overline{Y}_T}{M_1} \exp\left(M_0 \sum_{j=1}^d |Y_{\tau_k,j}|\right) \int_{B_R} |u(\tau_k,x) - u_k(\tau_k,x)| dx\right] \\
&\le \frac{1}{M_1}\left(\tilde{E}\left[\overline{Y}_T^2 \exp\left(2M_0 \sum_{j=1}^d |Y_{\tau_k,j}|\right)\right]\right)^{\frac{1}{2}} \left(\tilde{E}\left(\int_{B_R} |u(\tau_k,x) - u_k(\tau_k,x)| dx\right)^2\right)^{\frac{1}{2}} \\
&\le \frac{C_7}{M_1}\left(\tilde{E}\xi^2\right)^{\frac{1}{2}} \left(\tilde{E}\int_{B_R} |u(\tau_k,x)|^2 dx + \tilde{E}\int_{B_R} |u_k(\tau_k,x)|^2 dx\right)^{\frac{1}{2}}.
\end{aligned}$$

where $C_7 > 0$ is a constant which is related to the volume of the d-dimensional ball B_R, and ξ is the random variable given by $\xi = \overline{Y}_T \exp\left(M_0 \sum_{j=1}^d |Y_{\tau_k,j}|\right)$.

Thus,

(7.14)

$$\tilde{E}\xi^2 = \tilde{E}\left[\overline{Y}_T^2 \exp\left(2M_0 \sum_{j=1}^d |Y_{\tau_k,j}|\right)\right] \le \left(\tilde{E}\overline{Y}_T^4\right)^{\frac{1}{2}} \left(\tilde{E}\exp\left(4M_0 \sum_{j=1}^d |Y_{\tau_k,j}|\right)\right)^{\frac{1}{2}}$$

According to the Burkholder-Davis-Gundy inequality (cf. [15], Chapter 3, Theorem 3.28, for example), there exists $C_8 > 0$, such that

$$\tilde{E}\overline{Y}_T^4 \le C_8 \tilde{E}\sum_{j=1}^d |Y_{T,j}|^4 \le 3C_8 d T^2. \qquad (7.15)$$

and because $Y_{\tau_k,j}$ are normal, the expectation of $\exp(4M_0\sum_{j=1}^d Y_{\tau_k,j})$ is bounded.

For the value $\tilde{E}\int_{B_R} |u(\tau_k,x)|^2 dx$, because

$$u(t,x) = \exp\left(-\sum_{j=1}^d h_j(x) Y_{t,j}\right)\sigma(t,x), \qquad (7.16)$$

then

(7.17)

$$\begin{aligned}\tilde{E}\int_{B_R}|u(\tau_k,x)|^2dx &= \tilde{E}\int_{B_R}\exp\Big(-2\sum_{j=1}^d h_j(x)Y_{\tau_k,j}\Big)\sigma^2(\tau_k,x)dx\\ &\leq \tilde{E}\Bigg[\exp\left(2M_0\sum_{j=1}^d|Y_{\tau_k,j}|\right)\int_{B_R}\sigma^2(\tau_k,x)dx\Bigg]\\ &\leq \left(\tilde{E}\exp\left(4M_0\sum_{j=1}^d|Y_{\tau_k,j}|\right)\right)^{\frac{1}{2}}\left(\tilde{E}\left(\int_{B_R}|\sigma(\tau_k,x)|^2dx\right)^2\right)^{\frac{1}{2}}.\end{aligned}$$

Notice that $\sigma(t,x)$ is the solution to the stochastic partial differential equation

$$d\sigma(t,x)=\mathcal{L}^*\sigma(t,x)dt+\sum_{j=1}^d h_j\sigma(t,x)dY_{t,j}. \tag{7.18}$$

and the boundedness of $\tilde{E}\left(\int_{B_R}|\sigma(\tau_k,x)|^2dx\right)^2$ follows from the regularity theory of stochastic partial differential equation.

