Finite Dimensional Estimation Algebra for Time-Varying Filtering System and Optimal Transport Particle Filter: A Tangent Flow Point of View

JIAYI KANG
Tsinghua University, Beijing, China

XIAOPEI JIAO
Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, Huairou, China

STEPHEN S.-T. YAU, Life Fellow, IEEE
Tsinghua University, Beijing, China
Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, Huairou, China

Ever since the technique of the Kalman filter was popularized, there has been a lot of research interest in finding more classes of finite-dimensional recursive filters. In past research, the estimation algebra method could only be used for time-invariant systems. In this article, we extend the estimation algebra method so that it applies to a general class of time-varying filtering systems. Then, the Wei-Norman method can be used to derive the explicit solution of the posterior distribution of state estimation. As a special control law, tangent flow is derived for the nonlinear filtering system based on the Monge-Ampère equation in optimal transport. As a result, we propose an optimal transportation filter by applying stochastic tangent flow to Yau filtering systems. The numerical experiments demonstrate the higher efficacy and accuracy of the proposed optimal transportation filter compared to common traditional algorithms such as the extended Kalman filter and feedback particle filter.

Manuscript received 22 December 2022; revised 23 April 2023, 16 June 2023, and 12 July 2023; accepted 17 July 2023. Date of publication 4 August 2023; date of current version 8 December 2023.

DOI. No. 10.1109/TAES.2023.3299916

Refereeing of this contribution was handled by L. Chisci.

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 11961141005, in part by Tsinghua University start-up fund, and in part by the Tsinghua University Education Foundation fund under Grant 042202008.

Authors' addresses: Jiayi Kang is with the Department of Mathematical Sciences, Tsinghua University, Beijing 100084, China, E-mail: (kangjy19@mails.tsinghua.edu.cn); Xiaopei Jiao is with the Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, Huairou 101400, China, E-mail: (jiaoxiaopei0302@163.com); Stephen S.-T. Yau is with Department of Mathematical Science, Tsinghua University, Beijing 100084, China, and also with the Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, Huairou 101400, China, E-mail: (yau@uic.edu). *(Corresponding author: Stephen S.-T. Yau.)*

0018-9251

I. INTRODUCTION

In the 1960s, Kalman and Bucy [1], [2] first proposed linear filtering systems with Gaussian initial distributions. The general nonlinear filtering (NLF) problem is described by the following stochastic differential equations (SDE):

$$\begin{cases} dx_t = f(t, x_t)dt + \sigma_B(t, x_t)dB_t \\ dy_t = h(t, y_t)dt + dW_t \end{cases} \tag{1}$$

where $x_t \in \mathbb{R}^n$, $y_t \in \mathbb{R}^m$ are state and observation vectors, respectively, $f \in \mathbb{R}^n$, $\sigma_B \in \mathbb{R}^{n\times r}$, $h \in \mathbb{R}^m$ are C^∞ functions and B_t, W_t are Wiener processes of appropriate dimension with $E[dB_t dB_t^T] = \tilde{Q}(t)dt$, $\tilde{Q}(t) > 0$, $E[dW_t dW_t^T] = S(t)dt$, $S(t) > 0$. Here, we refer to $f(t, x_t)$ as the drift term, $\tilde{Q}(t)$, $S(t)$ as the covariance matrix of noises.

It is a central problem for the filtering algorithm to calculate the expectation of the conditional distribution $p(x_t|\mathcal{Y}_t)$, where $\mathcal{Y}_t$ is the σ-algebra generated by the process $\{y_s, 0 \le s \le t\}$. In the 1960s, Duncan, Mortensen, and Zakai [3], [4], [5] independently derived Duncan–Mortensen–Zakai (DMZ) equation, which is satisfied by the unnormalized conditional density. The DMZ equation can be further transformed into the robust form [6]. In the 1970s, Brockett and Clark [7], Brockett [8], and Mitter [9] independently proposed to construct finite-dimensional filters by using the estimation algebra method. Traditionally, the robust DMZ equation can be solved by the Wei-Norman approach [10] if one knows the explicit basis of estimation algebra. So seeking more filters with finite-dimensional estimation algebra (FDEA) is quite meaningful. Furthermore, there are many other ways to solve the DMZ equation. For example, Yau and Yau [11] proposed a new effective method to solve the “pathwise-robust” DMZ equation. Recently, an approximate real-time filtering algorithm was proposed to solve the robust DMZ equation based on “direct method” and Gaussian approximation [12], [13]. There are many other approximate algorithms to construct suboptimal filters to solve NLF problems such as the extended Kalman filter (EKF) [14], the unscented Kalman filter (UKF) [15], and particle filters (PF) [16]. Due to the importance of the estimation algebra method in NLF problems, a lot of work was focused on the classification of FDEA. Since the 1990s, in a series of works [17], [18], [19], [20], [21], [22], Yau and his collaborators have completely classified FDEAs with maximal rank on an arbitrary state space dimension, which include both Kalman–Bucy and Benés filtering systems as special cases. Since the 2000s, Yau and coworkers finished the complete classification of the estimation algebra of state dimension 2 and deeply studied nonmaximal rank FDEA with state dimension no more than four [23], [24], [25], [26], [27]. Recently, Jiao and Yau [28] made significant progress and constructed a new class of finite-dimensional filters with nonmaximal rank FDEA on an arbitrary state space dimension. However, exploring estimation algebra structure in the time-varying filtering system is still an open problem. The classification of FDEA made by Yau and his coworkers can naturally yield a formal explicit solution of the DMZ equation by the Wei-Norman method [29]. However, the formal solution derived by the Wei-Normal method is still hard to calculate numerically, especially in a high-dimensional situation.

With the consistent update and progress of sampling technology, the particle filter algorithm has been greatly developed in recent years. Several variants of the PF have emerged that employ resampling techniques to address the degeneracy problem. A framework for a gradual transition before the posterior was introduced in [30]. An important breakthrough came from feedback particle filter (FPF) [31]. The FPF applies a feedback structure to each particle and can be regarded as a generalization of the traditional linear regression filters such as the UKF or the smart-sampling Kalman filter [32].

Optimal transport (OT) is an old but very active research area. Starting from the original formulation of Monge, and the relaxation of Kantorovich [33], and following up with a sequence of important works, the OT has become a powerful tool in mathematics, physics, economics, engineering, and biology. Briefly speaking, OT deals with the problems of transporting from an initial distribution to a terminal distribution in a mass-preserving the manner with minimum cost. When the unit cost is the square of the Euclidean distance, the OT problem induces an extremely rich geometry for probability densities, especially Gaussian distribution [34]. And there are several filtering algorithms motivated by OT. Taghvaei and Mehta [35], [36], [37] proposed an optimal transportation method between prior and posterior distribution that can lead to unique control law under linear filtering with independent noise. Soon, the OT structure was extended to a linear filtering system with correlated noise in [38].

In this article, we first reformulate the controlled SDE for density evolution into the tangent flow equation based on optimal transportation using the Monge-Ampère equation. Generally speaking, it is hard to construct the corresponding tangent flow explicitly since it depends on the explicit expression of posterior densities. Second, we will extend the estimation algebra method to a new class of time-varying filtering systems so that the tangent flow can be calculated explicitly. And we shall prove this class of time-varying filtering system possesses FDEA and includes time-invariant Yau filter [17] as a special case. Finally, we explicitly calculate the tangent flow of the time-varying FDEA system. It can be considered as the nonlinear extension of [36]. The algorithm of the tangent flow for the general filtering problem is open. Taghvaei and Mehta [37] derived the corresponding tangent flow by first obtaining the exactness equation and then setting the optimal coupling condition to get the gradient form of control terms. Taghvaei and Mehta's series of works play an important role in the relation between nonlinear filtering and optimal transport. Inspired by their work about optimal transport filters, especially tangent flow, we exploit another new perspective to rederive deterministic/stochastic tangent flow. More specifically, our new derivation is based on the Monge-Ampère equation and uses the PDE expansion technique. The two derivations are ultimately equivalent. The benefit of our derivation is that it gives a connection between the estimation algebra and the tangent flow.

The map of optimal transportation used in this article is motivated by the several methods that are used in uncertainty propagation and Bayesian inference [39], [40], [41], [42].

The rest of this article is organized as follows. In Section II, we give some basic concepts including the filtering algorithm, estimation algebra, and optimal transportation. In Section III, the FDEA of a class of time-varying filtering systems is defined and an explicit solution of the DMZ equation is derived by the Wei-Norman method. In Section IV, tangent flow based on optimal transport is derived for the nonlinear filtering system. By applying stochastic tangent flow to a time-varying Yau filtering system, we proposed a new optimal transport filter. We gave a detailed analysis between the proposed method and the FPF. We present the numerical algorithm in Section IV, the numerical results in Section V. Finally, Section VI concludes this article.

II. BASIC CONCEPTS AND PRELIMINARY RESULTS

In this article, we use $\|\cdot\|_2$ to represent the L_2 norm of the vectors or the matrices, $\|\cdot\|_F$ to describe the Frobenius norm of the matrices. The $\mathrm{Tr}(A)$ is the matrix/operator trace. The $\mathbb{R}$ is the real-number field. The $\mathbb{E}[\cdot]$ is expectation operator. The space of continuous and smooth functions on $\mathbb{R}^n$ is denoted as $C^\infty(\mathbb{R}^n)$. The ∇ is the gradient operator of $C^\infty(\mathbb{R}^n)$ and $\nabla\cdot()$ is the divergence operator of $C^\infty(\mathbb{R}^n)$. And $\Delta(*) := \nabla\cdot(\nabla(*))$ denotes as the Laplace operator. However, the Δt denotes the small time step. If A and B are differential operators, the Lie bracket of A and B, $[A, B]$ is defined by $[A, B]\phi = A(B\phi) - B(A\phi)$ for any $C^\infty(\mathbb{R}^n)$ function ϕ. We denote $\circ$ as Stratonovich integral. Here, we define two new operators, $\bar{\nabla}(*)$ and $\bar{\nabla}\cdot(*)$, as follows:

$$\bar{\nabla}(\varphi_0(t,x)) := (\nabla\varphi_0^1(t,x), \quad \ldots, \quad \nabla\varphi_0^m(t,x))$$

where $\varphi_0(t,x) := (\varphi_0^1(t,x), \quad \ldots, \quad \varphi_0^m(t,x))$ is a m dimensional row vector value functions

$$\bar{\nabla}\cdot(K) := (\nabla\cdot(\mathcal{K}^1(t,x)), \quad \ldots, \quad \nabla\cdot(\mathcal{K}^m(t,x)))$$

where $\mathcal{K}^i(t,x)$ with $1 \leq i \leq m$ are all n dimensional column vector fields.

Combining the two new operators, we can define $\bar{\Delta}(*)$ as

$$\begin{aligned}\bar{\Delta}\varphi_0(t,x) &:= \bar{\nabla}\cdot(\bar{\nabla}\varphi_0(t,x))\\ &= (\Delta\varphi_0^1(t,x), \quad \ldots, \quad \Delta\varphi_0^m(t,x).\end{aligned}$$

A. DMZ Equation and Robust Form

In this section, we first introduce a classical filtering algorithm. In the continuous time-varying filtering system (1), we assume that $G(t) := \sigma_B\tilde{Q}(t)\sigma_B^T$ is a positive definite matrix. The posterior density $p(t,x)$ of X_t conditioned on the observation history $\mathcal{F}_t$ satisfies the Kushner equation [43] as follows:

$$\begin{cases} dp(t,x) = \mathcal{L}p(t,x)dt + (h(t,x) - \hat{h}_t)^T \\ \qquad\qquad \times S^{-1}(t)p(t,x) \circ (dy_t - \hat{h}_t dt) \\ p(0,x) = p_0(x) \end{cases} \tag{2}$$

where

$$\begin{aligned} \mathcal{L}(*) := & \frac{1}{2}\sum_{i,j=1}^{n} G_{ij}(t)\frac{\partial^2}{\partial x_i \partial x_j}(*) - \sum_{i=1}^{n} f_i \frac{\partial}{\partial x_i}(*) \\ & - \left(\sum_{i=1}^{n}\frac{\partial f_i}{\partial x_i} - \frac{1}{2}(h(t,x) - \hat{h}_t)^T \right. \\ & \left. \times\ S^{-1}(h(t,x) - \hat{h}_t) - c(t)\right)(*) \end{aligned} \tag{3}$$

and $\hat{h}_t = \int_{\mathbb{R}^n} h(t,x)p(t,x)dx$ and $c(t) = \int_{\mathbb{R}^n} \frac{1}{2}(h(t,x) - \hat{h}_t)^T S^{-1}(h(t,x) - \hat{h}_t)p(t,x)dx$, which can be considered as a normalizing constant.

The unnormalized density function $\sigma(t,x)$ of X_t conditioned on the observation history $\mathcal{F}_t$ satisfies the DMZ equation as follows [3], [4], [5]:

$$\begin{cases} d\sigma(t,x) = L_0\sigma(t,x)dt + h^T(t,x)S^{-1}(t)\sigma(t,x) \circ dy_t \\ \sigma(0,x) = \sigma_0(x) \end{cases} \tag{4}$$

where

$$\begin{aligned} L_0 := & \frac{1}{2}\sum_{i,j=1}^{n} G_{ij}(t)\frac{\partial^2}{\partial x_i \partial x_j} - \sum_{i=1}^{n} f_i \frac{\partial}{\partial x_i} \\ & - \sum_{i=1}^{n}\frac{\partial f_i}{\partial x_i} - \frac{1}{2}h^T(x,t)S^{-1}h(x,t). \end{aligned} \tag{5}$$

The DMZ equation can be transformed into a PDE with stochastic coefficients by using the following robust exponential transformation

$$\begin{aligned} u(t,x) = & \exp[-K(t,x)]\sigma(t,x) \\ K(x,t) = & h^T(x,t)S^{-1}(t)y_t. \end{aligned} \tag{6}$$

By using the robust transformation, we can obtain the robust DMZ equation as

$$\begin{cases} \frac{\partial u}{\partial t}(t,x) = L_0 u(t,x) + [L_0, K(t,x)]u(t,x) \\ + \frac{1}{2}[[L_0, K(t,x)], K(t,x)]u(t,x) \\ - \frac{\partial h^T(t,x)S^{-1}}{\partial t} y_t u(t,x) \\ u(0,x) = \sigma_0(x) \end{cases}$$

where $K(t,x) = h^T(t,x)S^{-1}(t)y_t$, $[L_0, K(t,x)] = \sum_{i,j=1}^{n} \frac{\partial K}{\partial x_i} G_{i,j} \frac{\partial}{\partial x_j} - \sum_{i=1}^{n} f_i \frac{\partial K}{\partial x_i}$, and $[[L_0, K(t,x)], K(t,x)] = \sum_{i,j=1}^{n} G_{i,j} \frac{\partial K}{\partial x_i}\frac{\partial K}{\partial x_j}$.

B. Estimation Algebra of Time-Invariant System

Next, we will introduce some basic concepts about the estimation algebra of a time-invariant system. In the following, in (1), we assume B_t, W_t are mutually independent standard Wiener processes and σ_B is a constant orthogonal matrix.

