ASIAN J. MATH.
Vol. 29, No. 1, pp. 051–076, February 2025

003

CONTINUOUS DISCRETE OPTIMAL TRANSPORTATION PARTICLE FILTER*

XIUQIONG CHEN†, JIAYI KANG‡, AND STEPHEN S.-T. YAU§

Abstract. In this paper, we employ optimal transportation principles to devise an innovative particle filter designed for linear time-varying systems featuring continuous state dynamics and discrete observations. This novel approach involves the optimal transport of the posterior distribution of the state from one time instant to its subsequent instant. Moreover, we conduct a thorough analysis of the estimation errors, examining the discrepancies between the actual conditional mean and empirical mean, as well as between the actual conditional covariance and empirical covariance. The efficiency of the proposed algorithm is demonstrated through numerical experiments.

Key words. Linear system, optimal transportation, feedback particle filter, convergence.

Mathematics Subject Classification. 93C05, 93D20, 93D23, 34A12.

1. Introduction. We investigate the filtering problems characterized by continuous state and discrete observation. The challenges within these situations arise due to the dynamic and continuous nature of the underlying systems, where the state evolves continuously over time. Simultaneously, discrete observations are acquired, adding a layer of complexity in reconciling the continuous system evolution with discrete and noisy measurements. To illustrate, in domains like robotics and autonomous vehicles, continuous state variables, such as position and velocity, undergo continuous evolution, while discrete sensors like cameras provide intermittent observations. Similarly, in aerospace engineering, the continuous trajectory of an aircraft or spacecraft is tracked using discrete radar or satellite measurements [9]. In financial markets, the modeling of stock prices as continuous processes involves tracking them through discrete observations, such as daily closing prices.

The filtering problems with continuous state and discrete observation can be modeled through the following stochastic differential equation system:

$$dX_t = f(X_t, t)dt + g(X_t, t)dB_t, \tag{1.1}$$
$$Y_n = h(X_{t_n}, t_n) + W_n, \tag{1.2}$$

Here, $X_t \in \mathbb{R}^{d_1 \times 1}$ represents the state at time t, B_t is a d_1-dimensional Brownian motion process with $\mathbb{E}[dB_t dB_t^\top] = Q_t dt$ and independent of X_0. The distribution of the initial state X_0 is denoted as $\pi_{0|0}$. The observation, denoted as $Y_n \in \mathbb{R}^{d_2 \times 1}$, arrives at discrete time instances $t = t_n = n\Delta t$ ($\Delta t > 0$), and W_n represents white noise with $\mathbb{E}[W_n W_n^\top] = R_n$ and independent of X_t. The probability space under consideration is denoted as $(\mathbf{\Omega}, \mathscr{F}, \mathbb{P})$, and the σ-algebra $\mathscr{F}_t \triangleq \sigma(Y_n : t_n \leq t)$ represents the observation history up to time t.

*Received February 12, 2024; accepted for publication October 23, 2024. This work is supported by National Natural Science Foundation of China (NSFC) grant (12201631, 42450242) and Tsinghua University Education Foundation fund (042202008).

†School of Mathematics, Renmin University of China, Beijing 100872, P. R. China (cxq0828@ruc.edu.cn).

‡Co-first author. Beijing Institute of Mathematical Sciences and Applications (BIMSA), Beijing 101408, P. R. China (kangjiayi@bimsa.cn).

§Corresponding author. Beijing Institute of Mathematical Sciences and Applications (BIMSA), Beijing 101408, P. R. China; and Department of Mathematical Sciences, Tsinghua University, Beijing, 100084, P. R. China (yau@uic.edu).

The objective of filtering problems is to seek an "optimal" estimate of the current state X_t given the observation history $\mathscr{F}_t$. It is known that the optimal estimate, in the minimum mean square error sense, is the conditional expectation $\mathbb{E}[X_t|\mathscr{F}_t]$ [7]. Moreover, the filtering problem can be fully resolved upon obtaining the posterior distribution $\mathbb{P}(X_t|\mathscr{F}_t)$. In the case where the system described by equations (1.1)-(1.2) is linear and Gaussian, the optimal solution is provided by the Kalman filter (KF) [8, 7]. However, for general nonlinear systems, obtaining the optimal solution is often intractable, necessitating the use of approximation techniques. Examples include the Taylor expansion method employed in the extended Kalman filter [7], finite dimensional spectral method used in Yau-Yau algorithms [19, 10] and the Monte Carlo approximation utilized in the particle filter (PF) [6] and ensemble Kalman filter [5].