In the monograph [21], the authors provided a similar regularity result, and proved that $\tilde{E}\int_{B_R}|\sigma(\tau_k,x)|^2dx$ is bounded by the initial values. Here in our case, we will prove that there exists $C_9>0$, such that

$$\tilde{E}\left(\int_{B_R}|\sigma(\tau_k,x)|^2dx\right)^2\leq C_9\left(\int_{B_R}|\sigma_0(x)|^2dx\right)^2. \tag{7.19}$$

The detailed proof of (7.19) can be found in the Appendix.

Therefore, we have

$$\tilde{E}\int_{B_R}|u(\tau_k,x)|^2dx\leq C_{10}, \tag{7.20}$$

where $C_{10}>0$ is a constant that does not depend on δ or M_1.

Furthermore, as we have discussed in the previous section, $\tilde{E}\int_{B_R}|u_k(\tau_k,x)|^2dx$ is also bounded above, and thus, we have

$$\tilde{E}\Bigg[1_{\Omega^c_{M_1}}\int_{B_R}e^{h^\top(x)Y_{\tau_k}}|u(\tau_k,x)-u_k(\tau_k,x)|dx\Bigg]\leq\frac{C_{11}}{M_1}, \tag{7.21}$$

where C_{11} is a constant which does not depend on M_1 or δ.

In summary, for each $\epsilon>0$, there exists $M_1>0$, such that $\frac{C_{11}}{M_1}<\frac{\epsilon}{2}$, and for this particular M_1, there exists $\delta>0$, such that $C_6e^{M_0M_1}\delta^{\frac{1}{2}}<\frac{\epsilon}{2}$. Therefore, for every $k=1,\cdots,K$,

$$\begin{aligned}&\tilde{E}\int_{B_R}e^{h^\top(x)Y_{\tau_k}}|u(\tau_k,x)-u_k(\tau_k,x)|dx\\ &=\tilde{E}\Bigg[1_{\Omega_{1,M_1}}\int_{B_R}e^{h^\top(x)Y_{\tau_k}}|u(\tau_k,x)-u_k(\tau_k,x)|dx\Bigg]\\ &\quad+\tilde{E}\Bigg[1_{\Omega^c_{1,M_1}}\int_{B_R}e^{h^\top(x)Y_{\tau_k}}|u(\tau_k,x)-u_k(\tau_k,x)|dx\Bigg]\\ &\leq C_6e^{M_0M_1}\delta^{\frac{1}{2}}+\frac{C_{10}}{M_1}<\epsilon.\end{aligned} \tag{7.22}$$

□

8. Conclusion. In this paper, we provide a novel convergence analysis of Yau-Yau algorithm from a probabilistic perspective. With very liberal assumptions only on the coefficients of the filtering systems and the initial distributions (without assumptions on particular paths of observations), we can prove that Yau-Yau algorithm can provide accurate approximations with arbitrary precision to a quite broad class of statistics for the conditional distribution of state process given the observations, which includes the most commonly used conditional mean and covariance matrix. Therefore, the capability of Yau-Yau algorithm to solve very general nonlinear filtering problems is theoretically verified in this paper.

In the process of deriving this probabilistic version of the convergence results, we study the properties of the exact solution, $\{\sigma(t,x) : 0 \leq t \leq T\}$, to the DMZ equation and the approximated solution $\{\tilde{u}_{k+1}(\tau_k, x) : 1 \leq k \leq K\}$, given by Yau-Yau algorithm, respectively.

For the exact solution $\sigma(t,x)$ of the DMZ equation, we have shown in Section 4 and Section 5 that most of the density of $\sigma(t,x)$ will remain in the closed ball B_R, and $\sigma(t,x)$ can be approximated well by the corresponding initial-boundary value problem of DMZ equation in B_R. This result also implies that it is very unlikely for the state process to reach infinity within finite terminal time.