Then DMZ equation (4) will become (Stratonovich form) as follows:

$$\begin{cases} d\sigma(t,x) = L_0\sigma(t,x)dt + \sum_{i=1}^{m} L_i\sigma(t,x) \circ dy_i(t) \\ \sigma(0,x) = \sigma_0(x) \end{cases} \tag{7}$$

where

$$L_0 = \frac{1}{2}\sum_{i=1}^{n}\frac{\partial^2}{\partial x_i^2} - \sum_{i=1}^{n} f_i\frac{\partial}{\partial x_i} - \sum_{i=1}^{n}\frac{\partial f_i}{\partial x_i} - \frac{1}{2}\sum_{i=1}^{m} h_i^2 \tag{8}$$

and L_i is the zero degree differential operator of multiplication by h_i, $1 \le i \le m$.

Let

$$D_i = \frac{\partial}{\partial x_i} - f_i, \quad \eta = \sum_{i=1}^{n}\left(\frac{\partial f_i}{\partial x_i} + f_i^2\right) + \sum_{i=1}^{m} h_i^2. \tag{9}$$

So that

$$L_0 = \frac{1}{2}\left(\sum_{i=1}^{n} D_i^2 - \eta\right). \tag{10}$$

The robust DMZ equation (7) will have the following form:

$$\begin{cases} \frac{\partial u}{\partial t}(t,x) = & L_0 u(t,x) + \sum_{i=1}^{m} y_i(t)[L_0, L_i]u(x,t) \\ & + \frac{1}{2}\sum_{i,j=1}^{m} y_i(t)y_j(t)[[L_0, L_i], L_j]u(x,t) \\ u(0,x) = & \sigma_0(x). \end{cases} \tag{11}$$

In the following, we give some basic concepts about estimation algebra.

Definition 2.1 A vector space $\mathcal{F}$ with the Lie bracket operation $\mathcal{F} \times \mathcal{F} \to \mathcal{F}$ denoted by $(x,y) \mapsto [x,y]$ is called a Lie algebra if the following axioms are satisfied.

(1) The Lie bracket operation is bilinear.

(2) $[x,x] = 0$ for all $x \in \mathcal{F}$.

(3) $[x,[y,z]] + [y,[z,x]] + [z,[x,y]] = 0$, $x, y, z \in \mathcal{F}$.

Definition 2.2 The estimation algebra E of a time-invariant filtering system is defined to be the Lie algebra generated by $\{L_0, L_1, \ldots, L_m\}$ with coefficient as real number $\mathbb{R}$, i.e., $E = \langle L_0, h_1, \ldots, h_m\rangle_{L.A.}$.

Definition 2.3 Let $L(E) \subset E$ be the vector space consisting of all the homogeneous degree 1 polynomial in E. Then the linear rank of estimation algebra E is defined by $r := \dim L(E)$. If r is equal to the dimension of state space, we call E has maximal rank.

Definition 2.4 ([19]) The Wong matrix is a antisymmetric matrix $\mathcal{O} = (\omega_{i,j})$ with $\frac{\partial f_i}{\partial x_j} - \frac{\partial f_j}{\partial x_i} = \omega_{j,i}$, $1 \le i, j \le n$.

C. Optimal Transportation

Finally, we introduce the basic concepts of the OT.

Let α and β be two probability measures on measure spaces Ω_X and Ω_Y, respectively. $\mathcal{P}(\Omega)$ denotes the set of probability measures on Ω. Let $c : \Omega_X \times \Omega_Y \to \mathbb{R}^+$ be a cost function and $c(x,y)$ measures the cost of transporting one unit of mass from $x \in \Omega_X$ to $y \in \Omega_Y$. The transport map is defined below.

DEFINITION 2.5 Let $\alpha \in P(\Omega_X)$ and $\beta \in P(\Omega_Y)$. We say that T is a transport map from α to β if

$$\beta(B) = \alpha\left(\mathcal{T}^{-1}(B)\right) \quad \text{for all } \beta\text{-measurable sets } B \subset \Omega_Y. \tag{12}$$

Equivalently, we write

$$\beta = \mathcal{T}_{\#}\alpha. \tag{13}$$

Monge's optimal transportation problem is formulated as follows.

THEOREM 2.1 MONGE'S OPTIMAL TRANSPORTATION PROBLEM [44] Given $\alpha \in \mathcal{P}(\Omega_X)$, $\beta \in \mathcal{P}(\Omega_Y)$, let

$$I[\mathcal{T}] = \int_{\Omega_X} c(x, \mathcal{T}(x))\mathrm{d}\alpha(x) \tag{14}$$

where $\mathcal{T} : \Omega_X \to \Omega_Y$ is a transport map from α to β, i.e., $\beta = \mathcal{T}_{\#}\alpha$. Then Monge's optimal transportation problem is to minimize the above integral among all transport maps from α to β.

If we further assume that the density functions of α and β are C^2 smooth, the optimal transportation map is the gradient form of some function Φ, i.e., $\nabla\Phi(x) = \mathcal{T}(x)$. The Φ function is determined by the following equation, which is the so-called Monge-Ampère equation [44]:

$$\det \nabla^2\Phi(x) = \frac{\beta_0(x)}{\alpha_0(\nabla\Phi(x))} \tag{15}$$

where $\nabla^2\Phi(x)$ is the Hessian matrix and α_0, β_0 are the density functions of α and β, respectively [45].

III. OPTIMAL TRANSPORTATION FILTERING ALGORITHM

In this section, we shall reformulate the known probability flow satisfied by controlled SDE as a tangent flow based on OT. Tangent flow is a type of gradient flow and it is hard to calculate explicitly for a general filtering system (1).

A. Optimal Transportation in SPDE

The key step of the Yau-Yau algorithm [46] is to discretize the time interval so that the PDE of posterior density between time steps can be established.

In this subsection, we shall first derive the tangent flow for the PDE (16) based on the Monge-Ampère equation in optimal transport

$$\frac{\partial p}{\partial t} = \mathcal{D}(p), \quad t \in [0, S] \tag{16}$$

where $\mathcal{D}(\cdot)$ is some differential operator and

$$\begin{aligned} p(t,\cdot) &\in \mathcal{P}^2(\mathbb{R}^d) \\ &:= \left\{q \text{ is a distribution on } \mathbb{R}^d | x_0 \in \mathbb{R}^d \right. \\ &\left. \text{s.t. } \mathbb{E}_q\left[\|x - x_0\|_2^2\right] < \infty\right\} \end{aligned}$$

$\mathcal{P}^2(\mathbb{R}^d)$ is the so-called Wasserstein space equipped with the well-known Wasserstein distance W^2 [44].

Considering the optimal transport problem between two distributions $p(t, x)$ and $p(t + \Delta t, x)$, we will get the following Monge-Ampère equation by letting the $\beta_0(x) = p(t + \Delta t, x)$ and $\alpha_0(x) = p(t, x)$:

$$\det \nabla^2\Phi_t = \frac{p(t + \Delta t, x)}{p(t, \nabla\Phi_t(x))}. \tag{17}$$

The Monge-Ampère equation is a nonlinear PDE that is hard to solve, but luckily the $p(t, x)$ and $p(t + \Delta t, x)$ are close when Δt is small. So $\nabla^2\Phi_t$ should be close to I_n, we can expand both sides of (17) in the asymptotic sense of $\Delta t \to 0$. First, we can expand $p(t + \Delta t, x)$ according to the following (16):

$$p(t + \Delta t, x) = p(t, x) + \mathcal{D}(p(t, x))\Delta t + O(\Delta t^2). \tag{18}$$

Divide $p(t, x)$ for both sides, which yields

$$\frac{p(t + \Delta t, x)}{p(t, x)} = 1 + \frac{\mathcal{D}(p(t, x))}{p(t, x)}\Delta t + O(\Delta t^2). \tag{19}$$

And, we consider expanding $\Phi_t(x)$. Since $\nabla^2\Phi_t$ should be close to I_n, then $\Phi_t(x)$ should be close to $\frac{|x|^2}{2}$. So, we have

$$\Phi_t(x) = \frac{|x|^2}{2} + \varphi_1(t, x) \cdot \Delta t + O(\Delta t^2) \tag{20}$$

where φ_1 is an undetermined function. So, submit (20) and (19) into (17), the left-hand side can be rewritten as

$$\begin{aligned} \det \nabla^2\Phi^{(n)}(t, x) = &\det(I_n + \nabla^2\varphi_1(t, x_t) \cdot \Delta t + O(\Delta t^2)) \\ &= 1 + \Delta\varphi_1(t, x_t) \cdot \Delta t + O(\Delta t^2)) \end{aligned} \tag{21}$$

where the second equality holds since Lemma A.1 in Appendix. And, the right-hand side can be rewritten as

$$\begin{aligned} &\frac{p(t + \Delta t, x)}{p(t, \nabla\Phi_t(x))} = \frac{p(t + \Delta t, x)}{p(t, x)} \cdot \frac{p(t, x)}{p(t, \nabla\Phi_t(x))} \\ &= \left(1 + \frac{\mathcal{D}(p(t, x))}{p(t, x)}\Delta t + O(\Delta t^2)\right) \\ &\cdot \left(1 - \frac{1}{p(t, x)}\nabla p(t, x)^T(\nabla\Phi_t(x) - x) + O(\Delta t^2)\right) \\ &= 1 + \frac{\mathcal{D}(p(t, x)}{p(t, x)}\Delta t - \frac{1}{p(t, x)}\nabla p(t, x)^T \\ &\times \nabla\varphi_1(t, x)\Delta t + O(\Delta t^2) \end{aligned} \tag{22}$$

where

$$\nabla\Phi_t(x) = x + \nabla\varphi_1(t, x)\Delta t + O(\Delta t^2). \tag{23}$$

We can combine the two sides (21) and (22), and take $\Delta t \to 0$, then we can get

$$\Delta\varphi_1(t, x) + (\nabla \log p(t, x))^T\nabla\varphi_1(t, x) = -\frac{\mathcal{D}(p(t, x))}{p(t, x)}. \tag{24}$$

Thus, we get optimal transport $\nabla\Phi_t(x)$ in Monge's optimal transportation problem in Theorem 2.1. We can design probability flow $x_{t+\Delta t} = \nabla\Phi_t(x_t)$ and by using (23) which leads to

$$x_{t+\Delta t} = x_t + \nabla\varphi_1(t, x_t)\Delta t + O(\Delta t^2). \tag{25}$$

In the asymptotic sense for (25), we get tangent flow

$$dx_t = \nabla\varphi_1(t,x)dt \tag{26}$$

where $x_0 \sim p(0,x)$. It is noticed that the posterior distribution of (26) satisfies the following Fokker-Planck equation:

$$\frac{\partial p(t,x)}{\partial t} = -\nabla\cdot(\nabla\varphi_1(t,x)p(t,x)). \tag{27}$$

And, combining with (24), we have

$$\begin{aligned}\frac{\partial p(t,x)}{\partial t} &= -\nabla\cdot(\nabla\varphi_1(t,x)p(t,x))\\ &= -\Delta\varphi_1(t,x)p(t,x) - \nabla\varphi_1(t,x)\cdot\nabla p(t,x)^T\\ &= -p(t,x)\left(-\frac{\mathcal{D}(p(t,x))}{p(t,x)}\right)\\ &= \mathcal{D}(p(t,x)).\end{aligned} \tag{28}$$

So, we show that the densities of dynamical system (26) satisfy (16). So, we call that (26) is the tangent flow for (16).

In the following, we can extend the tangent flow to the situation where density evolution satisfies an SPDE:

$$dp = \mathcal{D}(p)dt + \mathcal{H}(p)\circ dI_t \quad t\in[0,S] \tag{29}$$

where $\mathcal{D}(\cdot)$ and $\mathcal{H}(\cdot)$ are some differential operators, and I_t is $m-$dimensional Gaussian process. Generally speaking, I_t is multidimensional, and $\mathcal{H}(p)$ is a m-dimensional row vector function value operator.

Similarly, we consider a discretization of (29). Notice that if we want to apply the Monge-Ampere equation, $p(t+\Delta t,x)$ and $p(t,x)$ should be determined explicitly, which means that the path of dI_t is given in the analysis. In the following, we take the realization of a stochastic process $I_t = I_t(\omega)$ by fixing ω. For a time sequence $\{0=t_0\le t_1\le\cdots\le t_{2^n}=S\}$ with $n\in\mathbb{Z}^+$ and $t_i = iS2^{-n}$, we approximate the $\frac{dI_t}{dt}\approx \dot{I}_t^{(n)} = 2^n(I_{t_{i+1}}-I_{t_i})$ for $t\in[t_i,t_{i+1})$. So, we have

$$\frac{\partial p^{(n)}}{\partial t} = \mathcal{D}(p^{(n)}) + \mathcal{H}(p^{(n)})\cdot\dot{I}_t^{(n)} \tag{30}$$

where the $\dot{I}_t^{(n)} = 2^n(I_{t_{i+1}}-I_{t_i})$ if $t\in[t_i,t_{i+1})$. As before, we need to solve (17) and assume its solution has the following form:

$$\tilde{\Phi}_t(x) = \frac{|x|^2}{2} + \varphi_1(t,x)\cdot\Delta t + \varphi_0(t,x)\dot{I}_t^{(n)}\Delta t + O(\Delta t^2) \tag{31}$$

where $\varphi_1, \varphi_0 = (\varphi_0^1,\ldots,\varphi_0^m)$ are undetermined scale and vector value functions. Similarly with (26), after getting solution $\tilde{\Phi}_t$, by using Theorem 2.1, we can get tangent flow (30)

$$\frac{dx_t^{(n)}}{dt} = \nabla\varphi_1\left(t,x_t^{(n)}\right) + \bar{\nabla}\varphi_0\left(t,x_t^{(n)}\right)\cdot\dot{I}_t^{(n)} \tag{32}$$

where the $\dot{I}_t^{(n)} = 2^n(I_{t_{i+1}}-I_{t_i})$ if $t\in[t_i,t_{i+1})$ and $x_{t_i}^{(n)} = x$. Here, the $\bar{\nabla}\varphi_0(t,x_t^{(n)})$ is a slightly abuse of symbols. In fact $\bar{\nabla}\varphi_0(t,x) = (\nabla\varphi_0^1(t,x),\ \cdots,\ \nabla\varphi_0^m(t,x))$. In the following, we substitute the specific form of $\tilde{\Phi}_t$ to (17) and expand the right-hand side of (17) as a series in terms of Δt:

$$\begin{aligned}\det\nabla^2\tilde{\Phi}^{(n)}(t,x_t^{(n)}) = 1 &+ \Delta\varphi_0(t,x_t^{(n)})\cdot\Delta I_t\\ &+ \Delta\varphi_1(t,x_t^{(n)})\cdot\Delta t + O(\Delta t^2)\end{aligned} \tag{33}$$

where (33) holds by using Lemma A.1 in Appendix. And,

$$\begin{aligned}\frac{p^{(n)}(t+\Delta t,x)}{p^{(n)}(t,x_{t+\Delta t}^{(n)})} &= \frac{p^{(n)}(t+\Delta t,x)}{p^{(n)}(t,x)}\frac{p^{(n)}(t+\Delta t,x)}{p^{(n)}(t,x_{t+\Delta t}^{(n)})}\\ &= \left(1-\frac{\nabla p^{(n)}(t,x_t^{(n)})}{p^{(n)}(t,x_t^{(n)})}\frac{dx_t^{(n)}}{dt}\Delta t + O(\Delta t^2)\right)\\ &\quad\cdot\left(1+\frac{\mathcal{D}(p^{(n)})}{p^{(n)}}\Delta t + \frac{\mathcal{H}(p^{(n)})}{p^{(n)}}\cdot\dot{I}_t^{(n)}\Delta t + O(\Delta t^2)\right)\\ &= 1+\frac{\mathcal{D}(p^{(n)})}{p^{(n)}}\Delta t + \frac{\mathcal{H}(p^{(n)})}{p^{(n)}}\cdot\dot{I}_t^{(n)}\Delta t\\ &\quad - \frac{\nabla p^{(n)}(t,x_t^{(n)})}{p^{(n)}(t,x_t^{(n)})}\frac{dx_t^{(n)}}{dt}\Delta t + O(\Delta t^2).\end{aligned} \tag{34}$$

