shows great efficiency in numerical experiments, including both the scalar and high-dimensional cases.

Optimal Transportation Particle Filter for Linear Filtering Systems With Correlated Noises

JIAYI KANG
Tsinghua University, Beijing, China

XIUQIONG CHEN, Member, IEEE
Renmin University of China, Beijing, China

YANGTIANZE TAO
STEPHEN SHING-TOUNG YAU, Life Fellow, IEEE
Tsinghua University, Beijing, China

Dedicate to Professor Thomas Kailath on the occasion of his 87 Birthday.

In this article, we derive an optimal transportation particle filter for linear time-varying systems with correlated noises. This method can be regarded as the extension of the feedback particle filter with an optimal transportation structure. However, the particles in our method are evolved in a deterministic way, while we need to generate random particles in a feedback particle filter. Consequently, we only need a very few particles to obtain the satisfying results, and this property is especially significant for high-dimensional problems. The error analysis of our method and the feedback particle filter has been carried out when the system is time invariant. Compared with the feedback particle filter and the ensemble Kalman filter, our method

Manuscript received 18 August 2021; revised 9 December 2021 and 24 March 2022; released for publication 4 April 2022. Date of publication 12 April 2022; date of current version 6 December 2022.

DOI. No. 10.1109/TAES.2022.3166863

Refereeing of this contribution was handled by M. Efe.

This work was supported in part by the National Natural Science Foundation of China under Grant 11961141005, in part by the Tsinghua University start-up fund , and in part by the Tsinghua University Education Foundation fund under Grant 042202008. Professor Stephen Shing-Toung Yau is grateful to the National Center for Theoretical Sciences (NCTS) for providing an excellent research environment while part of this research was done.

Authors' addresses: Jiayi Kang, Yangtianze Tao, and Stephen Shing-Toung Yau are with the Department of Mathematical Sciences, Tsinghua University, Beijing 100084, China, E-mail: (kangjy19@mails.tsinghua.edu.cn; tytz19@mails.tsinghua.edu.cn; yau@uic.edu); Xiuqiong Chen is with the School of Mathematics, Renmin University of China, Beijing 100872, China, E-mail: (weizhicxq@163.com). *(Corresponding author: Stephen Shing-Toung Yau.)*

I. INTRODUCTION

Filtering has a long history, which can be traced back to the work of Wiener in 1949. In 1960, Kalman published a paper investigating a recursive solution to the discrete linear filtering problem [1], which is known as the Kalman filter (KF). A year later, Kalman and Bucy [2] proposed the continuous version of the KF, which has been widely used in various fields.

More generally, Crisan [3] and Davis [4] have studied a special kind of filtering systems called correlated systems, in which state noise and observation noise are correlated, i.e., $\sigma_W \neq 0$ in (2). The correlated system is more suitable to model the real-life problems, as explained in a substantial number of papers, such as [5]–[7].

Let $\mathcal{Y}_t$ be the σ-algebra generated by an observation process $\{y_s, 0 \leq s \leq t\}$, which is governed by some stochastic differential equations (SDEs) associated with x_t, and x_t is a hidden Markov process. Let p^* be the posterior density function, which is defined so that, for any measurable set $U \in \mathbb{R}^n$,

$$\int_U p^*(x,t)dx = P(x_t \in U|\mathcal{Y}_t). \tag{1}$$

In a word, the central problem of filtering is how to obtain the posterior density $p^*(x,t)$, which is governed by the so-called Kushner equation [8].

However, for the general filtering problems, it is not easy to obtain the explicit posterior density. In real applications, it seems that we often need to obtain some features of the posterior distributions instead of completely solving them. With the developments of Bayesian networks and Monte Carlo methods, the particle filter (PF) [9] was proposed to avoid solving Kushner equations and to simulate the distributions directly. However, in the traditional PF framework, the PF suffers from weight degeneracy, and the number of particles increases exponentially as the dimension increases. The weight degeneracy problem can only be alleviated by resampling steps under ten dimensions [8]. For higher dimensional problems, the PF will fail due to requiring too many particles.

An important breakthrough came from a feedback particle filter (FPF) [10], [11]. The FPF is based on a feedback control structure, and the number of particles can be reduced in this algorithm compared with the traditional PF [10]. In FPF's evolution, the weights of particles are not changed, and there is no importance sampling required as in the conventional PF. Therefore, a crucial distinction between the PF and the FPF is that there is no resampling of particles for the FPF. This property allows the FPF to be flexible with regard to implementation and does not suffer from particle degeneracy or sample impoverishment [10]. There are some works that give detailed comparisons between PFs and the FPF [12], [13]. The central problem of the FPF is how to numerically estimate the control function, which is governed by a family of partial differential equations (PDEs)

[10], [11]. However, there is no unified and simple numerical method for general high-dimensional PDEs at present. For more general nonlinear high-dimensional filtering problems, the FPF method is still facing the curse of the dimension problem. However, the stability and the convergence of the algorithm are still important problems. The error analysis of the FPF for a linear Gaussian system is given in [14]. Furthermore, the error analysis of the FPF for the general nonlinear system with continuous state and discrete observation is given in [15]. The FPF for the correlated scalar system was first proposed by Luo and Miao [16] in 2019.

It is unfortunate to see that the FPF does not perform well in high-dimensional numerical simulations, which can be checked through the experiments in Section V. In real applications, for example, the weather prediction problem [17], one needs to deal with the system of 10^{11} dimensions. The high dimension of the state provides a significant computational challenge even for the linear Gaussian cases.

Therefore, in this article, we mainly consider the following continuous linear filtering system:

$$\begin{cases} dx_t = A_t x_t dt + \sigma_W(t) dW_t + \sigma_B(t) dB_t \\ dy_t = H_t x_t dt + dW_t \end{cases} \tag{2}$$

where $x_t \in \mathbb{R}^n$ is the state and $y_t \in \mathbb{R}^m$ is the observation. The time t is in $[0, S]$, $S > 0$. $\{W_t\}$ and $\{B_t\}$ are independent Brownian motion processes with proper dimensions, and their covariance matrices are $E[dW_t^T dW_t] = Q(t)dt$ and $E[dB_t^T dB_t] = I_n dt$. Here, $\{W_t\}$, $\{B_t\}$, and the initial state x_0 are assumed to be independent of each other. A_t, H_t, $\sigma_W(t)$, and $\sigma_B(t)$ are assumed to be C^∞ functions of time, where $A_t, \sigma_B(t) \in \mathbb{R}^{n\times n}$. $H_t \in \mathbb{R}^{m\times n}$, and $\sigma_W(t) \in \mathbb{R}^{n\times m}$.

The ensemble Kalman filter (EnKF) [17]–[20] is a classical algorithm, which is popular in high-dimensional applications. The advantage of the EnKF is that it requires fewer computational resources than a KF that needs to store and propagate the error covariance matrix. The computational complexity and storage of the EnKF are $\mathcal{O}(n^2)$.

Sampling from the high-dimensional distribution is time consuming, and there are many works focusing on improving the sampling methods in the FPF and the EnKF [19]. Ignoring the difference in forms, many new sampling methods and inference methods are directly related to another important research field, optimal transportation (OT). Different sampling methods can be regarded as different dynamic evolution processes in the framework of OT. The map of transportation used in this article is motivated by the several methods that are used in uncertainty propagation and Bayesian inference [21]–[25]. The error analysis of the proposed algorithms is motivated by several methods appeared in [15] and [26]–[28].

The challenges in the FPF can be summarized as follows:

1) Sampling is time consuming in high-dimensional cases.
2) The number of particles is huge in high-dimensional cases.

The motivation of this article is twofold. First, with the vigorous developments of generative models in deep learning, a variety of sampling methods have been proposed due to the unified OT framework. The FPF provides a natural connection between sampling and filtering problems, which means that any new sampling method can correspond to a new EnKF. So far, there are already many different EnKFs [17]–[20], which are widely used in many real applications. However, a unified understanding of the EnKFs is still empty in generally correlated cases [28]. Therefore, it is natural to focus on the linear system in this article. Second, there are two major issues in the current FPF algorithms even in linear Gaussian cases. For example, the FPF in correlated noise is not efficient in high-dimensional cases, which can be seen from our experiments. Therefore, we want to construct a new FPF to overcome the slowness of high-dimensional sampling and combine it with a feedback control structure, which has stability with a small number of particles. Deterministic transfer in OT makes it possible. Besides, the equivalence of stochastic evolution and deterministic evolution gives a unified method to accelerate the FPF, which is only possible by setting up a unified framework in OT.

We make the following contributions in this article.

1) We propose a novel optimal transportation particle filter (OTPF) for linear filtering systems with correlated noises, and this new filter can work very efficiently for high-dimensional cases, which has been verified in several numerical experiments. Moreover, the so-called square-root EnKF [20] can be considered as the suboptimal filter of the OTPF, as discussed in Remark 2.
2) The OTPF can be considered as the extension of the FPF. Similarly, there is no importance sampling or resampling used in the OTPF, which allows the OTPF to not suffer from particle degeneracy or sample impoverishment.
3) We give a rigorous proof of the convergence of the proposed method when the system is time invariant. Furthermore, we show that the mean square error (MSE) is bounded by $\frac{Ce^{-\lambda t}}{\sqrt{N}}$, where the constant C has polynomial dependence on the dimension of state space, λ is governed by the system, and N is the number of particles.
4) We also give the error analysis of the FPF for linear time-invariant scalar systems with correlated noises.

In this article, we use $\|\cdot\|_2$ to represent the L_2 norm of the vectors or the matrices, $\|\cdot\|_F$ to represent the Frobenius norm (FN) of the matrices, and $\mathrm{Tr}(\cdot)$ to represent the trace of the matrix.

The rest of this article is organized as follows. In Section II, we recall the Kalman–Bucy filter (KBF) and the FPF for the linear time-invariant system with correlated noise. The motivation, derivation, and technique details of our OTPF are shown in Section III. The detailed understanding of the new algorithm is given at the end of Section III.

Furthermore, we prove the convergence of the OTPF and the FPF for linear systems with correlated noises in Section IV. In Section V, we compare our OTPF with the FPF, the EnKF, and the KBF in two numerical examples. Finally, Section VI concludes this article.

II. NOTATIONS AND PRELIMINARIES

In this section, we introduce some preliminary knowledge. In Section II-A, we listed the optimal filter, i.e., KBF. In Section II-B, we slightly generalized the FPF to the time-varying cases. Both the KBF and the FPF are used to compare with our new algorithms in Section V.

A. Kalman–Bucy Filter for (2)

It is well known that the optimal estimate of the state in (2) is given by the KBF. Let $\mu_t := E[x_t|\mathcal{Y}_t]$ and $P_t := E[(x_t - \mu_t)(x_t - \mu_t)^T|\mathcal{Y}_t]$. Then, the evolution equations of the conditional expectation μ_t and the conditional covariance P_t are given in the following lemma.