For the approximated solution $\tilde{u}_{k+1}(\tau_k, x)$ given by Yau-Yau algorithm, we have first proved in Section 6 that $\tilde{u}_{k+1}(\tau_k, x)$, which evolves in a recursive manner, will not explode in finite time interval, even if the time-discretization step $\delta \to 0$. And then, in Section 7, the convergence of $\tilde{u}_{k+1}(\tau_k, x)$ is proved and the convergence rate is also estimated to be $\sqrt{\delta}$.

It is clear that the properties of exact solutions and approximated solutions, which we have proved in this paper, highly rely on the nice properties of Brownian motion and Gaussian distributions, especially the Markov and light-tail properties. On the one hand, Brownian motion and Gaussian distribution are up to now, among the most commonly used objects in the mathematical modeling of many areas of applications, and can describe most scenarios in practice. On the other hand, for those systems driven by non-Markov or heavy-tailed processes, minimum mean square criteria, together with the conditional expectations (if exist), may not result in a satisfactory estimation of the state process. In this case, the studies of estimations based on other criteria, such as maximum a posteriori (MAP) [11][22][14], will be a promising direction.

Finally, in this paper, we only consider filtering systems and conduct convergence analysis in time interval $[0,T]$ with a fixed finite terminal time T. It is also interesting to study the behavior of the DMZ equation and the approximation capability of Yau-Yau algorithm in the case where the terminal time $T \to \infty$, especially for filtering systems with further stable assumptions. We will continue working on how to combine the existing studies on filter stability, such as [1][19], with our techniques developed in this paper. Furthermore, the study of the long-time behavior of Yau-Yau algorithm will lead to more precise techniques of determining the parameters (such as the radius R of the closed ball and the time-discretization step δ) in the algorithm design, and hopefully, some convergence results of Yau-Yau algorithm for the whole time line $(0,\infty)$ can be obtained.

Appendix A. Regularity Results of Parabolic Partial Differential Equation and Stochastic Evolution Equation. In this appendix, we will provide a detailed proof of the regularity results of the parabolic partial differential equation and the stochastic evolution equation.

For the purpose of deriving (6.10) and (7.19), the regularity results is slightly different from standard ones considered in square-integrable functional spaces.

THEOREM A.1. *Let $\sigma(t,x)$ be the solution of the following IBV problem:*

$$\begin{cases}\dfrac{\partial\sigma(t,x)}{\partial t}=\dfrac{1}{2}\displaystyle\sum_{i,j=1}^{d}\frac{\partial^2}{\partial x_i\partial x_j}(a^{ij}\sigma)-\sum_{i=1}^{d}\frac{\partial}{\partial x_i}(f_i\sigma)-\frac{1}{2}|h|^2\sigma,\ (t,x)\in[0,T]\times B_R,\\ \sigma(0,x)=\sigma_0(x),\ x\in B_R,\ \sigma(t,x)=0,\ (t,x)\in[0,T]\times\partial B_R,\end{cases} \tag{A.1}$$

where $B_R=\{x\in\mathbb{R}^d:|x|\le R\}$ is the ball in $\mathbb{R}^d$ with radius R; $a:\mathbb{R}^d\to\mathbb{R}^{d\times d}$, $f:\mathbb{R}^d\to\mathbb{R}^d$, $h:\mathbb{R}^d\to\mathbb{R}^d$ are smooth enough functions. Assume that the matrix-valued function $a(x)$ is uniformly positive definite, i.e., there exists $\lambda>0$, such that

$$\sum_{i,j=1}^{d}a^{ij}(x)\xi_i\xi_j\ge\lambda|\xi|^2,\ \forall\ x\in B_R,\ \xi\in\mathbb{R}^d. \tag{A.2}$$

If the initial value $\sigma_0(x)$ is quartic-integrable in B_R, then there exists a constant $C>0$, which depends on the coefficients of the system, such that

$$\int_{B_R}\sigma^4(T,x)dx\le e^{CT}\int_{B_R}\sigma_0^4(x)dx. \tag{A.3}$$

Remark A.2. In fact, Assumption **(A2)** in the main text will imply the coercivity condition (A.2). This is because the closed ball B_R is a compact set of $\mathbb{R}^d$, and the continuous function $\lambda(x)$ in Assumption **(A2)** will map B_R to a compact set. Therefore, there exists $\lambda>0$, such that $\lambda(x)\ge\lambda>0$, for all $x\in B_R$.