Then, we take the limit $n\to\infty$ which put $\Delta t\to 0$, the ODE (32) will converge to SDE according to Wong-Zakai approximation as follows [47]:

$$dx_t = \nabla\varphi_1(t,x_t)dt + \nabla\varphi_0(t,x_t)\circ dI_t. \tag{35}$$

And, the left-hand side of (17) will become the following:

$$\lim_{n\to\infty}\det\nabla^2\tilde{\Phi}_t^{(n)}(x_t^{(n)}) = 1 + \bar{\Delta}\varphi_0(t,x_t)\circ dI_t + \Delta\varphi_1(t,x_t)dt \tag{36}$$

the right-hand side of (17) will become the following:

$$\begin{aligned}&\frac{p(t,x)+dp(t,x)}{p(t,x)}\\ &= 1+\frac{1}{p(t,x)}\big[\mathcal{D}(p(t,x)dt + \mathcal{H}(p(t,x))\circ dI_t\\ &\quad - \nabla p(t,x)^T\left(\nabla\varphi_1(t,x)dt + \bar{\nabla}\varphi_0(t,x)\circ dI_t\right)\big].\end{aligned} \tag{37}$$

Therefore, substituting (36) and (37) to (17), we will get the constraint equation satisfied by φ_1 and φ_0, we summarized the result in the following Theorem.

Theorem 3.1 (Optimal Transportation Condition for the SPDE) Let $p(t,x)$ be the solution of (29). Then, the tangent flow for general SPDE (29) is

$$dx_t = \nabla\varphi_1(t,x_t)dt + \bar{\nabla}\varphi_0(t,x_t)\circ dI_t. \tag{38}$$

Here, $\varphi_0 = (\varphi_0^1,\ldots,\varphi_0^m)$ is the vector value function and satisfies the following:

$$\bar{\Delta}\varphi_0(t,x) + \nabla(\log p(t,x))^T\bar{\nabla}\varphi_0(t,x) = -\frac{\mathcal{H}(p)}{p} \tag{39}$$

where $\mathcal{H}$ is the m dimensional row vector operator.

Then, φ_1 is the scale function and satisfies the following:

$$\Delta\varphi_1(t,x) + \nabla(\log p(t,x))^T\nabla\varphi_1(t,x) = -\frac{\mathcal{D}p}{p}. \tag{40}$$

The tangent flow of Kushner (2) can be constructed as follows.

REMARK 1 (OPTIMAL TRANSPORTATION CONDITION FOR THE KUSHNER EQUATION) Let $p(t,x)$ be the solution of (2). By using Theorem 3.1, we assume that $\mathcal{H}(p) = (h(t,x) - \hat{h}_t)^T S^{-1}(t)p$ and $\mathcal{D}p = \mathcal{L}p$, where $\mathcal{L}$ is the operator defined in Kushner (2) and $\hat{h}_t = \int_{\mathbb{R}^n} h(t,x)p(t,x)dx$. Then, the tangent flow for Kushner (2) is

$$dx_t = \nabla\varphi_1(t,x_t)dt + \bar{\nabla}\varphi_0(t,x_t) \circ dI_t \tag{41}$$

where

$$\begin{aligned}\bar{\Delta}\varphi_0(t,x) + \nabla(\log p(t,x))^T \bar{\nabla}\varphi_0(t,x) \\ = (h(t,x) - \hat{h}_t)^T S^{-1}(t)\end{aligned} \tag{42}$$

and

$$\Delta\varphi_1(t,x) + \nabla(\log p(t,x))^T \nabla\varphi_1(t,x) = -\frac{\mathcal{L}p}{p}. \tag{43}$$

REMARK 2 The right-hand side of (42) and (43) is only determined by the estimation algebra. As the form of (42) and (43), the system of OT maps between the posterior distributions is highly relied on the estimation algebra due to the explicit expression for $p(t,x)$.

B. Optimal Transportation Corresponding Dynamic Flow

In the last subsection, we proposed a new way to optimally transfer the posterior distribution. In recent years, there has been another dynamic-based PF motivated by the mean-field control, which is the so-called FPF.

Similar to MCMCs [48], [49], although FPF [50] has the guarantee of asymptotic accuracy, the effect of their limited number of samples is usually difficult to guarantee in practice, and due to the stochastic simulation and the cross-correlation, its convergence is relatively slow [51]. Sampling from high-dimensional distribution, even Gaussian distribution, is slow, which violates the real-time requirements of high-dimensional filtering algorithms.

Numerous studies show that feedback control law is not unique. Taghvaei and Mehta [35] proposed an optimal transportation method between prior and posterior distribution, which can lead to a unique control law in a linear setting. Furthermore, Taghvaei and Mehta [36] proposed the optimal transportation formulation of EnKF which can construct a unique control law with noises. Taghvaei and Mehta [37] summarized the development from the viewpoint of optimal coupling. General-controlled SDE is formulated as follows:

$$dx_t = \mathcal{U}(t,x_t)dt + \mathcal{K}(t,x_t) \circ dI_t \tag{44}$$

where $\mathcal{K}(t,x) := (\mathcal{K}^1(t,x), \ldots, \mathcal{K}^m(t,x))$ is a $n \times m$-dimensional vector function.

LEMMA 3.1 (FOKKER-PLANCK DENSITY EQUATION) The forward Fokker-Planck density equation of (44) is determined by the following SPDE:

$$\begin{aligned}dp = -\nabla \cdot (\mathcal{U}(t,x_t)p(t,x))dt \\ - \bar{\nabla} \cdot (\mathcal{K}(t,x)p(t,x)) \circ dI_t\end{aligned} \tag{45}$$

where $\bar{\nabla} \cdot (\mathcal{K}(t,x)p(t,x)) := (\nabla \cdot (\mathcal{K}^1(t,x_t)p(t,x)), \ldots, \nabla \cdot (\mathcal{K}^m(t,x)p(t,x)))$.

Taghvaei et al. [37] gave a comprehensive review of FPF. They derived the exactness condition satisfied by the control term and from the optimization of optimal coupling revealed that control terms must be in gradient form. This result is important to design filtering algorithms for different nonlinear systems.

THEOREM 3.2 (FORWARD CONSISTENCY CONDITION [37]) Consider the process x_{t_i} that evolves according to (44). If the forward Fokker-Planck equation is exactly the Kushner equation of the filtering system. Then, the control terms satisfy the following:

$$\mathcal{L}p(t,x) = -\nabla \cdot (\mathcal{U}(t,x)p(t,x)) \tag{46}$$

and

$$\bar{\nabla} \cdot (\mathcal{K}(t,x)p(t,x)) = -(h(t,x) - \hat{h}_t)^T S^{-1}(t)p(t,x). \tag{47}$$

Due to that, the solutions of (46) and (47) are not unique. There is a natural requirement for the control term, which is to minimize the second-order moment

$$\min_{\mathcal{U}(t,x)} \int |\mathcal{U}(t,x)|^2 p(t,x)dx \tag{48}$$

$$\min_{\mathcal{K}(t,x)} \int \|\mathcal{K}(t,x)\|_F^2 p(t,x)dx \tag{49}$$

where the conditions (46) and (47) are needed, respectively. The optimal solutions of (48) are some gradient form [52].

REMARK 3 It is important to find that if we assume that $\nabla\varphi_0(t,x) = \mathcal{K}(t,x)$ and the $\nabla\varphi_1(t,x) = \mathcal{U}(t,x)$, then the (46) and (47) will become (42) and (43), respectively. The optimal solutions of (48) are exactly the optimal transportation conditions in Theorem 1.

C. Locally Optimal Transportation Method and General Dynamic Flow

In this subsection, we will summarize many control version PF methods into a uniform framework. There is a bunch of dynamic flows and their posterior distributions are exactly the posterior distributions of the filtering system, which is

$$dX_t^i = \underbrace{f(t,X_t^i)dt + \sigma_B(t,X_t^i)dB_t^i}_{\text{diffusion}} + \overbrace{\mathcal{K}(t,X_t^i) \circ dI_t}^{\text{update}} \tag{50}$$

such as FPF [50], Xiong's filter [53], and Reich filter [54].

The relationship with the Poisson equation in FPF is through the ensemble transform filter, which relies on a linear programming construction to approximate the optimal transport map. As discussed, the solution of the Poisson equation yields an optimal transport map from the "prior" to the "posterior." So, FPF is used as the optimal transportation method in the update step [55]. The new framework we proposed uses optimal transportation not only in the update step but also in the diffusion step.

In this section, we have established the conditions for the tangent flow. Now, we need to design numerical algorithms based on the tangent flow. However, the conditions for the tangent flow depend on the posterior density $p(t, x)$. Therefore, we are interested in filtering systems that allow for explicit representations of the posterior densities, i.e., the FDEA systems. Using the Wei-Norman approach, one can construct the posterior densities of the FDEA system explicitly. However, traditional FDEA systems are time-invariant, and in the next section, we shall extend the FDEA system to time-varying cases.

IV. Estimation Algebra in Time-Varying System

In this section, we aim to extend the estimation algebra structure to more general time-varying Yau filtering systems. This will allow us to construct the tangent flow of the general FDEA system.

First, we shall introduce the normalizing tricks. Since $G(t)$ is positive definite, $G(t)$ can be decomposed to $FF^T(t)$ by Cholesky decomposition, where $F(t)$ is a nonsingular matrix. Setting $F(t)z_t = x_t$, we obtain the SDE for z_t

$$dz_t = F^{-1}(t)f(t, F(t)z_t)dt + \frac{dF(t)}{dt}F^{-1}(t)z_t dt + F^{-1}(t)\sigma_B(t)dB_t.$$

It is easy to check that $F^{-1}(t)\sigma_B(t)dB_t$ is standard white noise. So, we can denote $F^{-1}(t)\sigma_B(t)dB_t = d\tilde{B}_t$, where $\tilde{B}_t$ is n-dimensional standard Brownian motion. Here, we can define a normalizing filtering system as

$$\begin{cases} dz_t = \tilde{f}(t, z_t)dt + d\tilde{B}_t \\ dy_t = \tilde{h}(t, z_t)dt + dW_t \end{cases} \tag{51}$$

where $\tilde{f}(t, z) := F^{-1}(t)f(t, F(t)z_t) + \frac{dF(t)}{dt}F^{-1}(t)z_t$ and $\tilde{h}(t, z_t) := h(t, F(t)z_t)$. Since the $F(t)z_t = x_t$ holds for all t, then The posterior distributions of z_t and x_t are mutually transformable by variable substitution. Thus, we can focus on the system modified by the normalizing trick to generalize the estimation algebra.

A. Conclusions and Main Assumptions of the Estimation Algebra in the Time-Varying System

In this subsection, we shall present the main assumptions of the general estimation algebra.

Assumption 4.1 The system function f satisfies

$$f(t, x) = L(t)x + l(t) + F(t)\nabla\phi(F(t)^{-1}x) \tag{52}$$

where $L(t) = (l_{ij}(t))$, $1 \le i, j \le n$, $l^T(t) = (l_1(t), \ldots, l_n(t))$. $F(t)$ is an invertible matrix satisfying $F(t)F^T(t) = G(t)$. $\phi(x)$ is a C^∞ function on $\mathbb{R}^n$.

Assumption 4.2 The observation function $h(t, x)$ is a linear function in variable x, i.e.,

$$h(t, x) = H(t)x \tag{53}$$

where $H(t) \in \mathbb{R}^{m\times n}$ has rank m.

Assumption 4.1 is equivalent to say that

$$\begin{aligned}\tilde{f}(t, z) &= F^{-1}f(t, Fz) + \frac{dF^{-1}}{dt}Fz \\ &= \tilde{L}z + \tilde{l} + \nabla_z\phi(z), \quad \tilde{L} = F^{-1}LF + \frac{dF^{-1}}{dt}F, \\ \tilde{l} &= F^{-1}l.\end{aligned}$$

Assumption 4.2 is equivalent to say that

$$\tilde{h}(t, z) := \tilde{H}(t)z = H(t)F(t)z.$$

Assumption 4.1 contains all the previous FDEA system with maximum rank as special cases.

Assumption 4.3 $\sum_{i=1}^n(\tilde{f}_i^2 + \frac{\partial \tilde{f}_i}{\partial z_i})$ is quadratic with respect to z, where $\tilde{f}_i$ is ith component of the $\tilde{f}(t, z)$.

And, in Lemma 4.2, we prove that if Assumptions 4.1, 4.2, and 4.3 are satisfied, then the estimation algebra associated with this filtering system will be finite-dimensional.

B. Derivation of the General Estimation Algebra

To derive estimation algebra for the time-varying system, we first give an important lemma for the analysis.

Lemma 4.1 [17] If A and B are two differential operators where the order of A is less than 1, then the following equation holds:

$$\begin{aligned} e^{-A}Be^{A} &= B + [B, A] + \frac{1}{2!}[[B, A], A] + \cdots \\ &\quad + \frac{1}{n!}[[[B, A], A]\cdots] + \cdots \\ e^{A}Be^{-A} &= B + [A, B] + \frac{1}{2!}[A, [A, B]] + \cdots \\ &\quad + \frac{1}{n!}[\cdots[A, [A, B]]] + \cdots. \end{aligned} \tag{54}$$

In order to extend the concept of estimation algebra based on robust DMZ equation, we define some notations

$$\begin{aligned} \tilde{D}_i &:= \frac{\partial}{\partial z_i} - \tilde{f}_i, \quad \tilde{D} = (\tilde{D}_1, \ldots, \tilde{D}_n)^T \\ \tilde{\omega}_{ij} &:= \frac{\partial \tilde{f}_j}{\partial z_i} - \frac{\partial \tilde{f}_i}{\partial z_j} \\ \tilde{\eta}(t, z) &:= \sum_{i=1}^n \left(\tilde{f}_i^2 + \frac{\partial \tilde{f}_i}{\partial z_i}\right) + h^T(t, Fz)S^{-1}h(t, Fz) \\ &\quad - tr\left(\frac{dF^{-1}}{dt}F\right) \\ \tilde{L}_0 &:= \frac{1}{2}\left(\sum_{i=1}^n \tilde{D}_i^2 - \tilde{\eta}\right), \tilde{L}_i = h_i, 1 \le i \le m. \end{aligned} \tag{55}$$

By applying above notations, the robust DMZ equation $v(t, z)$ of system (51) is as follows:

$$\frac{\partial v}{\partial t} = \tilde{L}_0 v + \nabla_z K^T(t, Fz)\tilde{D}v(t, z)$$

$$+\frac{1}{2}\nabla_z K^T(t, Fz)\nabla_z K(t, Fz)v(t, z)$$
$$-\frac{\partial h^T(t, Fz)S^{-1}}{\partial t}y_t v(t, z). \tag{56}$$

REMARK 4 Equation (56) is important in extending the estimation algebra method to time-varying systems. We use the techniques in [21] for (56).