LEMMA 1 (SEE[29]) The KBF of the system (2) is as follows:

$$d\mu_t = A_t\mu_t dt + [P_tH_t^T + \sigma_W(t)Q_t)]Q_t^{-1}(dy_t - H_t\mu_t dt) \tag{3}$$

$$\frac{dP_t}{dt} = A_tP_t + P_tA_t^T + \sigma_B(t)\sigma_B^T(t) - [P_tH_t^T\sigma_W^T(t) + \sigma_W(t)H_tP_t + P_tH_t^TQ_t^{-1}H_tP_t]. \tag{4}$$

Although the optimal estimate μ_t can be obtained by solving the ordinary differential equations (ODEs) in Lemma 1, ODEs in Lemma 1 are difficult to be solved fast in high-dimensional cases, since the size of P_t is $n \times n$.

Therefore, it is vital to propose an effective and fast numerical PF algorithm for this filtering system.

B. Feedback Particle Filter for (2)

In this part, we shall introduce the FPF for (2), which was proposed in [16]. The FPF for nonlinear filtering systems with correlated noises can be found in [16]. Since the proposed FPF is for time-invariant cases, we need to slightly generalize it to the time-varying cases.

The evolution equation of the ith particle $\tilde{x}_t^i$ in the FPF is given by the following controlled system:

$$d\tilde{x}_t^i = A_t\tilde{x}_t^i dt + \sigma_B(t)dB_t^i + u(\tilde{x}_t^i, t)dt + \tilde{\mathcal{K}}(\tilde{x}_t^i, t)dy_t \tag{5}$$

where $(u, \tilde{\mathcal{K}})$ is the control input to be determined. The initial particles $\{\tilde{x}_0^i\}$ are drawn from the initial distribution $p^*(x, 0)$ of x_0. Let $p(x, t)$ be the conditional density function of $\tilde{x}_t^i$, i.e., for any measurable set $U \in \mathbb{R}^n$

$$\int_U p(x, t)dx = P(\tilde{x}_t^i \in U|\mathcal{Y}_t). \tag{6}$$

DEFINITION 1 (OPTIMAL) We call the control $(\tilde{\mathcal{K}}, u)$ is optimal if $p(x, t) = p^*(x, t)$, given the same initial condition $p(x, 0) = p^*(x, 0)$, where $p^*(x, t)$ defined in (1) is the posterior density function of x_t.

It is vital to know that the posteriors of the states in the linear system (2) are Gaussian, and this result is given by the following Lemma 2.

LEMMA 2 The posterior distribution of the system (2) is Gaussian if the initial distribution of the state is Gaussian.

The proof of this lemma can be found in the Appendix.

For the linear system (2), the optimal control input $(u, \tilde{\mathcal{K}})$ can be explicitly computed, and therefore, we have Theorem 1.

THEOREM 1 For the system (2), the optimal control input in (5) is $\tilde{\mathcal{K}}_t := \tilde{K}_t + \sigma_W(t)$, and $u(t, x) = -\tilde{\mathcal{K}}_tH_t\mu_t + P_tH_t^TQ_t^{-1}\frac{H_t\mu_t - H_tx}{2}$, where $\tilde{K}_t = P_tH_t^TQ_t^{-1}$ is the Kalman gain.

PROOF We can directly know that the evolution equation (5) with the optimal control terms for the system (2) is

$$d\tilde{x}_t^i = A_t\tilde{x}_t^i dt + \sigma_W(t)(dy_t - H_t\mu_t dt) + \sigma_B(t)dB_t^i + \tilde{K}_t\left(dy_t - \frac{H_t\tilde{x}^i + H_t\mu_t}{2}dt\right). \tag{7}$$

Since (7) is a linear SDE with a Gaussian initial value, the expectation and the variance of it can be directly calculated. Naturally, it is not difficult to find that it is the same as the KBF in Lemma 1. □

REMARK 1 The proposed FPF in [16] can be generalized only in this linear Gaussian setting; the existence of multidimensional correlated nonlinear FPF is still an open problem.

Therefore, μ_t and $\tilde{K}_t$ are approximated by

$$\mu_t \approx \tilde{\mu}_t^{(N)} := \frac{1}{N}\sum_{i=1}^{N}\tilde{x}_t^i$$

$$P_t \approx \tilde{P}_t^{(N)} := \frac{1}{N-1}\sum_{i=1}^{N}(\tilde{x}_t^i - \tilde{\mu}_t^{(N)})(\tilde{x}_t^i - \tilde{\mu}_t^{(N)})^T$$

$$\tilde{K}_t \approx K_t^{(N)} := \tilde{P}_t^{(N)}H^TQ^{-1} \tag{8}$$

in numerical computations. The complete procedure of the FPF for the system (2) is shown in Algorithm 1.

Although the FPF has good theoretical characteristics, it needs to sample Gaussian distribution at each step, which requires lots of calculation time in high-dimensional cases. In the numerical experiments, it can be found that this algorithm is not feasible in high-dimensional cases. In addition, the FPF can only work for time-invariant systems.

III. OPTIMAL TRANSPORTATION PARTICLE FILTER

In this section, we propose a new PF, where the particles are evolved in a deterministic way.

As we discussed before, for high-dimensional systems, sampling Gaussian distributions in the FPF is quite time consuming. However, the particles can be evolved in a deterministic way instead of stochastic way in the FPF. Therefore, we need to construct the functions $\mathcal{U}_t$ and $\mathcal{G}_t$

Algorithm 1: FPF for the Correlated Linear System (2).

Initialization
S is the total time, and m is the number of the discretization steps.
Set $n = 0$, $dt = \frac{S}{m}$ and $t_n = n \times dt$ holds.
for $i = 1$ to N **do**
  Sample $\tilde{x}_0^i \sim p_0$, where p_0 is the initial distribution.
Iteration 1: Predicting
$\tilde{\mu}_{t_n} := \frac{1}{N}\sum_{i=1}^{N}\tilde{x}_{t_n}^i$ as the filtering result at time t_n
for $i = 1$ to N **do**
  From $N(0, \sigma_B(t_n)\sigma_B^T(t_n)dt)$ to get samplings $\{\sigma_B(t_n)dB_{t_n}\}_{i=1}^N$.
  $\tilde{x}_{t_n^-}^i = A_{t_{n-1}}\tilde{x}_{t_{n-1}^-}dt + \tilde{x}_{t_{n-1}^-}^i + dB_{t_n}^i$
Iteration 2: Feedback Updating
Let $\tilde{\mu}_{t_n^-}^N \approx \frac{1}{N}\sum_{i=1}^{N}\tilde{x}_{t_n^-}^i$, and $\tilde{P}_{t_n} \approx \tilde{P}_{t_n}^{(N)} = \frac{1}{N-1}\sum_{i=1}^{N}(\tilde{x}_{t_n^-}^i - \tilde{\mu}_{t_n}^N)(\tilde{x}_{t_n^-}^i - \tilde{\mu}_{t_n}^N)^T$.
And $\tilde{K}_{t_n} \approx \tilde{P}_{t_n}^{(N)}H_{t_n}^T Q_{t_n}^{-1}$.
Put $\tilde{x}_{t_{n-1}^-}$, $\tilde{\mu}_{t_{n-1}}^N$, $\tilde{\mu}_{t_n^-}^N$, $dy_{t_{n-1}}$, dy_{t_n}, and K_{t_n} in (7) with the forward Euler scheme to get $\tilde{x}_{t_n}^i$.
$n = n + 1$
If $n \leq m$
  Then, go to Iteration 1
Else End

in the following equation:

$$dx_t^i = A_t x_t^i dt + \mathcal{U}_t(x_t^i, t)dt + \mathcal{G}(x_t^i, t)dy_t \tag{9}$$

such that the variance and the mean of x_t^i in (9) satisfy the same evolution equation in Lemma 1.

In this new framework (9), one needs to construct a set of transfer maps that will transfer samples of the initial distribution to samples with any posterior distribution [22]–[25]. However, the transfer maps between posteriors are not unique [28]. As an obvious example, we consider that any n-dimensional orthogonal transformation V can transfer the n-dimensional standard normal distribution to itself. Because the transfer maps are not unique, how to determine the specific form of the transfer maps becomes an important problem. Therefore, the OT map between distributions can naturally solve this problem.

A. Background of the OT

Let α and β be two probability measures on measure spaces Ω_X and Ω_Y, respectively, and we use $\mathcal{P}(\Omega)$ to denote the set of probability measures on Ω. Let $c : \Omega_X \times \Omega_Y \to [0, +\infty]$ be a cost function decided by some structures, and $c(x, y)$ measures the cost of transporting one unit of mass from $x \in \Omega_X$ to $y \in \Omega_Y$. First, the definition of the transport map is as follows.

DEFINITION 2 We say that $T : \Omega_X \to \Omega_Y$ transports $\alpha \in \mathcal{P}(\Omega_X)$ to $\beta \in \mathcal{P}(\Omega_Y)$, and we call T a transport map if

$$\beta(B) = \alpha\left(T^{-1}(B)\right) \quad \text{for all } \beta\text{-measurable sets } B. \tag{10}$$

We write

$$X\beta = T_{\#}\alpha \tag{11}$$

if (10) is satisfied.

With these notations, Monge's OT problem is formulated as follows.

DEFINITION 3 (SEE[30]) Monge's OT problem: Given $\alpha \in \mathcal{P}(X)$ and $\beta \in \mathcal{P}(Y)$

$$\text{minimize } I[T] = \int_X c(x, T(x))d\alpha(x)$$

over α-measurable maps $T : X \to Y$ subject to $\beta = T_{\#}\alpha$.

In most cases, we cannot obtain the explicit form of T. However, the posterior distributions of the filtering system we considered are Gaussian; therefore, we only need to focus on the OT problem between Gaussian distributions. Let us use $\mathcal{N}(\mu, P)$ to denote the Gaussian distribution with mean μ and covariance P. The OT between two Gaussian distributions is given in the following theorem.

THEOREM 2 (SEE [31, REMARK 2.31]) If $\alpha = \mathcal{N}(\mu_\alpha, P_\alpha)$ and $\beta = \mathcal{N}(\mu_\beta, P_\beta)$ are two Gaussians in $\mathbb{R}^n$ with $P_\alpha, P_\beta \succ 0$, then one can show that the following map:

$$T : x \to \mu_\beta + V(x - \mu_\alpha) \tag{12}$$

is the OT with cost function

$$c(\alpha, \beta) := \|\mu_\alpha - \mu_\beta\|_2^2 + \|P_\alpha^{\frac{1}{2}} - P_\beta^{\frac{1}{2}}\|_F^2 \tag{13}$$

where

$$XV = P_\alpha^{-\frac{1}{2}}(P_\alpha^{\frac{1}{2}}P_\beta P_\alpha^{\frac{1}{2}})^{\frac{1}{2}}P_\alpha^{-\frac{1}{2}}. \tag{14}$$

B. OT Structure in the Correlated System

In the FPF, each step requires independent sampling, which undoubtedly increases the operation time greatly in higher dimensional problems. Since the posterior distribution at each instant and the noise between the observed instants are both Gaussian, we can take advantage of Theorem 2 and give another deterministic evolution equation of the particles by the OT. Before we proceed, we need the following lemma.