Proof. Let us define

$$\tilde{f}_i(x)=f_i(x)-\sum_{j=1}^{d}\frac{\partial a^{ij}(x)}{\partial x},\ i=1,\cdots,d. \tag{A.4}$$

Then the parabolic equation (A.1) can be written in a divergence form

$$\frac{\partial\sigma(t,x)}{\partial t}=\frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial}{\partial x_i}\left(a^{ij}(x)\frac{\partial}{\partial x_j}\sigma(t,x)\right)-\sum_{i=1}^{d}\frac{\partial}{\partial x_i}(\tilde{f}_i(x)\sigma(t,x))-\frac{1}{2}|h(x)|^2\sigma(t,x). \tag{A.5}$$

Hence,

$$\begin{aligned}\frac{d}{dt}\int_{B_R}\sigma^4(t,x)dx&=\int_{B_R}4\sigma^3(t,x)\frac{\partial\sigma}{\partial t}dx\\&=-6\int_{B_R}\sigma^2\sum_{i,j=1}^{d}a^{ij}\frac{\partial\sigma}{\partial x_i}\frac{\partial\sigma}{\partial x_j}dx+12\int_{B_R}\sum_{i=1}^{d}\tilde{f}_i\sigma^3\frac{\partial\sigma}{\partial x_i}dx-2\int_{B_R}\sigma^4|h|^2dx\\&\le-6\lambda\int_{B_R}\sigma^2|\nabla\sigma|^2dx+12\int_{B_R}\sum_{i=1}^{d}\frac{\tilde{f}_i\sigma^2}{\sqrt{\lambda}}\cdot\left(\sqrt{\lambda}\sigma\frac{\partial\sigma}{\partial x_i}\right)dx-2\int_{B_R}\sigma^4|h|^2dx\\&\le12\int_{B_R}\sum_{i=1}^{d}\left(\frac{\tilde{f}_i^2\sigma^4}{2\lambda}+\frac{\lambda}{2}\sigma^2\left|\frac{\partial\sigma}{\partial x_i}\right|^2\right)dx\le\int_{B_R}\left(\frac{6}{\lambda}\sum_{i=1}^{d}\tilde{f}_i^2-2|h|^2\right)\sigma^4(t,x)dx.\end{aligned}$$

In the bounded domain B_R, there exists a constant $C > 0$, such that $\left|\frac{6}{\lambda}\sum_{i=1}^d \tilde{f}_i^2 - 2|h|^2\right| \le C$. Thus,

$$\frac{d}{dt}\int_{B_R} \sigma^4(t,x)dx \le C\int_{B_R}\sigma^4(t,x)dx,\ t\in[0,T], \tag{A.6}$$

and by Gronwall's inequality, we have

$$\int_{B_R}\sigma^4(T,x)dx \le e^{CT}\int_{B_R}\sigma_0^4(x)dx. \tag{A.7}$$ ∎

Theorem A.3. *Consider the IBV problem of stochastic partial differential equation given by*

$$\begin{cases} d\sigma(t,x) = \mathcal{L}^*\sigma(t,x)dt + \sum_{j=1}^d h_j(x)\sigma(t,x)dY_{t,j},\ t\in[0,T] \\ \sigma(t,x) = 0,\ (t,x)\in[0,T]\times\partial B_R,\ \sigma(0,x)=\sigma_0(x),\ x\in B_R. \end{cases} \tag{A.8}$$

where $Y = \{Y_t : 0 \le t \le T\}$ is a standard d-dimensional Brownian motion in the filtered probability space $(\Omega,\mathcal{F},\{\mathcal{F}_t\}_{0\le t\le T},P)$; $B_R = \{x\in\mathbb{R}^d : |x|\le R\}$ is the ball in $\mathbb{R}^d$ with radius R, and

$$\mathcal{L}^*(\star) = \frac{1}{2}\sum_{i,j=1}^d \frac{\partial^2}{\partial x_i\partial x_j}(a^{ij}(x)\star) - \sum_{i=1}^d \frac{\partial}{\partial x_i}(f_i(x)\star). \tag{A.9}$$