In various PFs, the objective is to approximate the posterior distribution $\mathbb{P}(X_t|\mathscr{F}_t)$ through an empirical distribution formed by particles, as outlined in [4]. The feedback particle filter (FPF) is a novel Monte Carlo method incorporating a control law designed to minimize the Kullback-Leibler (K-L) divergence between the actual distribution and the posterior distribution of particles [18, 17]. Consequently, the posterior distribution is effectively approximated by the empirical distribution. The extension of FPF to continuous discrete systems is detailed by Yang et al. in [16], complemented by an explicit convergence analysis provided by Chen et al. in [2]. It is essential to note, however, that the control law in FPF is non-unique. The optimal control law can be uniquely determined through the optimal transportation between the posterior distribution $\mathbb{P}(X_t|\mathscr{F}_t)$ in the continuous case [14], and the estimation error was analyzed in [15].

Motivated by the work [14], in this paper, we construct a novel PF for linear continuous discrete systems using optimal transportation, denoted as the optimal transportation particle filter (OTPF). Compared with the traditional PF [6] and FPF, we need less particles to achieve the same accuracy.

The contributions of this work can be summarized as follows.

- In linear case, we propose a novel PF that utilizes optimal transportation between the actual posteriors in the form of a time-stepping procedure. This innovative approach is formally presented and proven in Theorem 4.3.
- The explicit estimations of the L^p error between the optimal estimates m_t, P_t and their approximations $m_t^{(N)}, P_t^{(N)}$ by OTPF are provided. Additionally, it is demonstrated that the L^p error of OTPF follows an order of $\mathcal{O}(1/\sqrt{N})$ for any $p \geq 1$, where N denotes the number of particles. Notably, this error decays exponentially fast as time $t \to \infty$. This analytical result is presented in Theorem 5.4.

This paper is organized as follows. In Section 2, we extend the FPF method to cover a broader class of time-varying systems. Section 3 is dedicated to the introduction of FPF for linear Gaussian systems, with particular emphasis on the evolution during the updating step. The subsequent section, Section 4, introduces the OTPF for linear Gaussian systems. In Section 5, we undertake a detailed analysis of the estimation error within the OTPF framework. Moving on to Section 6, we assess the efficiency of OTPF through the examination of two numerical examples. The concluding remarks are presented in the final section.

Notations: For two positive numbers a and b, the asymptotic inequality $a \lesssim_{p,q} b$ means that $a \leq C_{p,q} b$, where $C_{p,q}$ is a positive finite constant depending on the values of p and q. The notation $\|\cdot\|$ represents the 2-norms of vectors or matrices.

Additionally, let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ refer to the minimal and maximal eigenvalues of matrix A, respectively. Define the logarithmic norm $\mu(A)$ for a square matrix A of dimensions $n \times n$ as follows:

$$\begin{aligned}\mu(A) &:= \inf\left\{\alpha : \forall x \in \mathbb{R}^{n\times 1}, x^{\mathrm{T}}Ax \leqslant \alpha \left\|x\right\|^2\right\}\\ &= \lambda_{\max}\left(\left(A + A^{\mathrm{T}}\right)/2\right).\end{aligned}$$

It can be proved that [11]

$$\mu(A) \geqslant \varsigma(A) := \max\{\mathrm{Re}(\lambda) : \lambda \in \mathrm{Spec}(A)\}, \tag{1.3}$$

where $\mathrm{Re}(\lambda)$ denotes the real part of the eigenvalues λ. For all $p \geq 1$, the L^p norm, denoted as $\|\circ\|_p$, is defined for random vectors and matrices as $\mathbb{E}^{1/p}[\|\circ\|^p]$, provided that $\mathbb{E}[\|\circ\|^p]$ is finite. For two real symmetric matrices A and B with dimensions $r\times r$, the notation $A \geq B$ indicates that $A - B$ is positive semidefinite. For any function $f\colon \mathbb{R}^n \to \mathbb{R}$, its gradient $\nabla f\colon \mathbb{R}^n \to \mathbb{R}^n$ is defined at the point x as the column vector

$$\nabla f(x) = \left[\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n}\right]^{\top},$$

if it is differentiable at x. For any function $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^n$, its divergence is given by

$$\nabla \cdot \mathbf{f} = \frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} + \cdots + \frac{\partial f_n}{\partial x_n}.$$

For any function $\mathbf{F}\colon \mathbb{R}^n \to \mathbb{R}^{n\times m}$, its divergence is defined as

$$\nabla \cdot \mathbf{F} = [\nabla \cdot \mathbf{F}_1, \cdots, \nabla \cdot \mathbf{F}_m],$$

where $\mathbf{F}_i$ is the i-th column of $\mathbf{F}$, $1 \leq i \leq m$.