In what follows, we shall need the following assumption.

Next, the estimation algebra in time-varying filtering systems will be introduced and the structure will be analyzed.

DEFINITION 4.1 The estimation algebra $\tilde{E}$ of a filtering system (52) is defined to be the Lie algebra generated by $\{\tilde{L}_0, \tilde{h}_1, \ldots, \tilde{h}_m\}$, where $\tilde{h}_i$ is the ith row of $\tilde{h}(t, z) := h(t, F(t)z) = H(t)F(t)z$.

REMARK 5 In the traditional time-invariant finite dimensional filter, the coefficients of the basis in the estimation algebra are constant. But in the time-varying system, the basis is time-varying and the coefficient of the basis is a function only dependent on time which makes the corresponding estimation algebra be a $C^\infty([0, \infty))$- module.

LEMMA 4.2 Under Assumptions 4.1, 4.2, and 4.3, the estimation algebra $\tilde{E}$ of filtering system (52) and (53) is a subalgebra of $(2n+2)$ dimensional Lie algebra with a basis $\{\tilde{L}_0, \tilde{D}_1, \ldots, \tilde{D}_n, z_1, \ldots, z_n, 1\}$.

The proof is given in the appendix.

In the following, we shall try to extend the Baker-Campbell-Hausdorff type relations appears in [21].

LEMMA 4.3 If ζ is a C^∞ function in z and t, then for all $s \geq 0$, the following Baker-Campbell-Hausdorff type relations hold.

1) For $1 \leq j \leq n$

$$e^{s\tilde{D}_j}\tilde{L}_0\zeta = \left(\tilde{L}_0 + s\sum_{i=1}^{n}\tilde{\omega}_{i,j}\tilde{D}_i - \frac{s}{2}\frac{\partial\tilde{\eta}}{\partial z_j} + s^2 c_j\right) \times e^{s\tilde{D}_j}\zeta \tag{57}$$

where the $c_j = \frac{1}{2}[[\tilde{L}_0, \tilde{D}_j], \tilde{D}_j]$.

2) For $1 \leq j \leq n$

$$e^{sz_j}\tilde{L}_0\zeta = \left(\tilde{L}_0 - s\tilde{D}_j + \frac{s^2}{2}\right)e^{sz_j}\zeta. \tag{58}$$

3) Assume that the $E_1 = \tilde{D}_1, \ldots, E_n = \tilde{D}_n, E_{n+1} = z_1, \ldots, E_{2n} = z_n$. Then

$$e^{sE_i}E_j\zeta = (E_j + s\gamma_{i,j})e^{sE_i}\zeta \tag{59}$$

where the $\gamma_{i,j} = 0$ if $i, j > n$, $\gamma_{i,j} = 1$ if $i - j = n$ or $j - i = n$ and $\gamma_{i,j} = \tilde{\omega}_{j,i}$ if $i, j \leq n$.

The proof is given in the appendix.

In the following, we will use the structure of estimation algebra in Lemma 4.2 to derive a finite-dimensional filter by the Wei-Norman method.

THEOREM 4.1 Under Assumptions 4.1, 4.2, and 4.3, let $\tilde{E}$ be an FDEA of filtering system (52) with $\tilde{\eta}(t, x) = \sum_{i,j=1}^{n}\tilde{\eta}_{i,j}(t)z_i z_j + \sum_{i=1}^{n}\tilde{\eta}_i(t)z_i + \eta_0(t)$ where $\tilde{\eta}_{i,j}(t), \tilde{\eta}_i(t)$ and $\tilde{\eta}_0(t)$ are functions of t. Then the robust Duncan-Morten-Zakai equation (56) has a solution for all $t \geq 0$ of the following form:

$$v(t, z) = e^{T(t)}e^{r_n(t)z_n}\cdots e^{r_1(t)z_1} \times e^{s_n(t)\tilde{D}_n}\cdots e^{s_1(t)\tilde{D}_1}e^{\int_0^t L_0(\tau)d\tau}\sigma_0 \tag{60}$$

where $T(t), r_1(t), \ldots, r_n(t), s_1(t), \ldots, s_n(t)$ satisfy the following ordinary differential equations:

$$\frac{d\vec{s}}{dt} - \Omega(t)\vec{s} - \vec{r} = \nabla_z K(t, F(t)z) \tag{61a}$$

$$\frac{d\vec{r}}{dt} - \left(\frac{1}{2}\nabla_z^2\eta + \frac{dF^{-1}(t)}{dt}F(t) + F^{-1}(t)L(t)F(t)\right)^T\vec{s} = -y_t^T\frac{dS^{-1}(t)H(t)}{dt}F(t) \tag{61b}$$

and

$$\begin{aligned}\frac{dT(t)}{dt} = &\left(\sum_{i=1}^{n}\frac{ds_i}{dt}r_i(t) - \sum_{i=1}^{n}\frac{d\tilde{l}_i}{dt}\right)\\ &- \sum_{i=1}^{n-1}\sum_{m=i+1}^{n}\tilde{L}_{i,m}(t)s_m(t))\\ &- \sum_{j=1}^{n}\frac{s_j(t)^2}{2}\left(\sum_{i=1}^{n}\tilde{\omega}_{i,j}^2 + \frac{1}{2}\frac{\partial^2\tilde{\eta}}{\partial z_j^2}\right)\bigg)\\ &+ \sum_{j=1}^{n-1}\sum_{k=j+1}^{n}s_j s_k\left(\sum_{i=1}^{n}\tilde{\omega}_{i,j}\tilde{\omega}_{i,k}\right) + \frac{1}{2}\frac{\partial^2\tilde{\eta}}{\partial z_j\partial z_k})\\ &- \sum_{j=1}^{n}s_j\sum_{i=1}^{n}\tilde{\omega}_{i,j}r_i(t) + \frac{1}{2}\sum_{i=1}^{n}\frac{\partial K(t, F(t)z)}{\partial z_i}\frac{\partial K(t, F(t)z)}{\partial z_i}\end{aligned} \tag{62}$$

where $\vec{s} = (s_1, s_2, \ldots, s_n)^T$, $\vec{r} = (r_1, r_2, \ldots, r_n)^T$ are $n \times 1$ vectors.

The proof is given in the appendix.

Hence, the robust DMZ equation can be solved explicitly when the estimation algebra is finite-dimensional. But in the real application, ODEs (61a), (61b), and (62) are still hard to be solved accurately. Therefore, we need to design an effective numerical algorithm for time-varying finite-dimensional filters.

C. Explicit Solution of Tangent Flow

Although the tangent flow presents a general filtering framework, solving it for general systems can be challenging. However, the time-varying Yau system is a nonlinear case that admits an explicit solution of the tangent flow. Therefore, a new optimal transport particle algorithm for the time-varying Yau system can be designed based on tangent flow.

To start, we need to calculate φ_0, which can be numerically obtained by solving (42). Then, we recall the definition of $\tilde{H}(t)$ from Assumption 4.3, which is given by $H(t)F(t)$. Previous work on FPF [50] proposed the constant gain approximation to apply to high-dimensional filtering systems. This approximation is based on the fact that $\mathbb{E}[\bar{\nabla}\varphi_0(t,z)|\mathcal{Y}_t] = \Sigma_t \tilde{H}^T(t)S^{-1}(t)$ where Σ_t is the conditional variance. The constant gain approximation is to approximate $\nabla\varphi_0(z)$ by $\mathbb{E}[\bar{\nabla}\varphi_0(t,z)|\mathcal{Y}_t]$. According to (42), we can get the following:

$$\begin{aligned}&\bar{\Delta}\varphi_0(t,z) + \nabla(\log p(t,z))^T\bar{\nabla}\varphi_0(t,z)\\&\quad = -(z-\mu(t))^T\tilde{H}^T(t)S^{-1}(t)\end{aligned} \tag{63}$$

where $\mu(t) = \int_{\mathbb{R}^n} zp(t,z)dz$. We multiply term $zp(t,x)$ on both sides of (63) and take integration in terms of z, which yields

$$\int p(t,z)\bar{\nabla}\phi_0(t,z)dz = \int z(z-\mu(t))^T\tilde{H}^T(t)S^{-1}p(t,z)dz \tag{64}$$

by using integration by parts. Equivalently, we get $\mathbb{E}[\bar{\nabla}\varphi_0(t,z)|\mathcal{Y}_t] = \int z(z-\mu(t))^T\tilde{H}^T(t)S^{-1}(t)p(t,z)dz$. Then, we have

$$\begin{aligned}\mathbb{E}[\bar{\nabla}\varphi_0(t,z)|\mathcal{Y}_t] &= \int z(z-\mu(t))^T\tilde{H}^TS^{-1}(t)p(t,z)dz\\&= (\mathbb{E}[zz^T|\mathcal{Y}_t] - \mu(t)\mu(t)^T)\tilde{H}^TS^{-1}(t)\\&= \Sigma_t\tilde{H}^TS^{-1}(t)\end{aligned} \tag{65}$$

where Σ_t denotes the conditional covariance matrix. More details about constant gain can refer to [50] and [55].

Then, we shall calculate φ_1. Since the density function must be positive, we can assume the solution is

$$p(t,z) := e^{-u(t,z)}. \tag{66}$$

And, we substitute the condition (66) into the tangent flow condition (43). The new drift condition equation will be

$$\begin{aligned}&\Delta\varphi_1(t,z) - \nabla u(t,z)^T\nabla\varphi_1(t,z)\\&\quad= \frac{1}{2}\left[\Delta u(t,z) - |\nabla u(t,z)|_2^2\right]\\&\qquad - \tilde{f}\cdot\nabla u(t,z) + \sum_{i=1}^n \frac{\partial\tilde{f}_i}{\partial z_i}\\&\qquad + \frac{1}{2}(z-\mu(t))^T\tilde{H}^T(t)S^{-1}(t)\\&\qquad \times\tilde{H}(t)(z-\mu(t)) - c(t).\end{aligned} \tag{67}$$

1) *Linear System and Its OTPF:* This subsection will focus on the simplest situation: a linear system. Specifically, we will set $\phi = 0$ in (52), resulting in $\tilde{f}(t,z) = \tilde{L}(t)z + \tilde{l}(t)$. In addition, we will consider the filtering system with a Gaussian initial distribution. The OTPF for linear filtering systems was first proposed in [35], and it has been extended to correlated noise cases in [38].

For such a system, there is a well-known filtering Kalman-Bucy filter, and the posterior distributions are all Gaussian. Their means and variance satisfy the following:

$$\begin{aligned}d\mu(t) &= \tilde{L}(t)\mu(t)dt + \tilde{l}(t)dt\\&\quad + \Sigma_t\tilde{H}^T(t)S^{-1}(t)(dy_t - \tilde{H}(t)\mu(t)dt)\end{aligned} \tag{68}$$

$$\frac{d\Sigma_t}{dt} = \tilde{L}(t)\Sigma_t + \Sigma_t\tilde{L}^T(t) + I_n - \Sigma_t\tilde{H}^T(t)S_t^{-1}\tilde{H}(t)\Sigma_t. \tag{69}$$

And, (63) and (67) will become

$$\begin{aligned}\bar{\nabla}\varphi_0(t,z) &= \Sigma_t\tilde{H}^T(t)S^{-1}(t)\\\nabla\varphi_1(t,z) &= \tilde{L}(t)z + \tilde{l}(t) + \frac{1}{2}\Sigma_t^{-1}(z-\mu(t))\\&\quad - \frac{1}{2}\Sigma_t\tilde{H}^T(t)S(t)^{-1}\tilde{H}(t)(z-\mu(t)) + \xi(t,z)\end{aligned} \tag{70}$$

where $\nabla u(t,z)^T\xi(t,z) = 0$. The detailed derivation of (70) is given in Appendix.

The SDE (72) was first proposed in [35]. They define $\xi(t,z) := \Omega(t)\Sigma_t^{-1}(z-\mu(t))$, where $\Omega(t)$ satisfies the following matrix equation:

$$\begin{aligned}&\Omega(t)\Sigma_t^{-1} + \Sigma_t^{-1}\Omega(t)\\&\quad = \tilde{L}^T(t) - L(t) + \frac{1}{2}\left(\Sigma_t\tilde{H}^T(t)S(t)^{-1}\tilde{H}(t)\right.\\&\qquad \left.- \tilde{H}^T(t)S(t)^{-1}\tilde{H}(t)\Sigma_t\right).\end{aligned} \tag{71}$$

By using the result (70), the tangent flow of the linear filtering system (51) is as follows:

$$\begin{aligned}dz_t &= \tilde{L}(t)z_tdt + \tilde{l}(t)dt + \frac{1}{2}\Sigma_t^{-1}(z_t-\mu(t))dt\\&\quad\times K_t\left(dy_t - \frac{\tilde{H}(t)\mu(t) + \tilde{H}(t)z_t}{2}dt\right)\\&\quad + \Omega(t)\Sigma_t^{-1}(z-\mu(t))dt.\end{aligned} \tag{72}$$

REMARK 6 In the linear Gaussian case, it has been shown that different skew-symmetric $\Omega(t)$ choices result in the same posterior distributions [36]. However, a unique choice is defined in (72) that makes the dynamics symmetric and optimal in the optimal transportation sense. This choice can be seen as a correction term. If we set $\Omega(t)$ to be zero, then the SDE (72) becomes the same dynamical flow as the well-known square-root ensemble Kalman filter [36]. It is worth noting that the linear system is the simplest system in the FDEA system.

2) *Solution of FDEA System:* To construct the tangent flow of the nonlinear FDEA system satisfying Assumptions 4.1, 4.2, and 4.3, we need to characterize the conditional posterior densities of the FDEA system. The following theorem is summarized by several studies [13], [46]:

THEOREM 4.2 Assume that Assumptions 4.1, 4.2, and 4.3 are satisfied. Then, we consider the corresponding normalizing system (51) and denote the $p(t,x)$ as the posterior density. If the $p(0,z)e^{-\phi(z)}$ is Gaussian distribution, then the posterior distribution $p(t,z)$ at any time t satisfies the following:

$$p(t,x) = d(t)e^{-\frac{1}{2}(z-a(t)))^T\Theta^{-1}(t)(z-a(t))+\phi(z)} \tag{73}$$

where $d(t)$ is some normalized factor.