*Assume that the coefficients a, f, h are smooth enough and the Assumption **(A2)** holds for the matrix-valued function $a(x)$, which implies that $a(x)$ is uniformly positive definite in B_R, i.e., there exists $\lambda > 0$, such that*

$$\sum_{i,j=1}^d a^{ij}(x)\xi_i\xi_j \ge \lambda|\xi|^2,\ \forall\ x\in B_R,\ \xi\in\mathbb{R}^d. \tag{A.10}$$

If the initial value $\sigma_0(x)$ is square-integrable in B_R, then there exists a constant $C > 0$, which depends on T, R and the coefficients of the system, such that

$$E\left(\int_{B_R}|\sigma(T,x)|^2dx\right)^2 \le C\left(\int_{B_R}|\sigma_0(x)|^2dx\right)^2. \tag{A.11}$$

Proof. Let us define $\tilde{f}_i(x) = f_i(x) - \sum_{j=1}^d \frac{\partial a^{ij}(x)}{\partial x}$, $i = 1,\cdots,d$. Then the stochastic partial differential equation in (A.8) can be rewritten in divergence form:

$$d\sigma(t,x) = \frac{1}{2}\sum_{i,j=1}^d \frac{\partial}{\partial x_i}\left(a^{ij}\frac{\partial\sigma}{\partial x_j}\right) - \sum_{i=1}^d\frac{\partial}{\partial x_i}(\tilde{f}_i\sigma) + \sum_{j=1}^d h_j\sigma dY_{t,j}. \tag{A.12}$$

Let

$$\Phi(t) = \int_{B_R}\sigma^2(t,x)dx,\ t\in[0,T], \tag{A.13}$$

then according to Itô's formula,

$$d\Phi(t) = \left(\int_{B_R} (2\sigma\mathcal{L}^*\sigma + \sigma^2|h|^2)dx\right) dt + \sum_{j=1}^{d} \left(\int_{B_R} 2h_j\sigma^2 dx\right) dY_{t,j} \tag{A.14}$$

and

$$\begin{aligned} d\Phi^2(t) = {} & 2\left(\int_{B_R} \sigma^2 dx\right)\left(\int_{B_R} (2\sigma\mathcal{L}^*\sigma + \sigma^2|h|^2)dx\right) dt \\ & + 2\Phi(t)\sum_{j=1}^{d}\left(\int_{B_R} 2h_j\sigma^2 dx\right) dY_{t,j} + \sum_{j=1}^{d}\left(\int_{B_R} 2h_j\sigma^2 dx\right)^2 dt \end{aligned} \tag{A.15}$$

After taking expectations, we have

$$\frac{d}{dt}E\Phi^2(t) = E\Bigg[2\left(\int_{B_R} \sigma^2 dx\right)\left(\int_{B_R} (2\sigma\mathcal{L}^*\sigma + \sigma^2|h|^2)dx\right) + \sum_{j=1}^{d}\left(\int_{B_R} 2h_j\sigma^2 dx\right)^2\Bigg] \tag{A.16}$$