LEMMA 3 Let P_t be the solution of (3) in Lemma 1; then, we have

$$P_t^{-\frac{1}{2}}(P_t^{\frac{1}{2}}P_{t+\Delta t}P_t^{\frac{1}{2}})^{\frac{1}{2}}P_t^{-\frac{1}{2}} = I + G_t\Delta t + O(\Delta t^2) \tag{15}$$

where Δt is a small positive real number, A, H, σ_B, σ_W, and Q are the smooth functions of t, and

$$G_t = A_t - \sigma_W(t)H_t + \frac{1}{2}\sigma_B(t)\sigma_B^T(t)P_t^{-1} - \frac{1}{2}P_t H_t^T Q_t^{-1}H_t + \Omega_t P_t^{-1}$$

where Ω_t is the solution to

$$\Omega_t P_t^{-1} + P_t^{-1}\Omega_t = A_t^T - A_t + \frac{1}{2}(P_t H_t^T Q^{-1}H_t - H_t^T Q_t^{-1}H_t P_t)$$

$$+\frac{1}{2}(\sigma_B(t)\sigma_B^T(t)P_t^{-1}-P_t^{-1}\sigma_B(t)\sigma_B^T(t)). \tag{16}$$

Proof: The solution P_t is positive and bounded since the system is observable [29]. Fix $t \in [0, S]$, and define

$$F(s) := P_t^{-\frac{1}{2}}(P_t^{\frac{1}{2}}P_{t+s}P_t^{\frac{1}{2}})^{\frac{1}{2}}P_t^{-\frac{1}{2}}.$$

Taking the Taylor expansion of $F(s)$ at $s = 0$, we obtain

$$F(\Delta t) = I + F'(0)\Delta t + \frac{1}{2}F''(\tau)\Delta t^2, \quad \tau \in [0, \Delta t].$$

Then, the following equation:

$$F(s)P_tF(s) = P_{t+s} \tag{17}$$

holds. Taking the derivative of both the sides of (17) with respect to s, we obtain

$$F'(0)P_t + P_tF'(0) = A_tP_t + P_tA_t^T + \sigma_B(t)\sigma_B^T(t) - [P_tH_t^T\sigma_W^T(t) + \sigma_W(t)H_tP_t + P_tH_t^TQ_t^{-1}H_tP_t]. \tag{18}$$

The linear equation of (18) has a special solution $G_0 = A_t - \sigma_W(t)H_t + \frac{1}{2}\sigma_B(t)\sigma_B^T(t)P_t^{-1} - \frac{1}{2}P_tH_t^TQ_t^{-1}H_t$; therefore, the solution of (18) will be $G_t = G_0 + \Omega_tP_t^{-1}$, where Ω_t can be any skew-symmetric matrix.

However, the matrix G_t is a symmetric matrix, so that $G_t = G_t^T$ holds, which is (16). From Lemma 2, (16) has a unique solution. □

Now, the evolution equation of the particles in our new OTPF is given in Theorem 3.

THEOREM 3 (OTPF) The evolution equation of the particles in the OTPF for the continuous system (2) is as follows:

$$dx_t^i = A_tx_t^idt + \sigma_W(t)(dy_t - H_tx_t^idt) + \Omega_tP_t^{-1}(x_t^i - \mu_t)dt + \frac{1}{2}\sigma_B(t)\sigma_B^T(t)P_t^{-1}(x_t^i - \mu_t)dt + K_t\left(dy_t - \frac{H_tx_t^i + H_t\mu_t}{2}dt\right) \tag{19}$$

where $K_t = P_tH_t^TQ_t^{-1}$ and Ω_t is the solution to (16).

Proof: First, we need to consider a discrete partition $\{0 = t_0 < \cdots < t_n = S\}$ of time interval $[0, S]$ with $t_{k+1} - t_k = \Delta t, k = 0, \ldots, n-1$. By Lemmas 1 and 2, we know that the posterior distributions of states x_{t_k} and $x_{t_{k+1}}$ are $\mathcal{N}(\mu_{t_k}, P_{t_k})$ and $\mathcal{N}(\mu_{t_{k+1}}, P_{t_{k+1}})$, respectively. Following Theorem 2, it is known that the optimal transport map T_{t_k} between these two Gaussians is

$$T_{t_k}(x_{t_k}^i) = \mu_{t_{k+1}} + V_{t_k}(x_{t_k}^i - \mu_{t_k}) \tag{20}$$

where

$$V_{t_k} = P_{t_k}^{-\frac{1}{2}}(P_{t_k}^{\frac{1}{2}}P_{t_{k+1}}P_{t_k}^{\frac{1}{2}})^{\frac{1}{2}}P_{t_k}^{-\frac{1}{2}}.$$

Using Lemma 3 and (5), we have

$$G_tP_t + P_tG_t = (A_t - \sigma_W(t)H_t)P_t + P_t(A_t^T - H_t^T\sigma_W^T(t)) + \sigma_B(t)\sigma_B^T(t) - P_tH_t^TQ_t^{-1}H_tP_t. \tag{21}$$

Thus, if P_t is nonsingular, for any system, there is exactly one solution G_t for each time t by Lemma 3. We can get $V_{t_k} = I_n + G_{t_k}\Delta t + O(\Delta t^2)$. Therefore, the particles $\{x_t^i\}$ in the OTPF have the following map from t_k to t_{k+1}:

$$x_{t_{k+1}}^i = \mu_{t_{k+1}} + (x_{t_k}^i - \mu_{t_k}) + G_{t_k}(x_{t_k}^i - \mu_{t_k})\Delta t + O(\Delta t^2). \tag{22}$$

Let $\Delta t \to 0$; then, one has

$$dx_t^i = d\mu_t + G_t(x_t^i - \mu_t)dt \tag{23}$$

so that we complete the proof. □

In the numerical experiments, μ_t, P_t, and Ω_t in (19) are approximated by

$$\mu_t \approx \mu_t^{(N)} := \frac{1}{N}\sum_{i=1}^N x_t^i$$

$$P_t \approx P_t^{(N)} := \frac{1}{N-1}\sum_{i=1}^N (x_t^i - \mu_t^{(N)})(x_t^i - \mu_t^{(N)})^T$$

$$G_t \approx G_t^{(N)} = A_t - \sigma_W(t)H_t + \frac{1}{2}\sigma_B(t)\sigma_B^T(t)P_t^{(N)^{-1}} - \frac{1}{2}P_t^{(N)}H_t^TQ_t^{-1}H_t + \Omega_t^{(N)}P_t^{(N)^{-1}}$$

and $\Omega_t^{(N)}$ is the solution to

$$\Omega_t^{(N)}P_t^{(N)^{-1}} + P_t^{(N)^{-1}}\Omega_t^{(N)} = A_t^T - A_t + \frac{1}{2}(P_t^{(N)}H_t^TQ_t^{-1}H_t - H_t^TQ_t^{-1}H_tP_t^{(N)}) + \frac{1}{2}(\sigma_B(t)\sigma_B^T(t)P_t^{(N)^{-1}} - P_t^{(N)^{-1}}\sigma_B(t)\sigma_B^T(t)).$$

REMARK 2 The stochastic term $\sigma_B(t)dB_t$ in (7) is replaced with the deterministic term $\sigma_B(t)\sigma_B^T(t)P_t^{-1}(x_t - m_t)dt$. Given a Gaussian prior, the two terms yield the same posterior. If we just assume that $\sigma_W = 0$ and $\Omega_t = 0$, the OTPF will exactly become the square-root EnKF [20]. Therefore, the OTPF can be considered as the extension of the EnKF as well.

The complete procedure of the OTPF is listed in Algorithm 2.

Using different algorithms to calculate $P_{t_n}^{-1}$ can produce different algorithms, and the details are in the next subsection.

C. Calculation of P_t^{-1}

In practical applications, the inverse matrix P_t^{-1} can be calculated by using some linear algebra tricks. For instance, we can calculate $P_t^{-1}b$ by solving linear equation $P_ta = b$ instead of calculating P_t^{-1}, where a and b are some vectors. There are many famous algorithms for this question (such as singular value decomposition (SVD) and lower–upper decomposition) [32]. In general, it is not necessary to use the SVD method all the time, but SVD is an effective method in the appropriate dimension.

In addition, we can reduce the computational complexity by evolving the dual system.

Algorithm 2: OTPF for the Correlated Linear System.

Initialization
S is the total time and m is the number of discretization steps.
Set $n = 0$, $dt = \frac{S}{m}$, and $t_n = n \times dt$ holds.
for $i = 1$ to N **do**
  Sample $x_0^i \sim p_0$, where p_0 is the initial distribution.
  The filtering result of the initial time is $\mu_0 := \frac{1}{N}\sum_{i=1}^N x_0^i$
Iteration 1: Offline calculation.
Let $\mu_0^{(N)} \approx \frac{1}{N}\sum_{i=1}^N x_{t_0}^i$.
calculate P_{t_n} by Lemma 1 with $P_0 \approx P_0^{(N)} \frac{1}{N-1}\sum_{i=1}^N (x_0^i - \mu_0)(x_0^i - \mu_0)^T$. (Solve the ODE system by the forward Euler scheme)
Calculate $P_{t_n}^{-1}$.
$K_{t_n} \approx K_{t_n}^{(N)} = P_{t_n}^{(N)} H_{t_n}^T$.
Calculate Ω_{t_n} by (15).
Iteration 2: Online updating
Put the $x_{t_{n-1}}^i$, $\mu_{t_n}^{(N)}$, $\mu_{t_{n-1}}^{(N)}$, $dy_{t_{n-1}}$, dy_{t_n}, $\Omega_{t_n}^{(N)}$, and $K_{t_n}^{(N)}$ in (19) with the forward Euler scheme to get $x_{t_n}^i$.
Output $\frac{1}{N}\sum_{i=1}^N x_{t_n}^i$ as the filtering result at time t_n
$n = n + 1$
If $n \le m$
Then, go to Iteration 1
Else End

Since $P_t P_t^{-1} = I_n$, we take the derivative at both the sides and obtain

$$\begin{aligned}\frac{d(P_t^{-1})}{dt} &= -P_t^{-1}\mathrm{Ricc}(P_t)P_t^{-1} \\ &= P_t^{-1}(-L) + (-L^T)P_t^{-1} - P_t^{-1}C^T C P_t^{-1} + M^T M\end{aligned} \tag{24}$$

where $\mathrm{Ricc}(P_t)$ is defined in (30) with the initial condition P_0^{-1}.