The proof is given in the appendix. Here, we need to point out that $a(t)$ and $\Theta(t)$ may not be the mean and the covariance matrix of $p(t,z)$, respectively. By using Theorem 4.2, the solution of φ_1 can be rewritten as follows:

$$\nabla\varphi_1(t,z) = \tilde{f}(t,z) + \frac{1}{2}\Theta^{-1}(t)(z-a(t)) - \frac{1}{2}\nabla\phi(z) - \frac{1}{2}\Sigma_t\tilde{H}^T(t)S(t)^{-1}\tilde{H}(t)(z-\mu(t)) + \xi(t,z) \tag{74}$$

where we apply the constant gain approximation $\nabla\varphi_0(t,z) \approx \Sigma_t\tilde{H}^T S(t)^{-1}$. Motivated by the fact that two terms $\tilde{L}(t)z$ and $\frac{1}{2}\Sigma_t\tilde{H}^T(t)S(t)^{-1}(\tilde{H}(t)z - \tilde{H}(t)\mu(t))$ in (74) are not in gradient form, which is same in the linear filter case. So, we shall set $\xi(t,z)$ still as $\Omega(t)\Sigma^{-1}(t)(z-\mu)$, which is defined in (71). Then, the tangent flow of FDEA can be given as

$$\begin{aligned} dz_t &= \tilde{L}(t)z_t dt + \tilde{l}(t)dt + \frac{1}{2}\nabla\phi(z_t)dt \\ &\quad + \frac{1}{2}\Theta(t)^{-1}(z_t - a(t))dt \\ &\quad + K_t \circ \left(dy_t - \frac{(\tilde{H}z_t + \tilde{H}\mu(t))dt}{2}\right) \\ &\quad + \Omega(t)\Sigma_t^{-1}(z_t - \mu(t))dt \end{aligned} \tag{75}$$

where $K_t = \Sigma_t\tilde{H}^T(t)S^{-1}(t)$, and the $\Omega(t)$ satisfies (71). Next, we shall propose a method to numerically calculate the term $\Theta(t)$ and $a(t)$. According to Theorem 3.2, we have

$$a(t) = \frac{\int z e^{-\phi(z)}p(t,z)dz}{\int e^{-\phi(z)}p(t,z)dz} = \frac{\mathbb{E}[ze^{-\phi(z)}]}{\mathbb{E}[e^{-\phi(z)}]} \tag{76}$$

and similarly

$$\Theta(t) = \frac{\mathbb{E}[zz^T e^{-\phi(z)}]}{\mathbb{E}[e^{-\phi(z)}]} - \left(\frac{\mathbb{E}[ze^{-\phi(z)}]}{\mathbb{E}[e^{-\phi(z)}]}\right)\left(\frac{\mathbb{E}[ze^{-\phi(z)}]}{\mathbb{E}[e^{-\phi(z)}]}\right)^T. \tag{77}$$

V. NUMERICAL ALGORITHM

In this section, we will design a numerical algorithm based on Section IV. In the real application, we shall normalize the FDEA system as (51). Then, we can construct the tangent flow of this new system explicitly by using the constant gain approximation and (74), which is

$$\begin{aligned} dz_t^i &= \tilde{L}(t)z_t^i dt + \tilde{l}(t)dt + \frac{1}{2}\nabla\phi(z_t^i)dt \\ &\quad + \frac{1}{2}\left(\Theta^{(N)}(t)\right)^{-1}\left(z_t^i - a^{(N)}(t)\right)dt \\ &\quad + K_t^{(N)} \circ \left(dy_t - \frac{(\tilde{H}z_t^i + \tilde{H}\mu^{(N)}(t))dt}{2}\right) \\ &\quad + \Omega(t)\Sigma_t^{-1}(x - \mu(t))dt \end{aligned} \tag{78}$$

where $a^{(N)}(t) = \frac{\sum_{i=1}^N z_t^i e^{-\phi(z_t^i)}}{\sum_{i=1}^N e^{-\phi(z_t^i)}}$, $\Theta^{(N)}(t) = \frac{\sum_{i=1}^N z_t^i (z_t^i)^T e^{-\phi(z_t^i)}}{\sum_{i=1}^N e^{-\phi(z_t^i)}} - a^{(N)}(t)(a^{(N)}(t))^T$, $\mu^{(N)}(t) = \frac{1}{N}\sum_{i=1}^N z_t^i$, $\Sigma_t^{(N)} = \frac{1}{N-1}\sum_{i=1}^N (z_t^i - \mu^{(N)}(t))(z_t^i - \mu^{(N)}(t))^T$ and $K_t^{(N)} = \Sigma_t^{(N)}\tilde{H}^T(t)S^{-1}(t)$.

VI. SIMULATION

In this section, we will test the efficiency of the OT method [optimal transportation particle filter, (OTPF)] and show numerical results. We consider three numerical examples including the time-invariant scalar case, time-variant scalar case, and high dimensional vector case.

To measure the accuracy of the numerical algorithm, the definitions of the mean of mean norm error (MMNE) and the mean time (MT) is as follows:

$$MMNE := \frac{1}{20}\sum_{i=1}^{20}\left[\frac{1}{NS+1}\sum_{j=0}^{NS}\|\hat{x}_{t_j}^i - x_{t_j}^{true}\|_2^2\right] \tag{79}$$

$$MT := \frac{1}{20}\sum_{i=1}^{20} RT_i \tag{80}$$

where NS means the number of steps in the experiments and the RT_i means the total running time of the ith experiment, the $\hat{x}_{t_j}^i$ is the filtering result of the ith experiment at time t_j, and $x_{t_j}^{true}$ is the real state of the ith experiment at time t_j.

In this subsection, the following high-dimensional cases will be tested. In this subsection, we test the performance of the OTPF by considering high-dimensional cases, and the following filters are simulated for the comparison.

1) FPF: the constant gain FPF algorithm described in [50].
2) UKF: the unscent Kalman filter described in [15] with parameters $\alpha = 0.01$, $\beta = 1.8$.
3) EKF: the extended Kalman filter described in [14].
4) IEKF: the iterated extended Kalman filter described in [56].

EXAMPLE 1 (HIGH-DIMENSIONAL CASE) We choose the model

$$\begin{cases} dX_t = 0.8(1+0.2\cos(t))\cdot\begin{pmatrix} A_1^n & 0 & 0 \\ 0 & \cdots & 0 \\ 0 & 0 & A_k^n \end{pmatrix} X_t dt \\ \qquad + \nabla\phi(X_t)dt + 1.5 I_n dB_t \\ dY_t = H_n X_t dt + dW_t \end{cases} \tag{81}$$

where the nondiagonal positions is zero, $n = 2k$

$$A_i^n = \begin{pmatrix} 0 & 1-(i-1)\cdot\frac{2}{n} \\ (i-1)\cdot\frac{2}{n} - 1 & 0 \end{pmatrix} \tag{82}$$

and $\phi(x_t) = \sum_{i=1}^n \ln(e^{[x_t]_i} + e^{-[x_t]_i})$, $[\cdot]_i$ denotes i-components of the vector. $H_n \in \mathbb{R}^{n\times n}$ is defined as follows:

$$H_n = \left(H_{i,j}^{(n)}\right)_{1\le i,j\le n}, H_{i,j}^{(n)}$$

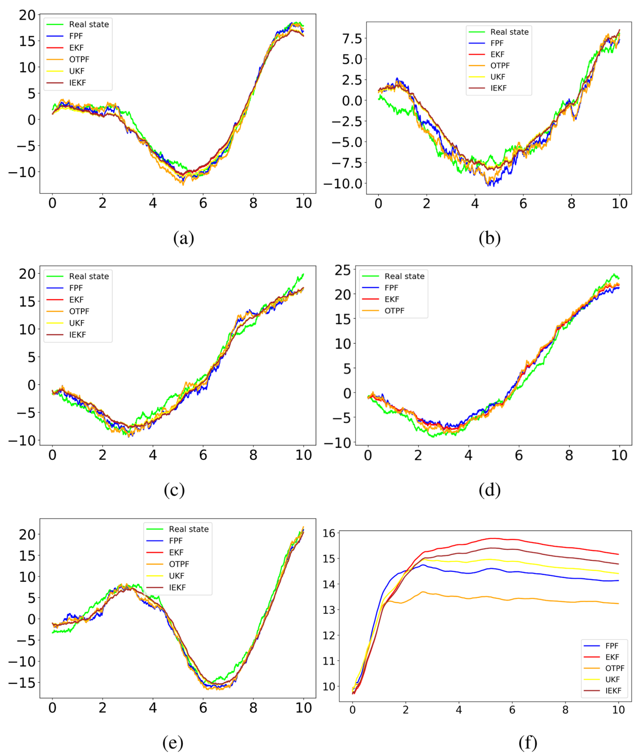

Fig. 1. 100 dimension case with 200 particles. (a) 1st dimension. (b) 25th dimension. (c) 50th dimension. (d) 75th dimension. (e) 100th dimension. (f) RMSE.

$$= \begin{cases} 1, \quad i = j \\ -0.1(i-j), \quad 0 < i - j \leq 5 \\ 0.003(i-j)^2 + 0.03, \; -5 \leq i - j < 0 \\ 0, \qquad \text{else} \end{cases}$$

and dB_t and dW_t are the standard n dimensional Brownian motions. The initial distribution σ_0 satisfies $\sigma_0 e^{-\phi(x_t)} \sim \mathcal{N}(\mu_n, 2I_n)$, where the first $[\frac{n}{2}]$ ($[\cdot]$ is the rounding function) components in μ_n are 1 and the others are -1.

The simulation results are discussed next.

1) *The figure of the filtering results:* The numerical results of the 100-D case with 200 particles are depicted in Fig. 1, where we chose to show the result of 1st dimension in (a), the result of 25th dimension in (b), the result of 50th dimension in (c), the result of 75th dimension in (d), the result of 100th dimension in (e), and the RMNE of time in (f). It is observed that OTPF provides a more accurate approximation than other algorithms.
2) *The comparison with different particle numbers:* In this simulation, we test the 10-D case system of (81) with different particle numbers $N \in \{10, 20, 40, 100\}$. The average results are shown in Table I. The OTPF with 20 particles performs better than EKF, UKF, and IEKF. The computation time of OTPF growth is much slower than FPF when the number of particles increases. It can be observed that the FPF with 100 particles performs similarly to the OTPF with 40 particles while the computational time of FPF is ten times that of OTPF. For comparison, the MMNE of the trivial filter obtained by simply inverting the observation matrix $\hat{X}_t = H_n^{-1}\frac{\Delta Y_t}{\Delta t}$ is 1107.4029 in this 10-D case.
3) *The comparison with different dimensions:* In this simulation, we test the performance of the proposed algorithm via the different dimensions of system (81), $n \in \{2, 4, 6, 8, 10, 20, 30, 40, 50, 100\}$ and the particle number are chosen as n and $2n$. The results are summarized in Tables II and III. The OTPF with $2n$ particles continues to exhibit the best performance via dimension n changes. The FPF with $2n$ particles has good performance but the computational time becomes unacceptable as the dimension increases (it takes over 90 s to simulate a 10 s system). We can conclude that OTPF is designed to deal with finite-dimensional Yau systems by utilizing particle evolution methodology. More importantly, OTPF is a real-time algorithm for the 100-D system.
4) *The dimensional comparisons where the dimension of the observation vector is half of the state vector*

TABLE I
Experiment Results 10-D Cases With Different Simulated Particle Number N

Method	EKF	UKF	IEKF	FPF	OTPF
MMNE	22.5179	19.4879	19.9417	29.3191	19.5838
MT	0.0932	0.5911	0.2441	1.1571	0.5708
N	–	–	–	10	10
MMNE	24.0187	19.2302	20.5642	20.7383	16.7212
MT	0.0905	0.5924	0.2417	2.006	0.6668
N	–	–	–	20	20
MMNE	23.1794	18.9965	20.1507	17.19127	15.7721
MT	0.0923	0.5921	0.2478	3.7742	0.8723
N	–	–	–	40	40
MMNE	21.67381	18.81763	19.74043	15.7026	15.3466
MT	0.0908	0.5858	0.2450	8.8179	1.4515
N	–	–	–	100	100

TABLE II
Dimensions in Simulation and Experiment Result

Dimension	2	4	6	8	10
EKF					
MMNE	3.8236	8.5848	12.5388	19.8047	21.8504
MT	0.08001	0.08204	0.08378	0.08476	0.08498
UKF					
MMNE	3.7512	6.1129	11.3806	17.4289	19.4156
MT	0.3097	0.3431	0.4372	0.54957	0.5766
IEKF					
MMNE	3.2714	6.9234	11.76432	17.9793	21.7878
MT	0.2159	0.2263	0.2321	0.23775	0.2466
FPF					
Particles	2	4	6	8	10
MMNE	9.9414	11.5608	18.4994	25.8121	27.6297
MT	0.3821	0.5667	0.7655	0.9773	1.1934
Particles	4	8	12	16	20
MMNE	5.4392	6.7979	9.5203	15.3946	19.6357
MT	0.5987	0.9192	1.2015	1.5735	1.9631
OTPF					
Particles	2	4	6	8	10
MMNE	5.46736	8.43777	11.4283	17.5371	20.9253
MT	0.4075	0.4378	0.5024	0.5636	0.6024
Particles	4	8	12	16	20
MMNE	3.0524	5.4055	8.5544	13.8154	16.0548
MT	0.44734	0.4942	0.5267	0.5707	0.7229

TABLE III
Dimensions in Simulation and Experiment Result

Dimension	20	30	40	50	100
EKF					
MMNE	43.8030	57.1328	88.0154	103.9819	225.8781
MT	0.11663	0.2363	0.3041	0.5641	0.7503
UKF					
MMNE	38.0983	53.1468	81.4468	90.6157	198.4468
MT	1.0423	1.8341	2.9509	5.7679	14.741923
IEKF					
MMNE	35.5404	53.7192	84.0154	95.9819	213.3688
MT	0.3037	0.4727	0.7488	1.2154	2.0660
FPF					
Particles	20	30	40	50	100
MMNE	53.1240	64.7332	105.9683	132.9215	242.4190
MT	2.1508	3.6975	6.7768	14.9229	45.7834
Particles	40	60	80	100	200
MMNE	40.9437	65.8355	78.0980	96.9388	180.7308
MT	4.00012	6.9910	13.6223	32.1665	95.9305
OTPF					
Particles	20	30	40	50	100
MMNE	34.1960	45.3911	65.8647	82.09158	161.7740
MT	0.8103	1.2493	2.0322	3.7389	7.2484
Particles	40	60	80	100	200
MMNE	31.0627	41.9929	64.4103	78.5034	157.0274
MT	1.02606	1.76695	2.90722	5.3000	9.7051

TABLE IV
Dimensions in Simulation and Experiment Result (With Half of Observation)

Dimension	2	4	6	8	10
EKF					
MMNE	6.2631	12.2631	20.725213	39.7239	61.8194
MT	0.0792	0.0916	0.1024	0.1045	0.1091
UKF					
MMNE	4.4237	9.7402	18.5519	23.2020	30.1169
MT	0.2675	0.3521	0.5099	0.5465	0.6111
IEKF					
MMNE	5.009	10.9393	19.8822	28.7317	33.2295
MT	0.2060	0.2178	0.2246	0.2306	0.2685
FPF					
Particles	2	4	6	8	10
MMNE	9.6125	14.5786	20.2332	24.5293	33.8893
MT	0.3215	0.4410	0.5969	0.9228	1.0998
Particles	4	8	12	16	20
MMNE	3.9765	9.2041	17.2041	22.8501	29.04011
MT	0.5663	1.017	1.329	1.5726	1.7932
OTPF					
Particles	2	4	6	8	10
MMNE	8.3041	12.7386	19.1126	23.0407	30.8748
MT	0.4001	0.4399	0.5019	0.5814	0.6363
Particles	4	8	12	16	20
MMNE	3.4376	6.0137	12.4365	18.6877	25.3408
MT	0.4347	0.4963	0.5528	0.6084	0.6965

dimension: In this simulation, we replaced the observation matrix with its first $m = n/2$ rows in Example 1. For such a new model, we test the performance of the algorithms via the different dimensions $n \in \{2, 4, 6, 8, 10\}$, and the particle number are chosen as n and $2n$. The numerical results are summarized in Table IV. By comparing these results with those presented in Tables II and III, we find that the EKF achieves relatively poor performance under these conditions. The OTPF, however, continues to display stable and effective performance.