Notice that

$$\begin{aligned} \int_{B_R} 2\sigma\mathcal{L}^*\sigma dx &= \int_{B_R} \sigma\sum_{i,j=1}^{d} \frac{\partial}{\partial x_i}\left(a^{ij}\frac{\partial\sigma}{\partial x_j}\right)dx - \int_{B_R} 2\sigma\sum_{i=1}^{d}\frac{\partial}{\partial x_i}(\tilde{f}_i\sigma)dx \\ &= -\int_{B_R}\sum_{i,j=1}^{d} a^{ij}\frac{\partial\sigma}{\partial x_i}\frac{\partial\sigma}{\partial x_j}dx + 2\int_{B_R}\sum_{i=1}^{d}\tilde{f}_i\sigma\frac{\partial\sigma}{\partial x_i}dx \\ &\leq -\lambda\int_{B_R}|\nabla\sigma|^2dx + 2\int_{B_R}\sum_{i=1}^{d}\frac{\tilde{f}_i\sigma}{\sqrt{\lambda}}\cdot\left(\sqrt{\lambda}\frac{\partial\sigma}{\partial x_i}\right)dx \\ &\leq -\lambda\int_{B_R}|\nabla\sigma|^2dx + \int_{B_R}\frac{1}{\lambda}\sum_{i=1}^{d}\tilde{f}_i^2\sigma^2dx + \lambda\int_{B_R}|\nabla\sigma|^2dx. \end{aligned} \tag{A.17}$$

Hence,

$$\begin{aligned} \frac{d}{dt}E\left(\int_{B_R}\sigma^2(t,x)dx\right)^2 \leq {} & 2E\Bigg[\left(\int_{B_R}\sigma^2dx\right)\left(\int_{B_R}\left(\frac{1}{\lambda}|\tilde{f}|^2+|h|^2\right)\sigma^2dx\right)\Bigg] \\ & + E\Bigg[\sum_{j=1}^{d}\left(\int_{B_R}2h_j\sigma^2dx\right)^2\Bigg] \end{aligned} \tag{A.18}$$

In the bounded domain B_R, there exists $M > 0$, such that

$$\frac{1}{\lambda}|\tilde{f}(x)|^2 + |h(x)|^2 \leq M,\ |h_j(x)| \leq M,\ \forall\ x \in B_R. \tag{A.19}$$

Thus,

$$\frac{d}{dt}E\left(\int_{B_R}\sigma^2(t,x)dx\right)^2 \leq (2M+4dM^2)E\left(\int_{B_R}\sigma^2(t,x)dx\right)^2 \tag{A.20}$$

According to Gronwall's inequality,

$$E\left(\int_{B_R}\sigma^2(T,x)dx\right)^2 \leq e^{(2M+4dM^2)T}\left(\int_{B_R}\sigma_0^2(x)dx\right)^2, \tag{A.21}$$

which is the desired result. □

References.

[1] R. Atar, *Exponential stability for nonlinear filtering of diffusion processes in a noncompact domain*, Annals of Probability, 26 (1998), pp. 1552–1574, https://doi.org/10.1214/aop/1022855873.

[2] A. Bain, *Fundamentals of Stochastic Filtering*, New York, NY : Springer New York : Imprint: Springer, 1st ed. 2009. ed., 2009.

[3] J. V. Candy, *Bayesian Signal Processing: Classical, Modern, and Particle Filtering Methods*, John Wiley & Sons, 2016.

[4] M. D. Chekroun, E. Park, and R. Temam, *The stampacchia maximum principle for stochastic partial differential equations and applications*, Journal of Differential Equations, 260 (2016), pp. 2926–2972, https://doi.org/https://doi.org/10.1016/j.jde.2015.10.022.

[5] K. Chen and J. Yu, *Short-term wind speed prediction using an unscented kalman filter based state-space support vector regression approach*, Applied energy, 113 (2014), pp. 690–705.

[6] X. Chen, Z. Sun, Y. Tao, and S. S.-T. Yau, *A uniform framework of yau–yau algorithm based on deep learning with the capability of overcoming the curse of dimensionality*, IEEE Transactions on Automatic Control, 70 (2025), pp. 339–354, https://doi.org/10.1109/TAC.2024.3424628.

[7] W. Dong, X. Luo, and S. S.-T. Yau, *Solving nonlinear filtering problems in real time by legendre galerkin spectral method*, IEEE Transactions on Automatic Control, 66 (2020), pp. 1559–1572.

[8] T. E. Duncan, *Probability densities for diffusion processes with applications to nonlinear filtering theory and detection theory*, PhD thesis, Stanford University, Stanford, California, May 1967.