From (24), it is obvious that P_t^{-1} can be considered as the variance matrix of the following filtering system:

$$\begin{cases} dx_t^* = A_t^* x_t^* dt + \sigma_W^*(t) dW_t^* + \sigma_B^*(t) dB_t^* \\ dy_t^* = H_t^* x_t^* dt + dW_t^* \end{cases} \tag{25}$$

where $A_t^* = -A_t^T$, $\sigma_W^*(t) = -H_t^T \sigma_W^T(t)\sigma_B^{T,-1}(t)$, $\sigma_B^*(t) = Q^{\frac{1}{2}} H_t$, $H_t^* = \sigma_B^T(t)$, and dB_t^* and dW_t^* are the m-dimensional and n-dimensional standard Brownian motion, respectively.

It can be easily verified that the variance of (25) satisfies (24) according to the KBF.

Since the system (25) is a linear system, we can easily construct the OTPF equation for it. The covariance of (25) is exactly P_t^{-1}. The particles $\{x_t^{*,i}\}_{i=1}^N$ that are determined by the OTPF of (25) can approximate P_t^{-1} by the following equation:

$$P_t^{-1} \approx P_t^{*,(N)}, \quad P_t^{*-1} \approx P_t^{(N)}. \tag{26}$$

Next, we summarize the dual-dynamic system for the OTPF in the following theorem.

THEOREM 4 (DUAL SYSTEM) The evolution equation of the particles in the dual system of the OTPF for the continuous system (2) is as follows:

$$\begin{aligned} dx_t^{*,i} &= A_t^* x_t^{*,i} dt + \sigma_W^*(t)(dy_t^* - H_t^* x_t^{*,i} dt) \\ &\quad + \Omega_t^* P_t^{*,-1}(x_t^{*,i} - \mu_t)dt \\ &\quad + \frac{1}{2}\sigma_B^*(t)\sigma_B^{*,T}(t) P_t^{*,-1}(x_t^{*,i} - \mu_t^*)dt \\ &\quad + K_t^*(dy_t^* - \frac{H_t x_t^{*,i} + H_t \mu_t^*}{2} dt). \end{aligned} \tag{27}$$

In the real application, there are no observation data dy_t^* for the dual system. However, it is reasonable to consider the innovation process of the original system can transfer to the innovation process of the dual system, since the variance is irrelevant with the observation process. $dI_t^* = dy_t^* - H^* \mu_t^* dt$ is a Brownian motion with $E[dI_t^* dI_t^{*,T}] = I_n dt$, and the original $dI_t = dy_t - H\mu_t dt$ is a Brownian motion with $E[dI_t dI_t^T] = Q_t dt$. dI_t^* is determined by dI_t with the transfer map $(\gamma_t Q_t^{-\frac{1}{2}})$, where γ_t is the projection matrix of the first n components. $(\gamma_t Q_t^{-\frac{1}{2}})$ naturally solves the equation $V_t Q_t V_t^T = I_n$ of V_t.

It naturally requires that the observation dimension cannot be less than the system dimension. Otherwise, there is no determined map from dI_t to dI_t^*, which means that the dual system needs extra information for the evolution.

Therefore, we can have the estimate of P_t^{-1} from the dual system

$$P_t^{-1} \approx P_t^{*,(N)} = \frac{1}{N-1}\sum_{i=1}^N (x_t^{*,i} - \mu_t^{*,(N)})(x_t^{*,i} - \mu_t^{*,(N)})^T \tag{28}$$

where $\mu_t^* = \frac{1}{N}\sum_{i=1}^N x_t^{*,i}$.

D. Singular Covariance of State

The derivation of the OTPF crucially relies on the assumption that P_t is positive definite. In general, when the covariance of Gaussian random variables x_{t_k} or $x_{t_{k+1}}$ is singular, the optimal transport map V does not exist. The simplest example is that there is no deterministic linear map that transfers a 1-D Gaussian random variable to a 2-D Gaussian random variable. However, this difficulty will be resolved as we reintroduce noise, which is the so-called Kantorovich relaxation.

Let Y_1 and Y_2 be the 1-D standard Gaussian and the 2-D standard Gaussian, respectively. There is no a deterministic map V, which satisfies $V(Y_1) = Y_2$. However, a stochastic map [30] exists and is given by

$$Y_2 = Y_1 \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} B$$

where B is a 1-D standard Gaussian and is independent of Y_1.

From the above analysis, even it is indeed impossible to construct an explicit expression of the OT when the system state covariance is singular, we can still transfer the particles in an FPF way. Besides, when some components of the distribution degenerate, we can still use the FPF structure in Algorithm 1.

In a word, the following formula can be obtained by rewriting (23) as the most general conclusion

$$dx_t^i = d\mu_t + \hat{G}_t(x_t^i - \mu_t)dt + \Gamma_t \sigma_B dB_t \tag{29}$$

where Γ_t is the projective operator to the kernel of P_t and $\hat{G}_t$ is the OT in the orthogonal complement of the kernel of P_t.

REMARK 3 So far, we can clearly describe the alternative relationship between stochastic evolution and deterministic evolution. There will be many hybrid methods derived from this alternative relationship.

IV. STABILITY AND CONVERGENCE FOR THE OTPF AND THE FPF

In this section, we focus on the system (2), in which A, H, σ_B, and σ_W are constants. We will discuss the convergence of our OTPF and FPF.

A. Riccati Flow

We start with the correlated Riccati flow.

DEFINITION 4 A Riccati quadratic operator is defined as

$$\text{Ricc}(P_t) := LP_t + P_t L^T - P_t M^T M P_t + C^T C \tag{30}$$

where

$$L = A - \sigma_W H, \quad M = Q^{-\frac{1}{2}} H, \quad C = \sigma_B. \tag{31}$$

A Riccati quadratic system is as follows:

$$\frac{dP_t}{dt} = \text{Ricc}(P_t). \tag{32}$$

For the error analysis, the following assumptions are made.

Assumption 1 The system (2) is detectable and $A - \sigma_W H, \sigma_B \sigma_B^T, \sigma_W Q, Q,$ and R are constant matrices and stable.

Assumption 2 The initial covariance matrix is positive definite.

We consider the bound condition

$$\text{Ricc}(P) = 0. \tag{33}$$

Let Ψ_t be the transition matrix of the following form and above notion:

$$\frac{d}{dt}\Psi_t = (L - P_t M^T M)\Psi_t, \Psi_0 = I \tag{34}$$

where P_t is the solution for (30). Therefore, we have the following lemma.

LEMMA 4 (SEE[29] AND [33]) Consider equations in Lemma 1 and (33). Then, under Assumption 1, we will have the following result.

1) There exists a unique positive-definite solution P_∞ to (33).
2) The explicit solution to the Riccati flow (30) with notation (31) is given by
$$P_t = P_\infty + e^{F_\infty t} D_t^{-1} e^{F_\infty t} \tag{35}$$
where $F_\infty = A - \sigma_W H - P_\infty H Q^{-1} H^T$ and $D_t = (P_0 - P_\infty)^{-1} + \int_0^t e^{F_\infty s} H^T H e^{F_\infty s} ds$. The real part of eigenvalue in F_∞ is all negative and less than $-\eta_0$ for some $\eta_0 > 0$.
3) If the initial covariance matrix P_0 is strictly positive definite, then there exist α_0 and β_0, such that the solution P_t satisfies
$$\alpha_0 I_n \prec P_t \prec \beta_0 I_n$$
where $\alpha_0 I_n \prec P_t$ means that $P_t - \alpha_0 I_n$ is a positive-definite matrix.
4) The covariance $P_t \to P_\infty$ exponentially fast which means that for any $0 < \eta < \eta_0$
$$\lim_{t\to\infty} \|P_t - P_\infty\|_F \leq ce^{-2\eta t}$$
where η_0 is the same as that in (2).

B. Convergence of the Correlated OT Filter

Now, we will give the error analysis of our OTPF, and the conditional expectation μ_t is approximated by μ_t^N satisfying the following equations:

$$\begin{aligned} dx_t^i &= A\mu_t^{(N)}dt + \sigma_W(dy_t - H\mu_t^{(N)}dt) \\ &\quad + K^{(N)}(dy_t - H\mu^{(N)} + G^{(N)}(x_t^i - \mu_t^{(N)})dt \end{aligned} \tag{36}$$

which is directly from (23). Then, we take the sum of (36); $\mu_t^{(N)}$ satisfies the following equations:

$$\begin{aligned} d\mu_t^{(N)} &= A\mu_t^{(N)}dt + \sigma_W(dy_t - H\mu_t^{(N)}dt) \\ &\quad + K^{(N)}(dy_t - H\mu^{(N)}dt) \end{aligned} \tag{37}$$

$$dP_t^{(N)} = \text{Ricc}(P_t^{(N)})dt \tag{38}$$

$$d\zeta_t^i = G^{(N)}\zeta_t^i dt \tag{39}$$

where $\mu_t^{(N)} := \frac{1}{N}\sum_{i=1}^N x_t^i$, $P_t^{(N)} := \frac{1}{N-1}\sum_{i=1}^N (x_t^i - \mu_t^{(N)})(x_t^i - \mu_t^{(N)})^T$, $K_t^{(N)} = P_t^{(N)} H^T$, $\zeta_t^i := x_t^i - \mu_t^{(N)}$, and

$$\begin{aligned} G_t^{(N)} &= A - \sigma_W H + \frac{1}{2}\sigma_B \sigma_B^T P_t^{(N)-1} \\ &\quad - \frac{1}{2} P_t^{(N)} H^T Q^{-1} H + \Omega_t^{(N)} P_t^{(N)-1} \end{aligned}$$

where $\Omega_t^{(N)}$ is the solution to

$$\begin{aligned} \Omega_t^{(N)} P_t^{(N)-1} + P_t^{(N)-1}\Omega_t^{(N)} &= A^T - A \\ &\quad + \frac{1}{2}(P_t^{(N)} H^T Q^{-1} H \\ &\quad - H^T Q^{-1} H P_t^{(N)}) \\ &\quad + \frac{1}{2}(\sigma_B \sigma_B^T P_t^{(N)-1} - P_t^{(N)-1}\sigma_B \sigma_B^T) \end{aligned}$$

with notations (31).

Next, we will introduce the results of consistency of $\mu_t^{(N)}$ and $P_t^{(N)}$.

THEOREM 5 Consider the correlated filtering system (2) initialized with the prior x_0, and the variance is strict positive definite. The estimate $\mu_t^{(N)}$ satisfies the system (37). Under Assumptions 1 and 2, we have the following convergence and error properties results.