5) *The comparison via different dimensions of the observation vector:* In this simulation, we replaced the observation matrix with its first m rows in the example 1, and the dimension of the state vector n is chosen as 10. For such a new model, we test the performance of the algorithms via the different dimensions $m \in \{1, 2, 3, 4, 5\}$, and the particle number is chosen as 20. The numerical results are summarized in Table V. As the dimension of observations changes, OTPF remains stable and maintains good performance in this numerical example.

TABLE V
Parameter Setting in Simulation 10-D Experiment Result With Different Dimensions of Observation (Dim Denotes the Dimension of the Observation Vector)

Method	EKF	UKF	IEKF	FPF	OTPF
MMNE	151.0089	58.2502	55.6451	59.4292	53.0294
MT	0.1043	0.5948	0.2269	1.3438	0.4673
Dim	1	1	1	1	1
MMNE	124.9765	51.6811	53.6734	51.8203	50.3095
MT	0.1077	0.6006	0.2319	1.4774	0.4992
Dim	2	2	2	2	2
MMNE	93.6929	43.8639	47.0998	40.8306	39.7969
MT	0.1076	0.6045	0.2495	1.5346	0.5433
Dim	3	3	3	3	3
MMNE	81.7734	37.2502	40.8886	34.4459	32.3435
MT	0.1087	0.6076	0.2535	1.6903	0.6138
Dim	4	4	4	4	4
MMNE	61.8194	30.1169	33.229	29.0401	25.3408
MT	0.1091	0.6111	0.2685	1.7932	0.6965
Dim	5	5	5	5	5

TABLE VI
Parameter Setting in Simulation 10-D Experiment Result With Different Time Step With 20 s

Method	EKF	UKF	IEKF	FPF	OTPF
MMNE	22.4212	20.4693	21.0772	20.4292	16.7926
MT	0.2395	1.1736	0.4845	3.9035	1.3609
dt	0.01	0.01	0.01	0.01	0.01
MMNE	24.0571	17.1975	17.3989	19.7797	16.8932
MT	0.1159	0.5816	0.2319	1.954	0.6702
dt	0.02	0.02	0.02	0.02	0.02
MMNE	21.7734	15.4423	14.8886	16.87457	14.8628
MT	0.0590	0.2910	0.1235	0.99034	0.3133
dt	0.05	0.05	0.05	0.05	0.05

6) *The comparison with different time-steps:* In this simulation, we test the performance of the proposed OTPF with 20 particles via the different time steps dt which are $0.01,\ 0.02,\ 0.05$, fixed the dimension as 10, and the total time as 20 seconds. The results are shown in Table VI. The OTPF still has the lowest MMNE in all situations. The IEKF becomes more efficient as the time step becomes larger. The OTPF is a stable algorithm for the FDEA system.

VII. CONCLUSION

We have successfully solved a class of time-varying filtering systems by using the Lie algebra method. The previous estimation algebra theory is only restricted to time-invariant filtering systems. In this work, we first extend the related estimation algebra theory to time-varying Yau filtering systems.

We extended a unified framework for FPF based on optimal transportation. Furthermore, we find the relationship between the operators of the Lie algebra method and tangent flow derived by optimal transport. Through the proposed framework, it is shown that the FPF is a particle filter updated by optimal transportation.

APPENDIX

A. PROOFS IN SECTION II

THEOREM 1.1 (REMARK 2.31 IN [34]) If $\alpha = \mathcal{N}(\mu_\alpha, \Sigma_\alpha)$ and $\beta = \mathcal{N}(\mu_\beta, \Sigma_\beta)$ are two Gaussians in $\mathbb{R}^n$ and Σ_α, Σ_β are positive-definite matrix, then one can show that the following map:

$$\mathcal{T} : x \mapsto \mu_\beta + V(x - \mu_\alpha) \tag{83}$$

is the optimal transportation with the cost function

$$c(\alpha, \beta) := \|\mu_\alpha - \mu_\beta\|_2^2 + \|\Sigma_\alpha^{\frac{1}{2}} - \Sigma_\beta^{\frac{1}{2}}\|_F^2 \tag{84}$$

where

$$V = \Sigma_\alpha^{-\frac{1}{2}}(\Sigma_\alpha^{\frac{1}{2}}\Sigma_\beta\Sigma_\alpha^{-\frac{1}{2}})^{-\frac{1}{2}}\Sigma_\alpha^{-\frac{1}{2}}. \tag{85}$$

□

B. PROOFS IN SECTION III

First, we introduce a vital lemma used in derivations of tangent flows.

LEMMA 1.1 (SPECIAL VERSION OF JACOBI'S FORMULA [57]) Let B be a $\mathbb{R}^{n\times n}$ matrix, then

$$\lim_{\epsilon\to 0} \frac{\det(I_n + \epsilon B) - \det(I_n)}{\epsilon} = Tr(B).$$

Next, we will give detailed proof in Section III.

PROOF OF THEOREM 3.1 For any arbitrary SPDE (29), there is an associate dynamical flow (35) after given the history path ω_t for dI_t where the (37) is equal to (36). Here, we can take out the coefficients of dI_t and dt at the left and right ends of the equation, and they must be equal, respectively. Finally, we can assume the SPDE is exactly Kushner (2) to complete the proof. □

PROOF OF LEMMA 3.1 We choose a test function $a(x) \in C^\infty$, then

$$\int_{\mathbb{R}^n} a(x)p(x)dx = \mathbf{E}[a(x_t)]. \tag{86}$$

Using the Ito lemma, we have that

$$\begin{aligned} da(x_t) &= \nabla a(x_t) \cdot dx_t \\ &= \nabla a(x_t)\{\mathcal{U}(t, x_t)dt + \mathcal{K}(t, x_t) \circ dI_t\}. \end{aligned} \tag{87}$$

Furthermore, (87) is equal to

$$\int_{\mathbb{R}^n} a(x)p(t,x)dx = \mathbf{E}\left[\int_0^t \nabla a(x_t)\cdot\mathcal{U}(t,x_t)dt\right] + \mathbf{E}\left[\int_0^t \nabla a(x_t)\cdot\mathcal{K}(t,x_t)\circ dI_t\right]. \tag{88}$$

The forward equation follows using integration by parts. □

PROOF OF THEOREM 3.2 The proof directly comes from the fact that the Fokker-Planck equation of (44) should be the same as (2). □

C. PROOFS IN SECTION IV

Next, we will give detailed proofs of Section IV.

PROOF OF LEMMA 4.2 First, we can consider the bigger basis which is $\{\tilde{L}_0, z_1, \cdots z_n\}$. And it is easy to find that $h_i \in \text{span}\langle z_1, \ldots, z_n\rangle$ with $1 \le i \le m$. We denote $\tilde{E}_1$ as the estimation algebra generated by $\{\tilde{L}_0, z_1, \cdots z_n\}$. So, we have $\tilde{E} \subset \tilde{E}_1$.

It can be shown that $z_i \in \tilde{E}_1$, $1 \le i \le n$. Then we calculate the following Lie bracket:

$$\begin{aligned} \left[\tilde{L}_0, z_i\right] &= \tilde{D}_i \in \tilde{E}_1, 1 \le i \le m \\ [\tilde{D}_i, \tilde{D}_j] &= \tilde{\omega}_{ji} \in \tilde{E}_1 \\ [\tilde{D}_i, z_j] &= \begin{cases} 1 & i = j \\ 0 & i \ne j \end{cases} \end{aligned} \tag{89}$$

and

$$\begin{aligned} \left[\tilde{L}_0, \tilde{D}_j\right] &= \frac{1}{2}\left[\sum_{i=1}^n \tilde{D}_i^2 - \tilde{\eta}, \tilde{D}_j\right] \\ &= \sum_{i=1}^n \tilde{\omega}_{j,i}\tilde{D}_i + \frac{1}{2}\frac{\partial\tilde{\eta}}{\partial z_j} \in \tilde{E}. \end{aligned} \tag{90}$$

Equation (55) yields $\tilde{\omega}_{j,i}$, $1 \le i, j \le n$ only depends on t. Notice that Assumption 4.3 yields that $\frac{\partial\tilde{\eta}}{\partial z_j}$, $1 \le j \le n$ is a degree 1 polynomial in z_i, $1 \le i \le n$. Therefore, estimation algebra $\tilde{E}_1$ is a $2n+2$ dimensional Lie algebra with basis given by $\{\tilde{L}_0, \tilde{D}_1, \ldots, \tilde{D}_n, z_1, \ldots, z_n, 1\}$. □

PROOF OF LEMMA 4.3: By Lemma 4.1 and direct calculation, the following results hold.

(1)

$$\begin{aligned} &e^{s\tilde{D}_j}\tilde{L}_0\zeta \\ &= e^{s\tilde{D}_j}\tilde{L}_0 e^{-s\tilde{D}_j}e^{s\tilde{D}_j}\zeta \\ &= \left(\tilde{L}_0 + s[\tilde{D}_j, \tilde{L}_0] + \frac{1}{2}s^2[\tilde{D}_j, [\tilde{D}_j, \tilde{L}_0]] + \cdots\right)e^{s\tilde{D}_j}\zeta \\ &= \left(\tilde{L}_0 + s\sum_{i=1}^n \tilde{\omega}_{i,j}\tilde{D}_i - \frac{s}{2}\frac{\partial\tilde{\eta}}{\partial z_j} + \frac{1}{2}s^2[\tilde{D}_j, [\tilde{D}_j, \tilde{L}_0]]\right)e^{s\tilde{D}_j}\zeta. \end{aligned} \tag{91}$$

(2)

$$\begin{aligned} &e^{sz_j}\tilde{L}_0\zeta \\ &= e^{sz_j}\tilde{L}_0 e^{-sz_j}e^{sz_j}\zeta \\ &= \left(\tilde{L}_0 + s[z_j, \tilde{L}_0] + \frac{1}{2}s^2[z_j, [z_j, \tilde{L}_0]] + \cdots\right)e^{sz_j}\zeta \\ &= \left(\tilde{L}_0 - s\tilde{D}_j + \frac{s^2}{2}\right)e^{sz_j}\zeta. \end{aligned} \tag{92}$$

The proof of (3) is similar. □

PROOF OF THEOREM 4.1 First, we consider the following derivative of operators. From the semigroup theory of partial differential equations, for any smooth function ζ, we have

$$\frac{\partial}{\partial t}\left(e^{s_i(t)\tilde{D}_i}\zeta(x)\right)=\left(\frac{ds_i(t)}{dt}\tilde{D}_i+s_i(t)\frac{d\tilde{D}_i}{dt}\right)e^{s_i(t)\tilde{D}_i}\zeta(x) \tag{93}$$

and

$$\frac{\partial}{\partial t}\left(e^{\int_0^t L_0(\tau)d\tau}\zeta(x)\right)=\tilde{L}_0(t)e^{\int_0^t \tilde{L}_0(\tau)d\tau}\zeta(x). \tag{94}$$

By differentiating (60), we have

$$\begin{aligned}
&\frac{\partial v}{\partial t}(t,x)\\
&=\frac{dT}{dt}e^{T(t)}e^{r_n(t)z_n}\cdots e^{r_1(t)z_1}\\
&\quad\times e^{s_n(t)\tilde{D}_n(t)}\cdots e^{s_1(t)\tilde{D}_1(t)}e^{\int_0^t L_0(\tau)d\tau}\sigma_0(x)\\
&\quad+e^{T(t)}\left(\frac{dr_n}{dt}z_n\right)e^{r_n(t)z_n}\cdots e^{r_1(t)z_1}\\
&\quad\times e^{s_n(t)\tilde{D}_n(t)}\cdots e^{s_1(t)\tilde{D}_1(t)}e^{\int_0^t L_0(\tau)d\tau}\sigma_0(x)\\
&\quad+\cdots\\
&\quad+e^{T(t)}e^{r_n(t)z_n}\cdots\left(\frac{dr_1}{dt}z_1\right)e^{r_1(t)z_1}\\
&\quad\times e^{s_n(t)\tilde{D}_n(t)}\cdots e^{s_1(t)\tilde{D}_1(t)}e^{\int_0^t L_0(\tau)d\tau}\sigma_0(x)\\
&\quad+e^{T(t)}e^{r_n(t)z_n}\cdots e^{r_1(t)z_1}\\
&\quad\times\left(\frac{ds_n(t)}{dt}\tilde{D}_n-s_n(t)\frac{\partial\tilde{f}_n}{\partial t}\right)e^{s_n(t)\tilde{D}_n(t)}\cdots e^{s_1(t)\tilde{D}_1(t)}\\
&\quad\times e^{\int_0^t L_0(\tau)d\tau}\sigma_0(x)\\
&\quad+\cdots\\
&\quad+e^{T(t)}e^{r_n(t)z_n}\cdots\times e^{r_1(t)z_1}\\
&\quad\times e^{s_n(t)\tilde{D}_n(t)}\cdots\left(\frac{ds_1(t)}{dt}\tilde{D}_1-s_1(t)\frac{\partial\tilde{f}_1}{\partial t}\right)e^{s_1(t)\tilde{D}_n(t)}\\
&\quad\times e^{\int_0^t L_0(\tau)d\tau}\sigma_0(x)\\
&\quad+e^{T(t)}e^{r_n(t)z_n}\cdots e^{r_1(t)z_1}\\
&e^{s_n(t)\tilde{D}_n(t)}\cdots e^{s_1(t)\tilde{D}_1(t)}L_0(t)e^{\int_0^t L_0(\tau)d\tau}\sigma_0(x).
\end{aligned} \tag{95}$$

Since $s_i(t), r_i(t), T(t)$ are all smooth functions of t, they can be exchanged with each other. In the following demonstration, we will exchange the operators $\frac{dr_i}{dt}z_i$, $\frac{ds_i(t)}{dt}\tilde{D}_i+s_i(t)\frac{\tilde{f}_i(t)}{dt}$, $1\le i\le n$ and $L_0(t)$ to the first term of the corresponding function, so that we can transform (60) to the equation in terms of $v(t,z)$.