[9] L. C. Evans, *Partial differential equations*, American Mathematical Society, 2nd ed.. ed., 2010.

[10] G. Galanis, P. Louka, P. Katsafados, I. Pytharoulis, and G. Kallos, *Applications of kalman filters based on non-linear functions to numerical weather predictions*, Ann. Geophys, 24 (2006), pp. 2451–2460.

[11] S. Godsill, A. Doucet, and M. West, *Maximum a posteriori sequence estimation using monte carlo particle filters*, Annals of the Institute of Statistical Mathematics, 53 (2001), pp. 82–96.

[12] C. Ichard, *Random media and processes estimation using non-linear filtering techniques: application to ensemble weather forecast and aircraft trajectories*, PhD thesis, Université de Toulouse, Université Toulouse III-Paul Sabatier, 2015.

[13] A. H. Jazwinski, *Stochastic processes and filtering theory*, Courier Corporation, 2007.

[14] J. Kang, A. Salmon, and S. S.-T. Yau, *Log-concave posterior densities arising in continuous filtering and a maximum a posteriori algorithm*, SIAM Journal on Control and Optimization, 61 (2023), pp. 2407–2424.

[15] I. Karatzas, *Brownian motion and stochastic calculus*, New York: Springer, 2nd ed., 1998.

[16] S. Li, Z. Wang, S. S.-T. Yau, and Z. Zhang, *Solving nonlinear filtering problems using a tensor train decomposition method*, IEEE Transactions on Automatic Control, (2022).

[17] X. Luo and S. S.-T. Yau, *Hermite spectral method to 1-d forward kolmogorov equation and its application to nonlinear filtering problems*, IEEE Transactions on Automatic Control, 58 (2013), pp. 2495–2507.

[18] R. E. Mortensen, *Optimal control of continuous-time stochastic systems.*, tech.

report, CALIFORNIA UNIV BERKELEY ELECTRONICS RESEARCH LAB, 1966.

[19] D. Ocone and E. Pardoux, *Asymptotic stability of the optimal filter with respect to its initial condition*, SIAM Journal on Control and Optimization, 34 (1996), pp. 226–243, https://doi.org/10.1137/S0363012993256617, https://doi.org/10.1137/S0363012993256617, https://arxiv.org/abs/https://doi.org/10.1137/S0363012993256617.

[20] M. Roth, G. Hendeby, C. Fritsche, and F. Gustafsson, *The ensemble kalman filter: a signal processing perspective*, EURASIP Journal on Advances in Signal Processing, 2017 (2017), pp. 1–16.

[21] B. L. Rozovsky, *Stochastic Evolution Systems Linear Theory and Applications to Non-Linear Filtering*, Cham : Springer, 2nd ed., 2018.

[22] S. Saha, P. K. Mandal, A. Bagchi, Y. Boers, and J. N. Driessen, *Particle based smoothed marginal map estimation for general state space models*, IEEE transactions on signal processing, 61 (2012), pp. 264–273.

[23] J. Sun, H. A. Blom, J. Ellerbroek, and J. M. Hoekstra, *Particle filter for aircraft mass estimation and uncertainty modeling*, Transportation Research Part C: Emerging Technologies, 105 (2019), pp. 145–162.

[24] Z. Wang, X. Luo, S. S.-T. Yau, and Z. Zhang, *Proper orthogonal decomposition method to nonlinear filtering problems in medium-high dimension*, IEEE Transactions on Automatic Control, 65 (2019), pp. 1613–1624.

[25] S.-T. Yau and S. S.-T. Yau, *Real time solution of nonlinear filtering problem without memory i*, Mathematical research letters, 7 (2000), pp. 671–693.

[26] S.-T. Yau and S. S.-T. Yau, *Real time solution of the nonlinear filtering problem without memory ii*, SIAM Journal on Control and Optimization, 47 (2008), pp. 163–195.

[27] M. Zakai, *On the optimal filtering of diffusion processes*, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 11 (1969), pp. 230–243.