1) For any $N > 1$

$$\lim_{t\to\infty} e^{2\eta t}\|\mu_t^{(N)} - \mu_t\|_2 = 0 \tag{40}$$

$$\lim_{t\to\infty} e^{2\eta t}\|P_t^{(N)} - P_t\|_F = 0. \tag{41}$$

2) For any $t > 0$ and as $N \to \infty$

$$E[\|\mu_t^{(N)} - \mu_t\|_2^2] \le C_m \frac{\mathrm{Tr}(P_0) + \mathrm{Tr}(P_0)^2}{N} e^{-2\eta t} \tag{42}$$

$$E[\|P_t^{(N)} - P_t\|_F^2] \le C_P \frac{\mathrm{Tr}(P_0)^2}{N} e^{-4\eta t}. \tag{43}$$

For all $0 < \eta < \psi(A - P_W H)$; $\psi(B) := \min\{-\lambda|\lambda$ is the eigenvalue of $B\}$; C_m and C_p are some constants.

Proof: The difference $P_t^{(N)} - P_t$ [see (32)–(37)] is

$$\begin{aligned} &d(P_t^{(N)} - P_t) \\ &= (A - \sigma_W H - P_t H^T Q^{-1} H)(P_t^{(N)} - P_t)dt \\ &\quad - (P_t^{(N)} - P_t)(A - \sigma_W H - P_t H^T Q^{-1} H)^T dt. \end{aligned}$$

Therefore, we need to analyze the system

$$\begin{aligned} \frac{d}{dt}X(t) &= (A - \sigma_W H - P_t H^T Q^{-1} H)X_t \\ &= F_t X_t = F_\infty X_t + (P_\infty - P_t)H^T Q^{-1} H \end{aligned} \tag{44}$$

where $F_t = A - \sigma_W H - P_t H^T Q^{-1} H$ and $F_\infty = A - \sigma_W H - P_\infty H^T Q^{-1} H$. This is a standard ODE; therefore, the solution can be expressed as

$$X_t = e^{tF_\infty} X_s + \int_s^t e^{(t-\tau)F_\infty}(P_\infty - P_t)H^T Q^{-1} H d\tau. \tag{45}$$

Grönwall's inequality is, then, used to conclude that

$$\begin{aligned} &\|X_t\|_2 \le \|e^{tF_\infty}\|_2 \|X_s\|_2 + \\ &\int_s^t \|e^{(t-\tau)F_\infty}\|_2 \|P_\infty - P_t\|_F \|H^T Q^{-1} H\|_2 \|X_\tau\| d\tau. \end{aligned} \tag{46}$$

Here, we need an error analysis for $\|e^{tF_\infty}\|$. First, we consider the Jordon standard form of $\|e^{tF_\infty}\|_2 = \|OJO^{-1}\|_2$. Owing to linear algebra and Lemma 4, the norm is bounded

$$\|e^{tF_\infty}\|_2 \le \|O\|_2 \|O^{-1}\|_2 \left(\max_{0\le k\le n} \frac{t^k}{k!}\right) e^{-\eta_0 t} \tag{47}$$

where η_0 is the largest multiplicity of the eigenvalues of F_∞. Then, for all $0 < \eta < \eta_0$, there exists a constant c

$$\|e^{tF_\infty}\|_2 \le ce^{-\eta t}.$$

Therefore, using $\|e^{tF_\infty}\|_2 \le ce^{-\eta t}$ and (46), we have

$$\begin{aligned} &\|X_t\|_2 \le ce^{-\eta t}\|X_s\|_2 + \\ &\int_s^t ce^{-\eta(t-\tau)}\|P_\infty - P_t\|_F \|H^T Q^{-1} H\|_2 \|X_\tau\|_2 d\tau. \end{aligned} \tag{48}$$

Grönwall's inequality is used to conclude that

$$\|X_t\|_2 \le c' e^{-\eta(t-s)}\|X_s\|_2 e^{c'\|H^T Q^{-1} H\|_2 \int_s^t \|P_t - P_\infty\|_F ds}. \tag{49}$$

This shows that the transition matrix Ψ_s^t, where $\Psi_s^t X_s = X_t$ holds, for the linear system (44) is bounded as follows:

$$\|\Psi_s^t\|_2 \le c' e^{-\eta(t-s)c'\|H^T Q^{-1} H\|_2 \int_s^t \|P_t - P_\infty\|_F ds}. \tag{50}$$

Now, because of the exponential convergence from Lemma 4, we have

$$\|\Psi_s^t\|_2 \le c' e^{-\eta(t-s)c'\|H^T Q^{-1} H\|_2 \frac{c\|P_0 - P_\infty\|_F}{2\eta}}. \tag{51}$$

The equation in Lemma 1 and the variance equation in (37) are similar. Furthermore, consider that the empirical counterpart of the linear system replaces F_t, P_t, and Ψ_s^t with $F_t^{(N)}$, $P_t^{(N)}$, and $\Psi_s^{t,(N)}$, respectively. Now, we can simplify (44) by the above approach

$$P_t^{(N)} - P_t = \Psi_0^t(P_0^{(N)} - P_0)(\Psi_0^{t,(N)})^T. \tag{52}$$

Therefore

$$\begin{aligned} \|P_t^{(N)} - P_t\|_F &\le \|\Psi_0^t\|_2 \|(\Psi_0^{t,(N)})^T\|_2 \|P_0^{(N)} - P_0\|_F \\ &\le c^2 e^{-2\eta t}\|P_0^{(N)} - P_0\|_F. \end{aligned} \tag{53}$$

Taking the expectation of both the sides, we obtain

$$\begin{aligned} E[\|P_t^{(N)} - P_t\|_F] &\le c^4 e^{-4\eta t} E[\|P_0^{(N)} - P_0\|_F] \\ &= c^4 e^{-4\eta t}\frac{1}{N}E[\|\mathrm{Tr}(\zeta_0^i \zeta_0^{i\,T}) - P_0)^2] \\ &\le c^4 e^{-4\eta t}\frac{E[\|\zeta_0^i\|_2^4]}{N} = c^4 e^{-4\eta t}\frac{3\mathrm{Tr}(P_0)^2}{N}. \end{aligned} \tag{54}$$

Then, the error analysis of the mean square estimation can be given. Subtracting conditional mean from (37) for Lemma 1, we have

$$\begin{aligned} d\mu_t^{(N)} - d\mu_t &= (A - \sigma_W H - P_t^{(N)} H^T Q^{-1} H)(\mu_t^{(N)} - \mu_t)dt \\ &\quad + (P_t^{(N)} - P_t)H^T Q^{-1} H dI_t \end{aligned} \tag{55}$$

where $dI_t = dy_t - H\mu_t dt$ is the innovation process. Its solution is given by

$$\begin{aligned} \mu_t^{(N)} - \mu_t &= \Psi_0^{t,(N)}(\mu_0^{(N)} - \mu_0) \\ &\quad + \int_0^t \Psi_s^{t,(N)}(P_s^{(N)} - P_s)H^T Q^{-1} H dI_s. \end{aligned} \tag{56}$$

The norm of the first term of (55) is bounded by

$$\begin{aligned} E[\|\Psi_0^{t,(N)}(\mu_0^{(N)} - \mu_0)\|_2^2] &\le c^2 e^{-2\eta t} E[\|\mu_0^{(N)} - \mu_0\|_2^2] \\ &\le c^2 e^{-2\eta t}\frac{\mathrm{Tr}(P_0)}{N}. \end{aligned} \tag{57}$$

The norm of the second term of (55) is bounded by

$$\begin{aligned} &E\left[\|\int_0^t \Psi_s^{t,(N)}\left(P_s^{(N)} - P_s\right)H^T Q^{-1} H dI_s\|_2^2\right] \\ &= \int E\left[Tr\left(\Psi_s^{t,(N)}(P_s^{(N)} - P_s)H^T Q^{-1} H\right.\right. \\ &\qquad \left.\left.(P_s^{(N)} - P_s)^T\right)\left(\Psi_s^{t,(N)}\right)^T\right] ds \end{aligned}$$

$$\leq \int_0^t E[\|\Psi_s^{t,(N)}(P_s^{(N)} - P_s)\|_F^2 \|H\|_2^2]ds$$
$$\leq \|H\|_2^2 \int_0^t c^2 e^{-2\eta(t-s)} c^4 e^{-4\eta s} E\left[\|P_0^{(N)} - P_0\|_F^2\right] ds$$
$$\leq c^6 \|H\|_2 \frac{3\mathrm{Tr}(P_0)^2}{N} \frac{e^{-2\eta t}}{2\eta}. \quad (58)$$

The first row leads to the second row using a quadratic variation process [34], which is $\langle dI_s \rangle = Q_s ds$. Adding (56) and (58), we complete the proof. □

C. Convergence of the Correlated FPF

In this subsection, we prove the convergence of the 1-D linear correlated FPF.

Consider the N particle system. We define the error processes $\tilde{\zeta}_t^i := \tilde{x}_t^i - \mu_t$. The evolutions for the mean $\tilde{\mu}_t^{(N)}$, the covariance $\tilde{P}_t^{(N)}$, and the error $\tilde{\zeta}_t^i := \tilde{x}_t^i - \mu_t$ are determined by the following equations.

LEMMA 5 The error process $\tilde{\zeta}_t^i = \tilde{x}_t^i - \mu_t$ for the FPF is determined by the following equation:

$$d\tilde{\zeta}_t^i = \left\{A - \sigma_W H - \frac{P_t H^T Q^{-1} H}{2}\right\} \tilde{\zeta}_t dt + \sigma_B dB_t^i. \quad (59)$$

$\tilde{\zeta}_t^i$ is the Gaussian at any time t if the initial $\tilde{\zeta}_0^i$ is the Gaussian.

Proof: It can be obtained directly from (7) and Lemma 1. □

Similar to (37), we have

$$d\tilde{\mu}_t^{(N)} = A\tilde{\mu}_t^{(N)} dt + \sigma_W(dy_t - H_t \tilde{\mu}_t^{(N)} dt) + \sigma_B dB_t^{(N)} + K^{(N)}(dy_t - H\tilde{\mu}_t^{(N)} dt) \quad (60)$$
$$d\tilde{P}_t^{(N)} = \mathrm{Ricc}(\tilde{P}_t^{(N)})dt + dZ_t + dZ_t^T \quad (61)$$
$$d\tilde{\zeta}_t^i = G_t \tilde{\zeta}_t^i dt + \sigma_B dB_t^i - \sigma_B dB_t^{(N)} \quad (62)$$

where $dB_t^{(N)} := \frac{1}{N}\sum_{i=1}^N dB_t^i$ and $dZ_t := \frac{1}{N-1}\sum_{i=1}^N (\sigma_B dB_t^i \zeta_t^{T,i} + \zeta_t^i dB_t^{i,T} \sigma_B^T)$.

Then, we can conclude the following convergence analysis.

THEOREM 6 Consider the SDE system (60). Under Assumptions 1 and 2 and for the state dimension with 1, we have the following.