First, we focus on the $\frac{dr_i}{dt}z_i$. They are commutative with any other function multiplication operator. So we have

$$\begin{aligned}
&e^{T(t)}\frac{dr_n}{dt}z_n e^{r_n(t)z_n}\cdots e^{r_1(t)z_1}\\
&\quad\times e^{s_n(t)\tilde{D}_n(t)}\cdots e^{s_1(t)\tilde{D}_1(t)}e^{\int_0^t L_0(\tau)d\tau}\sigma_0(x)\\
&\quad+\cdots\\
&\quad+e^{T(t)}e^{r_n(t)z_n}\cdots\left(\frac{dr_1}{dt}z_1\right)e^{r_1(t)z_1}\\
&\quad\times e^{s_n(t)\tilde{D}_n(t)}\cdots e^{s_1(t)\tilde{D}_1(t)}e^{\int_0^t L_0(\tau)d\tau}\sigma_0(x)\\
&=\left(\sum_{l=1}^n\frac{dr_l}{dt}z_l\right)e^{T(t)}e^{r_n(t)z_n}\cdots e^{r_1(t)z_1}\\
&\quad\times e^{s_n(t)\tilde{D}_n(t)}\cdots e^{s_1(t)\tilde{D}_1(t)}e^{\int_0^t L_0(\tau)d\tau}\sigma_0(x).
\end{aligned} \tag{96}$$

Second, we exchange $\frac{ds_i(t)}{dt}\tilde{D}_i-s_i(t)\frac{\partial\tilde{f}_i}{\partial t}$. By using Lemma 4.3, we have

$$\begin{aligned}
&e^{s_n(t)\tilde{D}_n(t)}\cdots\left(\frac{ds_i(t)}{dt}\tilde{D}_i-s_i(t)\frac{\tilde{f}_i(t)}{dt}\right)e^{s_i(t)\tilde{D}_i(t)}\cdots e^{s_1(t)\tilde{D}_1(t)}\\
&\quad\times e^{\int_0^t L_0(\tau)d\tau}\sigma_0(x)\\
&=\left(\frac{ds_i(t)}{dt}\tilde{D}_i-s_i(t)\frac{\partial\tilde{f}_i}{\partial t}+\sum_{j=i+1}^n\frac{ds_i(t)}{dt}s_j(t)\tilde{\omega}_{j,i}\right)\\
&\quad\times e^{s_n(t)\tilde{D}_n(t)}\cdots e^{s_1(t)\tilde{D}_1(t)}e^{\int_0^t L_0(\tau)d\tau}\sigma_0(x).
\end{aligned} \tag{97}$$

So we will have

$$\begin{aligned}
&e^{T(t)}e^{r_n(t)z_n}\cdots e^{r_1(t)z_1}\\
&\quad\times\left(\frac{ds_n(t)}{dt}\tilde{D}_n-s_n(t)\frac{\partial\tilde{f}_n}{\partial t}\right)e^{s_n(t)\tilde{D}_n(t)}\cdots e^{s_1(t)\tilde{D}_1(t)}\\
&\quad\times e^{\int_0^t L_0(\tau)d\tau}\sigma_0(x)\\
&\quad+\cdots\\
&\quad+e^{T(t)}e^{r_n(t)z_n}\cdots e^{r_1(t)z_1}\\
&\quad\times e^{s_n(t)\tilde{D}_n(t)}\cdots\left(\frac{ds_1(t)}{dt}\tilde{D}_1-s_1(t)\frac{\partial\tilde{f}_1}{\partial t}\right)e^{s_1(t)\tilde{D}_n(t)}\\
&\quad\times e^{\int_0^t L_0(\tau)d\tau}\sigma_0(x)\\
&=e^{T(t)}e^{r_n(t)z_n}\cdots e^{r_1(t)z_1}\\
&\left[\sum_{i=1}^n\left(\frac{ds_i(t)}{dt}\tilde{D}_i-s_i(t)\frac{\partial\tilde{f}_i}{\partial t}\right)+\sum_{k=1}^{n-1}\sum_{j=k+1}^n\frac{ds_k(t)}{dt}s_j(t)\tilde{\omega}_{j,k}\right]\\
&\quad\times e^{s_n(t)\tilde{D}_n(t)}\cdots e^{s_1(t)\tilde{D}_1(t)}e^{\int_0^t L_0(\tau)d\tau}\sigma_0(x).
\end{aligned} \tag{98}$$

In the following, we denote:

$$\begin{aligned}
\mathcal{M}:=&\sum_{i=1}^n\left(\frac{ds_i(t)}{dt}\tilde{D}_i-s_i(t)\frac{\partial\tilde{f}_i}{\partial t}\right)\\
&+\sum_{k=1}^{n-1}\sum_{j=k+1}^n\frac{ds_k(t)}{dt}s_j(t)\tilde{\omega}_{j,k}.
\end{aligned} \tag{99}$$

Next, we only need to exchange $\mathcal{M}$ and the function multiplication operator $e^{T(t)}e^{r_n(t)z_n}\cdots e^{r_1(t)z_1}$. Again using Lemma 4.3 (3), we can transform (98) to

$$\begin{aligned}
&\left[\sum_{i=1}^n\left(\frac{ds_i(t)}{dt}\tilde{D}_i-s_i(t)\frac{\tilde{f}_i(t)}{dt}\right)+\frac{ds_i(t)}{dt}r_i(t)\right.\\
&\left.+\sum_{k=1}^{n-1}\sum_{j=k+1}^n\frac{ds_k(t)}{dt}s_j(t)\tilde{\omega}_{j,k}\right]
\end{aligned}$$

$$\times e^{T(t)}e^{r_n(t)z_n}\cdots e^{r_1(t)z_1}$$
$$\times e^{s_n(t)\tilde{D}_n(t)}\cdots e^{s_1(t)\tilde{D}_1(t)}e^{\int_0^t L_0(\tau)d\tau}\sigma_0(x). \quad (100)$$

Finally, by a similar method, we can switch the operator $\tilde{L}_0(t)$ to the first term of the function by using Lemma 4.3

$$\begin{aligned} &e^{s_n(t)\tilde{D}_n(t)}\cdots e^{s_1(t)\tilde{D}_1(t)}\tilde{L}_0(t)e^{\int_0^t \tilde{L}_0(\tau)d\tau}\sigma_0(x) \\ &= \Bigg[\tilde{L}_0 + \sum_{l=1}^n\left(s_l(t)\sum_{i=1}^n\tilde{\omega}_{i,l}(t)\tilde{D}_i - \frac{s_l(t)}{2}\frac{\partial\tilde{\eta}}{\partial z_l} + s_l^2(t)c_l(t)\right) \\ &\quad + \sum_{j=1}^{n-1}\sum_{k=j+1}^n s_j(t)s_k(t)\left(\tilde{\omega}_{i,j}\tilde{\omega}_{i,k} - \frac{\partial^2\eta}{\partial z_j\partial z_k}\right)\Bigg] \\ &\quad\times e^{s_n(t)\tilde{D}_n(t)}\cdots e^{s_1(t)\tilde{D}_1(t)}e^{\int_0^t \tilde{L}_0(\tau)d\tau}\sigma_0(x) \end{aligned} \quad (101)$$

where we denote that

$$\begin{aligned} \mathcal{N} := &\tilde{L}_0 + \sum_{l=1}^n s_l(t)\sum_{i=1}^n\tilde{\omega}_{i,l}(t)\tilde{D}_i - \frac{s_l(t)}{2}\frac{\partial\tilde{\eta}}{\partial z_l} \\ &+ s_l^2(t)c_l(t)\sum_{j=1}^{n-1}\sum_{k=j+1}^n s_j(t)s_k(t)\left(\tilde{\omega}_{i,j}\tilde{\omega}_{i,k} - \frac{\partial^2\eta}{\partial z_j\partial z_k}\right). \end{aligned} \quad (102)$$

Next we only need to switch the $\mathcal{N}$ and the function multiplication operator $e^{T(t)}e^{r_n(t)z_n}\cdots e^{r_1(t)z_1}$. Using the fact that the function term in M is commutative with any other function multiplication operator and (101), we have the following:

$$\begin{aligned} &e^{T(t)}e^{r_n(t)z_n}\cdots e^{r_1(t)z_1} \\ &\quad\times e^{s_n(t)\tilde{D}_n(t)}\cdots e^{s_1(t)\tilde{D}_1(t)}\tilde{L}_0(t)e^{\int_0^t L_0(\tau)d\tau}\sigma_0(x) \\ &= e^{T(t)}e^{r_n(t)z_n}\cdots e^{r_1(t)z_1}e^{s_n(t)\tilde{D}_n(t)} \\ &\quad\times \mathcal{N}e^{s_n(t)\tilde{D}_n(t)}\cdots e^{s_1(t)\tilde{D}_1(t)}e^{\int_0^t L_0(\tau)d\tau}\sigma_0(x) \\ &= \left[\mathcal{N} - \sum_{l=1}^n s_l(t)\sum_{i=1}^n\tilde{\omega}_{i,l}(t)r_i(t) - \left(\sum_{i=1}^n r_i\tilde{D}_i - \frac{1}{2}r_j^2\right)\right] \\ &\quad\times e^{T(t)}e^{r_n(t)z_n}\cdots e^{r_1(t)z_1} \\ &\quad\times e^{s_n(t)\tilde{D}_n(t)}\cdots e^{s_1(t)\tilde{D}_1(t)}e^{\int_0^t L_0(\tau)d\tau}\sigma_0(x). \end{aligned} \quad (103)$$

Put (103), (100), and (96) to (95), we obtain the following:

$$\begin{aligned} \frac{\partial v}{\partial t} = &\Bigg[\mathcal{N} + \mathcal{M} + \sum_{l=1}^n\frac{dr_l}{dt}z_l \\ &- \sum_{l=1}^n s_l(t)\sum_{i=1}^n\tilde{\omega}_{i,l}(t)r_i(t) - \left(\sum_{i=1}^n r_i\tilde{D}_i - \frac{1}{2}r_i^2\right) + \frac{dT}{dt}\Bigg]v. \end{aligned} \quad (104)$$

By basic calculations, it can be obtained that in the right-hand side of (104) coefficient vector of $D_i v$ is $\frac{d\vec{s}}{dt} - \Omega(t)\vec{s} - \vec{r}$ and coefficient vector of $z_i v$ is $\frac{d\vec{r}}{dt} - (\frac{1}{2}\nabla_z^2\eta + \frac{dF^{-1}(t)}{dt}F(t) + F^{-1}(t)L(t)F(t))^T\vec{s}$.

By comparing (56) and (104), it is clear that $v(t,z)$ in (60) is a solution to (56) if (61a), (61b), and (62) are satisfied. □

Proof of Theorem 4.2 First, we shall consider the DMZ equation of the $p(t,z)$, and we consider the following density transformation $\tilde{p}(t,z)e^{\phi(z)} = p(t,z)$. So, we can have the following SPDE:

$$\begin{aligned} d(\tilde{p}(t,z)e^{\phi(z)}) = &\tilde{L}_0(\tilde{p}(t,z)e^{\phi(z)})dt \\ &+ z^T\tilde{H}(t)^T(\tilde{p}(t,z)e^{\phi(z)})\circ dy_t. \end{aligned} \quad (105)$$

By using Lemma 4.1, (105) can be rewritten as follows:

$$\begin{aligned} d(\tilde{p}(t,z)) &= e^{-\phi(z)}\tilde{L}_0e^{\phi(z)}(\tilde{p}(t,z))dt \\ &\quad + z^T\tilde{H}(t)^T(\tilde{p}(t,z)\circ dy_t \\ &= L_0(\tilde{p}(t,z))dt + [L_0,\phi(z)](\tilde{p}(t,z))dt \\ &\quad + \frac{1}{2}[[L_0,\phi(z)],\phi(z)]p(t,z))dt \\ &\quad + z^T\tilde{H}(t)^T(\tilde{p}(t,z)\circ dy_t. \end{aligned} \quad (106)$$

Then, we shall focus on the operator of the drift term in (106), which is as follows:

$$\begin{aligned} &\tilde{L}_0 + [\tilde{L}_0,\phi(z)] + \frac{1}{2}[[\tilde{L}_0,\phi(z)],\phi(z)] \\ &\quad= \tilde{L}_0 + \nabla\phi(z)^T\nabla - \tilde{f}(t,z)^T\nabla\phi(z) \\ &\quad\quad + \frac{1}{2}|\nabla\phi(z)|^2. \end{aligned} \quad (107)$$

It is easy to check that the second-order operator in (107) is $\frac{1}{2}\Delta$, the coefficient of the first-order operator is $(\nabla\phi - f)$ (it is the linear function of z), and the function term is $-\sum_{i=1}^n(\tilde{f}_i^2 + \frac{\partial\tilde{f}_i}{\partial z_i}) + \|(\nabla\phi - f)\|^2$, which is a quadratic function of z according to Assumption 4.3. Finally, we can finish the proof by using the sample fact that if the initial is Gaussian density, then the posteriors of (107) are all Gaussian. □

Derivation of (70): First, by using $\nabla u(t,z)^T = (z-\mu(t))\Sigma^{-1}(t)$, we can easily verify $\bar{\nabla}\varphi_0(t,z) = \Sigma_t\tilde{H}^T(t)S^{-1}(t)$.