1) For any $t > 0$, $p > 1$, and $N > 4p$

$$E[(\tilde{P}_t^{(N)} - P_t)^{2p}]^{\frac{1}{p}} \leq \frac{C_1}{N} e^{-\beta t} + \frac{C_2}{N}. \quad (63)$$

2) For any $t > 0$ and as $N \to \infty$

$$E[(\tilde{\mu}_t^{(N)} - \mu_t)^2] \leq \frac{P_0}{N} e^{-\psi(A - \sigma_W H)t} + \frac{C_3}{N} \quad (64)$$

where $\psi(B) := \min\{-\lambda | \lambda$ is the eigenvalue of $B\}$ and C_1, C_2, and C_3 are some constant numbers.

Proof: We can obtain the results following the similar procedure in the proof of Theorem 5. A, H, σ_B, σ_W, μ_t, P_t, and Q are scales.

The evolution of the difference $\tilde{P}_t^{(N)} - P_t$ is

$$d(\tilde{P}_t^{(N)} - P_t) = 2\left(A - \sigma_W H - \frac{\tilde{P}_t^{(N)} + P_t}{2} H^2 Q^{-1}\right)(\tilde{P}_t^{(N)} - P_t)dt + 2dZ_t.$$

Define $\tilde{R}_t := E[(\tilde{P}_t^{(N)} - \tilde{P}_t)^{2p}]$, and by Itô's formula, we have

$$\frac{d\tilde{R}_t}{dt} = E\left[4p\left(A - \sigma_W H - \frac{\tilde{P}_t^{(N)} + P_t}{2} H^2 Q^{-1}\right)\left(\tilde{P}_t^{(N)} - P_t\right)^{2p}\right] + p(2p-1)\frac{4\sigma_B^2}{N-1} E[\tilde{P}_t^{(N)}(\tilde{P}_t^{(N)} - P_t)^{2p-2}]. \quad (65)$$

It is easy to see that

$$\left(A - \sigma_W H - \frac{\tilde{P}_t^{(N)} + P_t}{2} H^2 Q^{-1}\right) \leq G_t(P_t).$$

By mean inequality and $P_t \leq P_\infty + P_0$, we have

$$\tilde{P}_t^{(N)} = (\tilde{P}_t^{(N)} - P_t) + P_t \leq \frac{(\tilde{P}_t^{(N)} - P_t)^2}{2P_\infty} + \tilde{P}_0 + 2P_\infty.$$

From Hölder's inequality, we have $E[(\tilde{P}_t^{(N)} - P_t)^{2p-2}] \leq R_t^{\frac{p-1}{p}}$. Therefore

$$\frac{d\tilde{R}_t}{dt} \leq \left(4pG_t(P_t) + \frac{4p(p-1)\sigma_B^2}{2P_\infty(N-1)}\right) R_t + \frac{4p(p-1)\sigma_B^2}{(N-1)}(2P_\infty + P_0)\tilde{R}_t^{\frac{p-1}{p}}. \quad (66)$$

Therefore

$$\frac{d\tilde{R}_t^{\frac{1}{p}}}{dt} \leq 4(G_t(P_t) + \frac{4p(p-1)\sigma_B^2}{2P_\infty(N-1))}\tilde{R}_t^{\frac{1}{p}} + \frac{4(2p-1)\sigma_B^2(2P_\infty + P_0)}{N-1}. \quad (67)$$

Using the linear transition result from points 3 and 4 of Lemma 4 with the same notation, we have

$$\frac{d\tilde{R}_t^{\frac{1}{p}}}{dt} \leq c^4 e^{-4\eta t + \frac{4(2p-1)\eta_0}{N-1}t} \tilde{R}_t^{\frac{1}{p}} + \frac{c^4}{4\eta_0 - \frac{4(2p-1)\eta_0}{N-1}} \frac{4(2p-1)\sigma_B^2(2P_\infty + P_0)}{N-1} \quad (68)$$

using $N > 4p$, which proves the first part.

The difference $\mu_t^{(N)} - \mu_t$ satisfies

$$d(\tilde{\mu}_t^{(N)} - \mu_t) = (A - \sigma_W H - \tilde{P}_t^{(N)} H^2 Q^{-1}) + (\tilde{P}_t^{(N)} - P_t)HQ^{-1}(dy_t - H\mu_t dt) + \sigma_B dB_t^{(N)}. \quad (69)$$

Therefore, by the application of Itô's rule, we have

$$\frac{d}{dt}E[(\tilde{\mu}_t^{(N)} - \mu_t)^2] = E[2(A - \sigma_W H - \tilde{P}_t^{(N)} H^2 Q^{-1})(\tilde{\mu}_t^{(N)} - \mu_t)]$$

$$+ E\left[(\tilde{P}_t^{(N)} - P_t)^2 H^2 Q^{-1}\right] + \frac{\sigma_B^2}{N}. \tag{70}$$

P_t is bounded by using the bounded condition assumption, we can obtain

$$\frac{d}{dt}E[(\tilde{\mu}_t^{(N)} - \mu_t)^2] \leq -2\psi(A - \sigma_W H)E[(\tilde{\mu}_t^{(N)} - \mu_t)^2] + \frac{c_1 e^{-\eta t + c_2}}{N}H^2 Q^{-1} + \frac{\sigma_B^2}{N}. \tag{71}$$

The application of Grönwall's inequality concludes the second part. □

REMARK 4 Here, the analysis of the OTPF and the FPF for the continuous system is given. Furthermore, the new convergence order of the MSE is $\mathcal{O}(\frac{1}{N})$, which is better than $\mathcal{O}(\frac{1}{N^{2-\frac{2}{r}}})$ with $1 \leq r < 2$ in general result for the PF [35] [27].

V. NUMERICAL EXPERIMENTS FOR ALGORITHMS

In this part, we will test the efficiency of our proposed OTPF. We consider two numerical examples. The first one is a scalar case and the second one is a high-dimensional case with the dimension varying from 25 to 200. We repeated each experiment 20 times and take the average of the MSE and time. The definitions of the mean of mean square error (MMSE) and the mean of time (MT) are as follows:

$$\text{MMSE} := \frac{1}{20}\sum_{i=1}^{20}\left[\frac{1}{\text{NS}+1}\sum_{j=0}^{NS}\left(\hat{x}_{t_j}^i - \mu_{t_j}^i\right)^2\right] \tag{72}$$

$$\text{MT} := \frac{1}{20}\sum_{i=1}^{20}\text{RT}_i \tag{73}$$

where NS means the number steps of the experiments, the RT means the total running time of the ith experiment, $\hat{x}_{t_j}^i$ is the filtering result of the ith experiment at time t_j, and $\hat{x}_{t_j}^i$ is the estimation of $\mu_{t_j}^i$. The EnKF used in the experiments is from [18].

There are two kinds of OTPF in our simulations, which the OTPF uses the linear algebra method to calculate P_t^{-1} and the OTPF DP uses the dual-particle flow to approximate P_t^{-1}. The selection of initial particles has a great influence on the approximation of P_t by the dual system, and it is important to make the initial $P_0^{*,(N)}$ close enough to P_0^{-1}.

We run our simulations on CPU clusters with Intel Core i9-9880H(2.3CGz/L3 16M) equipped with 16-GB memory. Besides, we use NumPy, which is a python package for scientific computing in algorithms, and we use time, which is a python package for obtaining the CPU running time of algorithms.

A. Scalar Case

The scalar system is as follows:

$$\begin{cases} dx_t = Ax_t dt + \sigma_B dB_t + \sigma_W dW_t \\ dy_t = Hx_t dt + dW_t \end{cases} \tag{74}$$

TABLE I
Parameter Setting in Simulation and Experiment Result

Parameters	N	σ_B	σ_W	Q	MMSE	MT
EnKF	5	20	0.2	2	3.0576	0.414
FPF	5	20	0.2	2	55.6329	0.5415
OTPF	5	20	0.2	2	0.6335	0.1997
EnKF	5	10	0.2	2	0.7752	0.4136
FPF	5	10	0.2	2	3.5107	0.5230
OTPF	5	10	0.2	2	0.4331	0.3041
EnKF	5	5	0.2	2	0.4336	0.4233
FPF	5	5	0.2	2	1.8106	0.5437
OTPF	5	5	0.2	2	0.3138	0.2987
EnKF	5	2	0.2	2	0.1650	0.4261
FPF	5	2	0.2	2	0.6967	0.5428
OTPF	5	2	0.2	2	0.1527	0.3034
EnKF	5	1	0.2	1	0.1025	0.4261
FPF	5	1	0.2	1	0.3643	0.5428
OTPF	5	1	0.2	1	0.0906	0.3034
EnKF	5	1	0.2	5	0.0864	0.4261
FPF	5	1	0.2	5	0.3567	0.5428
OTPF	5	1	0.2	5	0.0848	0.3034
EnKF	5	2	0.2	5	0.1740	0.4287
FPF	5	2	0.2	5	0.7222	0.5354
OTPF	5	2	0.2	5	0.2072	0.3010

TABLE II
Parameter Setting in Simulation and Experiment Result

Parameters	N	σ_B	σ_W	Q	MMSE	MT
EnKF	5	2	0.2	2	0.1650	0.4261
FPF	5	2	0.2	2	0.6967	0.5428
OTPF	5	2	0.2	2	0.1527	0.3034
EnKF	10	2	0.2	2	0.1547	0.8267
FPF	10	20	0.2	2	0.4213	1.500
OTPF	10	20	0.2	2	0.1369	0.4861
EnKF	15	2	0.2	2	0.1376	1.1937
FPF	15	2	0.2	2	0.1618	2.722
OTPF	15	2	0.2	2	0.1351	0.6560

where B_t is a standard Brownian motion and the initial distribution of x_0 is $\mathcal{N}(1, 1)$. Here, we investigate the system (74) with varying A, σ_B, σ_W, H, and Q, where Q is the covariance matrix of W_t.

In Table I, the total time of the simulation is 40˜s and the number of steps is 4000, so that dt is 0.01 s, the system matrix $A = -0.5$, and observation matrix $H = 1$. From Table I, we can find out our new method OTPF is highly robust for noise term, such as σ_W, σ_B, and Q. Our algorithm still maintains high efficiency even if the standard deviation of the noise is 20, and the other algorithms are all failed.

In Table II, the total time of the simulation is 40 s and the number of steps is 4000, so that dt is 0.01 s, the system matrix $A = -0.5$, and observation matrix $H = 1$. From Table II, our new algorithm requires fewer particles than other algorithms. The FPF needs three times of the particle number to have the same performance. As the dimensional grows, the computational complexity of the OTPF does not grow as fast as that of the other two algorithms.

In Table III, the total time of the simulation is 4 s and the number of steps is 400, so that dt is 0.01 s, $\sigma_B = 1$, $Q = 1$, and $\sigma_W = 0.2$. The particle number is 5. We can find out

TABLE III
Parameter Setting in Simulation and Experiment Result

Parameters	N	A	H	MMSE	MT
EnKF	5	-0.5	1	0.0695	0.0827
FPF	5	-0.5	1	0.2803	0.2918
OTPF	5	-0.5	1	0.0595	0.0820
EnKF	5	-0.5	2	0.1483	0.0528
FPF	5	-0.5	2	0.2671	0.0867
OTPF	5	-0.5	2	0.1287	0.0382
EnKF	5	0	1	0.2094	0.2081
FPF	5	0	1	0.7336	0.5815
OTPF	5	0	1	0.1629	0.1086
EnKF	5	0	2	0.7019	0.2045
FPF	5	0	2	0.8062	0.6049
OTPF	5	0	2	0.5482	0.1127
EnKF	5	0.5	1	0.4138	0.2038
FPF	5	0.5	1	1.064	0.5797
OTPF	5	0.5	1	0.1901	0.1060
EnKF	5	2	0.2	0.1376	1.1937
FPF	5	2	0.2	0.1618	2.722
OTPF	5	2	0.2	0.1351	0.6560

that the OTPF is better than any other algorithms in several situations.

B. Vector Case

In the vector system, the total time of the simulation is 10 s and the number of steps is 1000, so that dt is 0.01 s.

The high-dimensional example considered here is as follows:

$$\begin{cases} dx_t = A_n x_t dt + 1.5 I_n dB_t + 0.3 I_n dW_t \\ dy_t = H_n x_t dt + dW_t \end{cases} \quad (75)$$

where

$$A_n = \begin{pmatrix} a & b & 0 & \cdots & 0 \\ c & a & b & \cdots & 0 \\ 0 & c & a & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & c & a \end{pmatrix} \in M_n(\mathbb{R})$$

with $a = -0.2,\ b = -0.1,\ c = 0, H_n = I_n \in \mathbb{R}^n$ is the identity matrix, and dB_t and dW_t are the standard n dimensional Brownian motions. The initial distribution of x_0 is $\mathcal{N}(\mu_n, 2I_n)$, where the first $[\frac{n}{2}]$ ($[\cdot]$ is the rounding function) components in μ_n are 1, and the others are -1.

REMARK 5 The online–offline approach can be applied to the OTPF. The offline part of the OTPF is to compute P_t and Ω_t, and the others are the online part. It is easy to see that P_t and Ω_t in the OTPF are independent of the observation.

In Table IV, the total time of the simulation is 40 s with 1000 number steps. The OLT in Table V means the online time.

From Table IV, it is not difficult to see that the time required by the FPF increases significantly as the dimension increases compared with the other two methods, because of the complexity of high-dimensional sampling. It can also be seen that the EnKF cannot effectively improve the accuracy of the algorithm with the increase in the number of particles. The OTPF performs well in any situation. After the improvement of the online and offline approaches, the computational complexity of the OTPF is almost the same as that of the EnKF, and it performs better than the FPF with the same parameters.

In order to show the convergence of the ensemble variance, we give the FN of the error between the optimal variance and the ensemble variance. The time steps in Table IV are 0.01 s, and the system is 100-D in (75). We

TABLE IV
Error of the Ensemble Variance in 100 Dimension With Time Steps 0.01 s

Algorithms	Norm for error	100 steps	500 steps	1000 steps
OTPF	FN	24.1743	12.178	9.337
OTPF	CA	0.242	0.122	0.0934
OTPF DP	FN	30.1743	20.028	17.174
OTPF DP	CA	0.3017	0.2003	0.172
FPF	FN	57.3348	56.0055	56.5605
FPF	CA	0.5733	0.560055	0.565605
EnKF	FN	68.9241	62.6420	51.0139
EnKF	CA	0.6892	0.6264	0.5101

TABLE V
Parameter Setting in Simulation and Experiment Result

Parameters	N	n	MMSE	MT	OLT
EnKF	200	100	9.5317	2.4886	–
FPF	200	100	8.9391	152.2366	–
OTPF	200	100	8.2453	5.6911	3.2925
EnKF	100	100	10.4381	2.252	–
FPF	100	100	9.0786	74.1399	–
OTPF	100	100	8.3256	4.3674	2.4927
EnKF	50	100	10.6123	2.0838	–
FPF	50	100	9.2524	38.4147	–
OTPF	50	100	8.9160	3.4068	2.0474
EnKF	25	100	12.0417	1.8137	–
FPF	25	100	11.4513	14.7709	–
OTPF	25	100	9.7026	2.7506	1.8065
KBF	-	100	8.2782	–	–
EnKF	100	50	6.8058	0.9232	–
FPF	100	50	5.8454	30.6788	–
OTPF	100	50	5.6499	2.2912	1.3323
EnKF	50	50	6.8404	0.7844	–
FPF	50	50	6.1477	15.6919	–
OTPF	50	50	5.6509	1.7307	1.0061
EnKF	25	50	7.1740	0.7150	–
FPF	25	50	6.9573	8.0514	–
OTPF	25	50	6.4021	1.4407	0.8076
KBF	-	50	5.5776	–	–
EnKF	50	25	4.8392	0.3035	–
FPF	50	25	4.7400	2.8714	–
OTPF	50	25	4.5751	0.7162	0.5135
EnKF	25	25	5.6113	0.3299	–
FPF	25	25	5.0915	5.4705	–
OTPF	25	25	4.5605	0.9481	0.4470
EnKF	15	25	5.6429	0.2771	–
FPF	15	25	4.9941	1.8839	–
OTPF	15	25	4.7841	0.6488	0.3987
KBF	-	25	4.5782	–	–

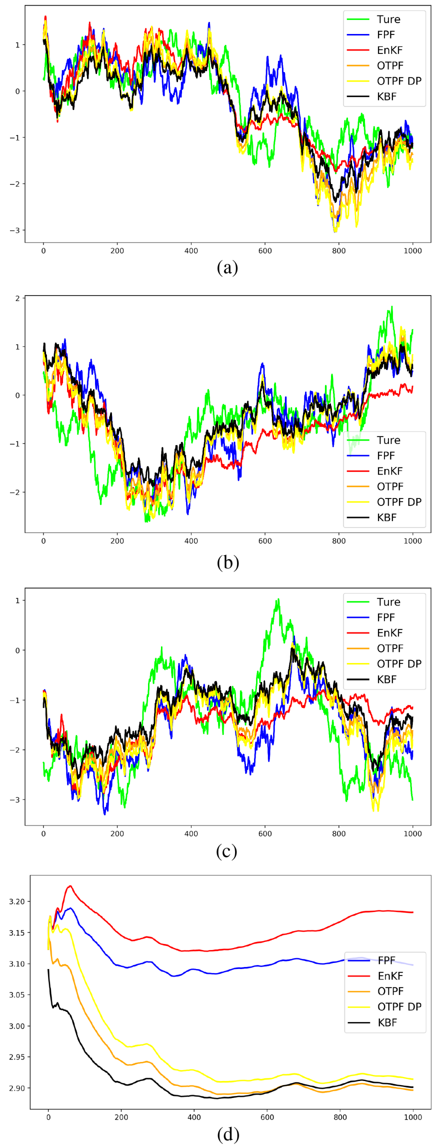

Fig. 1. Estimation results of algorithms in the 100-D case with 100 particles at different time steps $0 \leq k \leq 1000$. (a) First dimension. (b) 50th dimension. (c) 100th dimension. (d) Total RMSE.

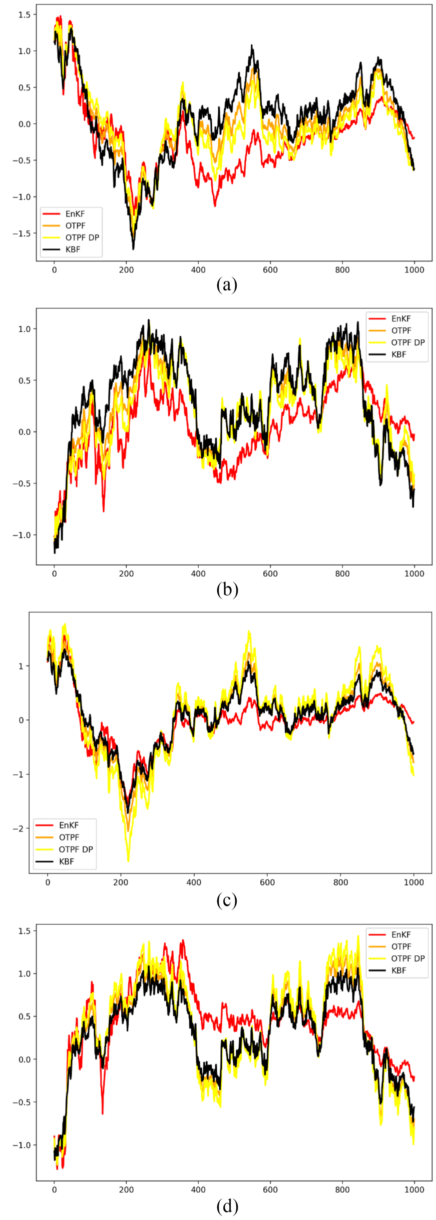

Fig. 2. Estimation results of algorithms in the 200-D case with 100 particles and 200 particles at different time steps $0 \leq k \leq 1000$. (a) First dimension with 100 particles. (b) 200th dimension with 100 particles. (c) First dimension with 200 particles. (d) 200th dimension with 200 particles.

set P_{40} as P_∞, and we calculate the FN of error between the ensemble variance and the optimal variance at time steps 100, 500, and 1000. The "Component average" (CA) is the mean average of the FN, i.e., $\text{CA} = \text{FN}/100$. The convergence of the ensemble variance will be effectively improved by shortening the time step.

Fig. 1 shows the comparison of the three algorithms in 100-D filtering, where the parameters are the same with the 100-D case in Table V.

Then, we compared the ensemble means of the two OTPFs based on the different approximation methods for P_t^{-1} through 200-D experiments. The ensemble means of the OTPFs, the EnKF, and the optimal result KBF are displayed in Fig. 2. The MMSEs of the 200-D case are given in Table VI, where N_1 and N_2 are particle numbers. It is easy to find that the OTPFs perform better than the traditional EnKF given the same number of particles, and the OTPF based on the dual system also has good accuracy. As the number of particles increases, the EnKF performs better, but its results are not as good as the OTPF with a low number of particles. In higher dimensional cases, we have reason to believe that the OTPF can use a smaller number of particles than the EnKF to obtain better numerical results.

TABLE VI
MMSE in 200 Dimensions

Algorithms	N_1	MMSE	N_2	MMSE
EnKF	100	14.54447	200	13.96709
OTPF	100	12.8897	200	12.7366
OTPF DP	100	13.2798	200	12.9912
KBF	–	12.3194	–	12.4593

VI. CONCLUSION

In this article, we reinterpret the evolution of the posterior distribution as the geometric evolution and give transfer maps by OT. Based on this idea, we propose a new effective OTPF for linear time-varying systems with correlated noise. Furthermore, we give the rigorous convergence analysis of our method as well as the FPF when the system is time invariant. In the simulation part, it is shown that our OTPF is especially efficient for high-dimensional problems compared with the FPF and the EnKF. More specifically, we only need 100 particles for 100-D problems in our OTPF. However, the new method is not limited to 100 dimensions and can solve higher dimensional numerical experiments with the aid of suitable computing equipment.