We can submit the $\tilde{f}(t,z)$, $\frac{1}{2}\nabla u$, and $\frac{1}{2}\bar{\nabla}\varphi_0(Hz - H\mu(t))$ into $\nabla\cdot(*) - \nabla u(t,z)^T(*)$, which yield

$$\begin{aligned} &\nabla\cdot(\tilde{f}(t,z)) - \nabla u(t,z)^T(\tilde{f}(t,z)) \\ &\quad= \tilde{f}\cdot\nabla u(t,z) + \sum_{i=1}^n\frac{\partial\tilde{f}_i}{\partial z_i}, \end{aligned} \quad (108)$$

$$\begin{aligned} &\nabla\cdot\left(\frac{1}{2}\nabla u\right) - \nabla u(t,z)^T\left(\frac{1}{2}\nabla u\right) \\ &\quad= \frac{1}{2}\left[\Delta u(t,z) - |\nabla u(t,z)|_2^2\right] \end{aligned} \quad (109)$$

and

$$\begin{aligned} &\nabla\cdot\left(-\frac{1}{2}\bar{\nabla}\varphi_0H(t)(z-\mu(t))\right) \\ &\quad - \nabla u(t,z)^T\left(\frac{1}{2}\bar{\nabla}\varphi_0(t,z)H(t)(z-\mu(t))\right) \\ &\quad= \frac{1}{2}\bar{\Delta}\varphi_0(t,z)H(t)(z-\mu(t))) + \frac{1}{2}Tr(\varphi_0(t,z)H(t)) \\ &\quad\quad + \nabla u(t,z)^T\left(\frac{1}{2}\bar{\nabla}\varphi_0H(t)(z-\mu(t))\right) \\ &= \frac{1}{2}(z-\mu(t))^T\tilde{H}^T(t)S^{-1}(t)\tilde{H}(t)(z-\mu(t)) \end{aligned}$$

$$-\frac{1}{2}Tr(\varphi_0 H(t)) \tag{110}$$

where the last equality holds according to (63). Here, by using $\frac{1}{2}Tr(\varphi_0\tilde{H}(t)) = Tr(\Sigma_t\tilde{H}^T(t)S^{-1}(t)\tilde{H}(t))$, and

$$\begin{aligned} c(t) &= \int_{\mathbb{R}^n}\frac{1}{2}(z-\mu(t))^T\tilde{H}^T(t)S^{-1}(t) \\ &\quad \times \tilde{H}(t)(z-\mu(t))p(t,z)dz \\ &= \frac{1}{2}Tr\left(\int_{\mathbb{R}^n}\tilde{H}^T(t)S^{-1}(t)\tilde{H}(t)(z-\mu(t))\right. \\ &\quad \left.\times (z-\mu(t))^T p(t,z)dz\right. \\ &= Tr(\Sigma_t\tilde{H}^T(t)S^{-1}(t)\tilde{H}(t)). \end{aligned} \tag{111}$$

By using (111), (110) equals to $\frac{1}{2}(z-\mu(t))^T\tilde{H}^T(t)S^{-1}(t)\tilde{H}(t)(z-\mu(t)) - c(t)$. Finally, adding (108)–(110), then the right-hand side of the sums is the same as the right-hand side of (112). However, the $\tilde{f}(t,z)$ and $-\frac{1}{2}\bar{\nabla}\varphi_0 H(t)(z-\mu(t))$ are not in gradient form. So, we can add a divergence-free term $\xi(t,z)$ satisfied $\nabla\cdot(\xi(t,z)) - \nabla u(t,z)^T\xi(t,z) = 0$. In this case, the solution of φ_1 can be given as

$$\begin{aligned} \nabla\varphi_1(t,z) &= \tilde{f}(t,z) + \frac{1}{2}\nabla u(t,z) \\ &\quad - \frac{1}{2}\bar{\nabla}\varphi_0(t,z)(Hz - H\mu(t)) + \xi(t,z). \end{aligned} \tag{112}$$

The proof can be completed by using $\tilde{f}(t,z) := \tilde{L}(t)z + \tilde{l}(t)$. □

ACKNOWLEDGMENT

Prof. S. S.-T. Yau would like to thank the National Center for Theoretical Sciences (NCTS) for providing an excellent research environment while part of this research was done.

REFERENCES

[1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME–J. Basic Eng.*, vol. 82, no. Series D, pp. 35–45, 1960.

[2] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *ASME Trans., Part D, J. Basics Eng.*, vol. 83, pp. 95–103, 1961.

[3] T. E. Duncan, "Probability densities for diffusion processes with applications to nonlinear filtering system," Ph.D. dissertation, Dept. Elect. Eng., Stanford Univ., Stanford, CA, USA, 1967.

[4] R. E. Mortensen, "Optimal control of continuous time stochastic systems," Ph.D. dissertation, Dept. Elect. Eng., Univ. California, Berkeley, Berkeley, CA, USA, 1966.

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

[6] M. Davis, "On a multiplicative functional transformation arising in nonlinear filtering theory," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 54, no. 2, pp. 125–139, 1980.

[7] R. W. Brockett and J. M. C. Clark, "The geometry of the conditional density equation," in *Proc. Int. Conf. Anal. Optimisation Stochastic Syst.*, 1980, pp. 299–309.

[8] R. Brockett, "Stochastic Systems: The mathematics of filtering and identification and applications," *Nonlinear Systems and Nonlinear Estimation Theory*, M. Hazewinkel, J. C. Willems, Eds., pp. 441–477, 1981.

[9] S. K. Mitter, "On the analogy between mathematical problems of nonlinear filtering and quantum physics," *Ricerche Autom.*, vol. 30, no. 2, pp. 163–216, 1979.

[10] J. Wei and E. Norman, "On global representations of the solutions of linear differential equations as a product of exponentials," *Proc. Amer. Math. Soc.*, vol. 15, no. 2, pp. 327–327, 1964.

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

[12] J. Shi, Z. Yang, and S.-T. Yau, "Direct method for Yau filtering system with nonlinear observations," *Int. J. Control*, vol. 91, no. 3, pp. 678–687, 2017.

[13] X. Chen, J. Shi, and S. S.-T. Yau, "Real-time solution of time-varying Yau filtering problems via direct method and Gaussian approximation," *IEEE Trans. Autom. Control*, vol. 64, no. 4, pp. 1648–1654, Apr. 2019.

[14] L. Ljung, "Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems," *IEEE Trans. Autom. Control*, vol. 24, no. 1, pp. 36–50, Feb. 1979.

[15] E. A. Wan and R. Van Der Merwe, "The unscented Kalman filter for nonlinear estimation," in *Proc. Adaptive Syst. Signal Process., Commun., Control Symp.*, 2000, pp. 153–158.

[16] C. Fitzpatrick, "Observations of particle detachment during filter backwashing," *Water Sci. Technol.*, vol. 27, no. 10, pp. 213–221, 1993.

[17] S. S. T. Yau, "Finite-dimensional filters with nonlinear drift. I: A class of filters including both Kalman-Bucy and Benes filters," *J. Math. Syst., Estimation, Control*, vol. 4, pp. 181–203, 1994.

[18] J. Chen and S. S. T. Yau, "Finite-dimensional filters with nonlinear drift, VI: Linear structure of Ω," *Math. Control, Signals Syst.*, vol. 9, no. 4, pp. 370–385, 1996.

[19] W. Wong, X. Wu, and S. S. T. Yau, "Hessian matrix non-decomposition theorem," *Math. Res. Lett.*, vol. 6, no. 6, pp. 663–673, 1999.

[20] S. T. Yau and S.-T. Yau, "Finite-dimensional filters with nonlinear drift. III: Duncan-Mortensen-Zakai equation with arbitrary initial condition for the linear filtering system and the Benes filtering system," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, no. 4, pp. 1277–1294, Oct. 1997.

[21] S. S.-T. Yau, "Complete classification of finite-dimensional estimation algebras of maximal rank," *Int. J. Control*, vol. 76, no. 7, pp. 657–677, 2003.

[22] S. S.-T. Yau and G. Q. Hu, "Classification of finite-dimensional estimation algebras of maximal rank with arbitrary state–space dimension and Mitter conjecture," *Int. J. Control*, vol. 78, no. 10, pp. 689–705, 2005.

[23] X. Wu and S. S.-T. Yau, "Classification of estimation algebras with state dimension 2," *SIAM J. Control Optim.*, vol. 45, no. 3, pp. 1039–1073, 2006.

[24] W. Dong, X. Chen, and S. S. T. Yau, "The novel classes of finite dimensional filters with non-maximal rank estimation algebra on state dimension four and rank of one," *Int. J. Control*, vol. 94, no. 2, pp. 1–20, 2019.

[25] J. Shi, X. Chen, W. Dong, and S. S. T. Yau, "New classes of finite dimensional filters with non-maximal rank," *IEEE Contr. Syst. Lett.*, vol. 1, no. 2, pp. 233–237, Oct. 2017.

[26] J. Shi and S. S.-T. Yau, "Finite dimensional estimation algebras with state dimension 3 and rank 2, I: Linear structure of wong matrix," *SIAM J. Control Optim.*, vol. 55, no. 6, pp. 4227–4246, 2017.

[27] J. Shi and S. S.-T. Yau, "Finite dimensional estimation algebras with state dimension 3 and rank 2, Mitter conjecture," *Int. J. Control*, vol. 93, no. 9, pp. 2177–2186, 2020.

[28] X.-P. Jiao and S. S.-T. Yau, "New classes of finite dimensional filters with nonmaximal rank estimation algebra on state dimension n and linear rank n-2," *SIAM J. Control Optim.*, vol. 58, no. 6, pp. 3413–3427, 2020.

[29] S.-T. Yau and Y. T. Lai, "Explicit solution of DMZ equation in nonlinear filtering via solution of ODEs," *IEEE Trans. Autom. Control*, vol. 48, no. 3, pp. 505–508, Mar. 2003.

[30] J. L. Beck and S.-K. Au, "Bayesian updating of structural models and reliability using Markov chain Monte Carlo simulation," *J. Eng. Mechanics*, vol. 128, no. 4, pp. 380–391, 2002.
[31] T. Yang, P. G. Mehta, and S. P. Meyn, "Feedback particle filter," *IEEE Trans. Autom. Control*, vol. 58, no. 10, pp. 2465–2480, Oct. 2013.
[32] J. Steinbring and U. D. Hanebeck, "S 2 KF: The smart sampling Kalman filter," in *Proc. 16th Int. Conf. Inf. Fusion*, 2013, pp. 2089–2096.
[33] L. V. Kantorovich, "On the translocation of masses," *Dokl. Akad. Nauk. USSR (NS)*, vol. 37, pp. 199–201, 1942.
[34] G. Peyré et al., "Computational optimal transport: With applications to data science," *Found. Trends Mach. Learn.*, vol. 11, no. 5/6, pp. 355–607, 2019.
[35] A. Taghvaei and P. G. Mehta, "An optimal transport formulation of the linear feedback particle filter," in *Proc. Amer. Control Conf.*, 2016, pp. 3614–3619.
[36] A. Taghvaei and P. G. Mehta, "An optimal transport formulation of the ensemble Kalman filter," *IEEE Trans. Autom. Control*, vol. 66, no. 7, pp. 3052–3067, Jul. 2021.
[37] A. Taghvaei and P. G. Mehta, "Optimal transportation methods in nonlinear filtering," *IEEE Control Syst. Mag.*, vol. 41, no. 4, pp. 34–49, Aug. 2021.
[38] J. Y. Kang, X. Chen, Y. Tao, and S. T. S. Yau, "Optimal transportation particle filter for linear filtering systems with correlated noises," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 6, pp. 5190–5203, Dec. 2022.
[39] S. Reich, "A nonparametric ensemble transform method for Bayesian inference," *SIAM J. Sci. Comput.*, vol. 35, no. 4, pp. A2013–2024, 2013.
[40] T. A. E. Moselhy and Y. M. Marzouk., "Bayesian inference with optimal maps," *J. Comput. Phys.*, vol. 231, no. 23, pp. 7815–850, 2012.
[41] S. K. D. A. Mesa, J. M. Mendoza, and T. P. Coleman, "A distributed framework for the construction of transport maps," *Neural Comput.*, vol. 31, no. 4, pp. 613–52, 2019.
[42] J. Heng, "On the use of transport and optimal control methods for Monte Carlo simulation," Ph.D. dissertation, St Cross College, Univ. Oxford, Oxford, U.K., 2016.
[43] H. J. Kushner, "Dynamical equations for optimal nonlinear filtering," *J. Differ. Equ.*, vol. 3, no. 2, pp. 179–190, 1967.
[44] C. Villani, *Topics in Optimal Transportation*. Providence, RI, USA: American Mathematical Society, 2003, vol. 58.
[45] Y. Brenier, "Polar factorization and monotone rearrangement of vector-valued functions," *Commun. Pure Appl. Math.*, vol. 44, no. 4, pp. 375–417, 1991.
[46] X. Luo and S. S. -T. Yau, "Complete real time solution of the general nonlinear filtering problem without memory," *IEEE Trans. Autom. Control*, vol. 58, no. 10, pp. 2563–2578, Oct. 2013.
[47] F. Konecny, "On Wong-Zakai approximation of stochastic differential equations," *J. Multivariate Anal.*, vol. 13, no. 4, pp. 605–611, 1983.
[48] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *J. Appl. Statist.*, vol. 20, no. 5/6, pp. 25–62, 1993.
[49] N. Ding, Y. Fang, R. Babbush, C. Chen, R. Skeel, and H. Neven, "Bayesian sampling using stochastic gradient thermostats," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3203–3211. [Online]. Available: http://papers.nips.cc/paper/5592-bayesian-sampling-using-stochastic-gradient-thermostats
[50] P. G. M. T. Yang, R. S. Laugesen, and S. P. Meyn, "Multivariable feedback particle filter," *Automatica*, vol. 71, pp. 10–23, 2016.
[51] C. Liu, J. Zhuo, P. Cheng, R. Zhang, and J. Zhu, "Understanding and accelerating particle-based variational inference," in *Proc. 36th Int. Conf. Mach. Learn.*,, 2019, pp. 4082–4092.
[52] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization With PDE Constraints*. Berlin, Germany: Springer-Verlag, 2008.
[53] D. Crisan and J. Xiong, "Approximate Mckean–Vlasov representations for a class of SPDEs," *Stochastics*, vol. 82, no. 1, pp. 53–68, 2010.
[54] S. Reich, "A dynamical systems framework for intermittent data assimilation," *BIT Numer. Math.*, vol. 51, no. 1, pp. 235–249, 2011.
[55] A. Taghvaei, P. G. Mehta, and S. P. Meyn, "Diffusion map-based algorithm for gain function approximation in the feedback particle filter," *SIAM/ASA J. Uncertainty Quantification*, vol. 8, no. 3, pp. 1090–1117, 2020.
[56] R. P. Wishner, J. A. Tabaczynski, and M. Athans, "A comparison of three non-linear filters," *Automatica*, vol. 5, no. 4, pp. 487–496, 1969.
[57] R. Bellman, *Introduction to Matrix Analysis*. Philadelphia, PA, USA: SIAM, 1997.

Jiayi Kang received the B.S. degree in pure and applied mathematics from the College of Mathematics, Sichuan University, Chengdu, China, in 2019. He is currently working toward the Ph.D. degree in applied mathematics from the Department of Mathematical Sciences, Tsinghua University, Beijing, China.

His research interests include machine learning, nonlinear filtering, and bioinformatics.

Xiaopei Jiao received the B.S. degree in physics from Shanghai Jiao Tong University, Shanghai, China, in 2017 and the Ph.D. degree in applied mathematics from the Department of Mathematical Sciences, Tsinghua University, Beijing, China, in 2022.

He is currently a Postdoctoral Researcher with the Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, Beijing, China. His current research interests include nonlinear filtering, optimal control, and bioinformatics.

Stephen S.-T. Yau (Life Fellow, IEEE) received the Ph.D. degree in mathematics from the State University of New York at Stony Brook, Stony Brook, NY, USA, in 1976.

From 1976 to 1977 and from 1981 to 1982, he was a Member of the Institute of Advanced Study, Princeton, NJ, USA. From 1977 to 1980, he was a Benjamin Pierce Assistant Professor with Harvard University, Cambridge, MA, USA. 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 over 30 years. From 2005 to 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, CR geometry, and singularities theory.

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