The experimental results show that the sparsity of the system in high-dimensional situations means that the system can be approximated by a small number of particles. The computational complexity will be reduced from $\mathcal{O}(n \times n)$ to $\mathcal{O}(N \times n)$, where N is much smaller than n.

However, all our works are limited in linear systems, and how to derive the efficient OTPF for general nonlinear systems is our future work.

APPENDIX

Proof of Lemma 2

In this appendix, we will give the proof of Lemma 2.
First of all, we need the following lemma.

LEMMA 6 (SEE [29]) The two Gaussian distribution a and b, where $a \sim \mathcal{N}(\mu_a, P_{aa})$ $b \sim \mathcal{N}(\mu_b, P_{bb})$ and

$$\begin{pmatrix} a \\ b \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}, \begin{pmatrix} P_{aa} & P_{ab} \\ P_{ba} & P_{bb} \end{pmatrix}\right).$$

Then, the conditional random variable $a|b$ is also Gaussian and $a|b \sim \mathcal{N}(\mu_{a|b}, P_{a|b})$

$$\mu_{a|b} = \mu_a + P_{ab}P_{bb}^{-1}(b - \mu_b) \quad (76)$$

$$P_{a|b} = P_{aa} - P_{ab}P_{bb}^{-1}P_{ba}. \quad (77)$$

Proof of Lemma 2: For the filtering system (2), we have

$$x_t = e^{\tilde{A}(t)}x_0 + \sigma_W \int_0^t e^{\tilde{A}(t-s)}dW_s + \sigma_B \int_0^t e^{\tilde{A}(t-s)}dB_s \quad (78)$$

where $\tilde{A}(t) := \int_0^t A(s)ds$. It can be easily concluded that x_t is Gaussian since x_0, $\{W_s\}$ and $\{B_s\}$ are independent and Gaussian.

Then, for the observation term, we have

$$\begin{aligned} y_t &= y_0 + \int_0^t H(t)x_t dt + \int_0^t dW_t \\ &= y_0 + A^{-1}(t)(I_n - e^{\tilde{A}(t)})x_0 + \\ &\quad + \sigma_W(t)A^{-1}(t)\int_0^t (I_n - e^{\tilde{A}(t-s)})dW_s \\ &\quad + \sigma_B(t)A^{-1}(t)\int_0^t (I_n - e^{\tilde{A}(t-s)})dB_s + W_t. \end{aligned} \quad (79)$$

Then, we can know that y_t is also Gaussian. Then, this lemma holds by Lemma 6. □

Acknowledgment

Professor Stephen Shing-Toung Yau is grateful to the National Center for Theoretical Sciences 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
J. Basic Eng., vol. 82, pp. 35–45, 1960.

[2] R. E. Kalman and R. S. Bucy
New results in linear filtering and prediction theory
J. Basic Eng., vol. 83, pp. 95–108, 1961.

[3] D. Crisan
Robust filtering: Correlated noise and multidimensional observation
Ann. Appl. Probab., vol. 23, no. 5, pp. 2139–2160, 2013.

[4] M. H. A. Davis
Pathwise nonlinear filtering with correlated noise
in *The Oxford Handbook of Nonlinear Filtering*. Oxford, U.K.: Oxford Univ. Press, 2011, pp. 403–424.

[5] J. M. C. Clark
The design of robust approximations to the stochastic differential equations of nonlinear filtering
Commun. Syst. Random Process Theory, vol. 25, pp. 721–734, 1978.

[6] M. H. A. Davis
On a multiplicative functional transformation arising in nonlinear filtering theory
Ztschrift Wahrschnlichktstheorie Verwandte Gebiete, vol. 54, no. 2, pp. 125–139, 1980.

[7] H. J. Kushner
A robust discrete state approximation to the optimal nonlinear filter for a diffusion
Stochastics, vol. 3, nos. 1–4, pp. 75–83, 1980.

[8] A. Bain and D. Crisan
Fundamentals of Stochastic Filtering, vol. 60. Berlin, Germany: Springer, 2008.

[9] N. J. Gordon, D. J. Salmond, and A. F. M. Smith
Novel approach to nonlinear/non-Gaussian Bayesian state estimation
Proc. Inst. Elect. Eng. F, Radar Signal Process., vol. 140, no. 2, pp. 107–113, 2002.

[10] 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.

[11] S. M. P. G. Yang, T. Laugesen R, and S. P. Meyn
Multivariable feedback particle filter
Automatica, vol. 71, pp. 10–23, 2016.

[12] A. K. Tilton, S. Ghiotto, and P. G. Mehta
A comparative study of nonlinear filtering techniques
in *Proc. IEEE 16th Int. Conf. Inf. Fusion*, 2013, pp. 1827–1834.

[13] K. Berntorp
Feedback particle filter: Application and evaluation
in *Proc. 18th Int. Conf. Inf. Fusion*, 2015, pp. 1633–1640.

[14] A. Taghvaei and P. G. Mehta
Error analysis for the linear feedback particle filter
in *Proc. IEEE Annu. Amer. Control Conf.*, 2018, pp. 4261–4266.

[15] X. Chen, X. Luo, J. Shi, and S. S.-T. Yau
General convergence result for continuous-discrete feedback particle filter

Int. J. Control, pp. 1–15, 2021, doi: 10.1080/00207179.2021.1948105.
[16] X. Luo and H. Miao
Feedback particle filter with correlated noises
in *Proc. IEEE 58th Conf. Decis. Control*, 2019, pp. 1637–1643.
[17] P. L. Houtekamer and H. L. Mitchell
A sequential ensemble Kalman filter for atmospheric data assimilation
Monthly Weather Rev., vol. 129, no. 1, pp. 123–137, 2001.
[18] J. Curn
Correlated estimation problems and the ensemble Kalman filter
Ph.D. dissertation, Dept. Comput. Sci., Univ. Dublin, Dublin, Ireland, 2014.
[19] N. K. Chada, A. Jasra, and F. Yu
Multilevel ensemble Kalman-Bucy filters
2020, *arXiv:2011.04342*.
[20] T. Lange and W. Stannat
On the continuous time limit of ensemble square root filters
Commun. Math. Sci., vol. 19, no. 7, pp. 1855–1880, 2021.
[21] A. Halder and T. T. Georgiou
Gradient flows in uncertainty propagation and filtering of linear Gaussian systems
in *Proc. IEEE 56th Annu. Conf. Decis. Control*, 2017, pp. 3081–3088.
[22] S. Reich
A nonparametric ensemble transform method for Bayesian inference
SIAM J. Sci. Comput., vol. 35, no. 4, pp. A2013–2024, 2013.
[23] E. M. Tarek and Y. M. Marzouk
Bayesian inference with optimal maps
J. Comput. Phys., vol. 231, no. 23, pp. 7815–850, 2012.
[24] D. Mesa, J. Tantiongloc, M. Mendoza, and T. P. Coleman
A distributed framework for the construction of transport maps
Neural Comput., vol. 31, no. 4, pp. 613–652, 2019.
[25] J. Heng
On the use of transport and optimal control methods for Monte Carlo simulation
Ph.D. dissertation, St Cross College, Univ. Oxford, Oxford, England, 2016.
[26] J. Mandel, L. Cobb, and J. D. Beezley
On the convergence of the ensemble Kalman filter
Appl. Math., vol. 56, no. 6, pp. 533–541, 2011.
[27] X.-L. Hu, T. Schon, and L. Ljung
A general convergence result for particle filtering
IEEE Trans. Signal Process., vol. 59, no. 7, pp. 3424–3429, Jul. 2011.
[28] 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.
[29] A. H. Jazwinski
Stochastic Processes and Filtering Theory. Chelmsford, MA, USA: Courier Corporation, 2007.
[30] C. Villani
Topics in Optimal Transportation, vol. 58. Providence, RI, USA: American Math. Soc., 2003.
[31] G. Peyré and M. Cuturi
Computational optimal transport: With applications to data science
Found. Trends Mach. Learn., vol. 11, no. 5/6, pp. 355–607, 2019.
[32] C. D. Meyer
Matrix Analysis and Applied Linear Algebra, vol. 71. Philadelphia, PA, USA: SIAM, 2000.
[33] T. Kaczorek
Linear Control Systems: Analysis of Multivariable Systems. Hoboken, NJ, USA: Wiley, 1992.
[34] I. Karatzas and S. Shreve
Brownian Motion and Stochastic Calculus. Berlin, Germany: Springer, 2012.
[35] X.-L. Hu, T. B. Schon, and L. Ljung
A basic convergence result for particle filtering
IEEE Trans. Signal Process., vol. 56, no. 4, pp. 1337–1348, Apr. 2008.

Jiayi Kang received the B.S. degree in mathematics 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 with the Department of Mathematical Sciences, Tsinghua University, Beijing, China, under the supervision of Prof. Stephen Shing-Toung Yau.

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

Xiuqiong Chen (Member, IEEE) received the B.S. degree in applied mathematics from the School of Mathematics and Systems Science, Beihang University, Beijing, China, in 2014, and the Ph.D. degree in applied mathematics from the Department of Mathematical Sciences, Tsinghua University, Beijing, in 2019.

She was a Postdoctoral Scholar with the Yau Mathematical Sciences Center, Tsinghua University, from 2019 to 2021, and was selected as the "Shuimu Tsinghua Scholar." She is currently an Assistant Professor with the School of Mathematics, Renmin University of China, Beijing. Her research interests include nonlinear filtering and deep learning.

Yangtianze Tao received the B.S. degree in mathematics 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 with the Department of Mathematical Sciences, Tsinghua University, Beijing, China, under the supervision of Prof. Stephen, Yau.

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

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

He was a Member of the Institute of Advanced Study, Princeton, NJ, USA, from 1976 to 1977 and 1981 to 1982. He was a Benjamin Pierce Assistant Professor with Harvard University, Cambridge, MA, USA, from 1977 to 1980. He then joined the Department of Mathematics, Statistics and Computer Science (MSCS), University of Illinois at Chicago (UIC), Chicago, IL, USA, and served for more than 30 years. From 2005 to 2011, he was 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 currently a Full Time Professor with the Department of Mathematical Sciences. His research interests include nonlinear filtering, bioinformatics, complex algebraic geometry, Cauchy–Riemann geometry, and singularities theory.

Dr. Yau has been the Managing Editor and Founder of *Journal of Algebraic Geometry* since 1991 and the Editor-in-Chief and Founder of *Communications in Information and Systems* since 2000. He was the General Chairman of the 1995 IEEE International Conference on Control and Information. He received the Sloan Fellowship in 1980, the Guggenheim Fellowship in 2000, and the American Mathematical Society Fellow Award in 2013. In 2005, he was entitled the UIC Distinguished Professor.