2. Preliminary.

When constructing the OTPF for a continuous-discrete system, we aim for the conditional density function to evolve continuously. However, it evolves discretely during the Bayesian update step. Alternatively, FPF provides an equivalent continuous evolution equation for the update step. Therefore, we first introduce the FPF for general nonlinear time-varying systems, which can be straightforwardly obtained from [16]. These results will later be used in the construction of the OTPF.

2.1. Exact Feedback particle filter.

Let $\overline{X}_t$ represent the state of the i.i.d. particles used in FPF, with initial particles $\overline{X}_0 \sim p(X_0)$, the density of the initial actual state. The evolution of $\overline{X}_t$ can be divided into two iterative steps:

1. Prediction: The particles evolve according to (1.1) in the time interval $t \in [t_{n-1}, t_n)$:

$$d\overline{X}_t = f(\overline{X}_t, t)dt + g(\overline{X}_t, t)d\overline{B}_t, \tag{2.1}$$

   with initial value $\overline{X}_{t_{n-1}}$, and $\overline{B}_t$ is a independent copy of B_t. We denote the left limit as:

$$\overline{X}_{t_n^-} := \lim_{t \nearrow t_n} \overline{X}_t. \tag{2.2}$$

2. Updating: let $\overline{S}_n(0) := \overline{X}_{t_n^-}$, $\overline{S}_n(\lambda)$ evolves according to the following equation

$$\frac{d\overline{S}_n}{d\lambda}(\lambda) = \underbrace{\mathcal{K}_n(\overline{S}_n(\lambda), \lambda)Y_n + u_n(\overline{S}_n(\lambda), \lambda)}_{optimal\ U_n(\lambda)}, \tag{2.3}$$

with initial condition $\overline{S}_n(0)$, and the pseudo-time $\lambda \in [0, 1]$. **The control input $U_n(\lambda)$ (or $\{\mathcal{K}, u\}$) is *optimal* if it is designed such that the posterior distribution of $\overline{S}_n(1)$ equals to the actual posterior distribution**, i.e.,

$$\mathbb{P}(X_{t_n} \in A|\mathscr{F}_{t_n}) = \mathbb{P}(\overline{S}_n(1) \in A|\mathscr{F}_{t_n}).$$

The initial condition for the next interval is assigned as $\overline{X}_{t_n} = \overline{S}_n(1)$.

More specifically, for any measurable set $A \in \mathscr{F}$,

$$\int_A p^*(x,t)dx = \mathbb{P}(X_t \in A|\mathscr{F}_t), \tag{2.4}$$

where $p^*(x, 0)$ is the density function of the initial state X_0. For $n = 1, 2, \cdots$,

$$\int_A p(x,t)dx = \mathbb{P}(\overline{X}_t \in A|\mathscr{F}_t). \tag{2.5}$$

Given the initial density $p^*(x, 0)$ and the increasing filtration $\mathscr{F}_t$, the evolution of the posterior $p^*(x, t)$ is obtained by two alternative steps: prediction and updating, which is shown in the following proposition.

Proposition 2.1 (Proposition 4.2.1 in [16]). *Consider the filtering problem (1.1)-(1.2) over time interval $[t_{n-1}, t_n]$. For $t \in [t_{n-1}, t_n)$, $p^*(x, t)$ satisfies the following Fokker-Planck equation [7]:*

$$\frac{\partial p^*}{\partial t}(x,t) = -\sum_{i=1}^{d_1} \frac{\partial(p^* f_i)}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{d_1} \frac{\partial^2 \left[p^*(gQg^\top)_{ij}\right]}{\partial x_i \partial x_j}. \tag{2.6}$$

Then we have

$$p^*(x, t_n^-) := \lim_{t \nearrow t_n} p^*(x,t).$$

Note $p^*(x, t_n^-)$ is the apriori distribution of X_{t_n} given $\mathscr{F}_{t_{n-1}}$.

At the discrete time instant $t = t_n$ when the observation is made, the posterior density is updated using Bayes' rule:

$$p^*(x,t_n) = p^*(x,t_n^-) \cdot \exp\left[-\frac{1}{2}(Y_n - h(x,t_n))^\top R_n^{-1}(Y_n - h(x,t_n))\right] / C_n, \tag{2.7}$$

where C_n is the normalization constant.

The two equations (2.6)-(2.7) define the mapping of p_X from t_{n-1} to t_n.

Let us take a logarithm of both sides of (2.7):

$$\ln p^*(x,t_n) = \ln p^*(x,t_n^-) + h^\top(x,t_n)R_n^{-1}\left(Y_n - \frac{1}{2}h(x,t_n)\right) - \ln C_n', \tag{2.8}$$

where C'_n is a constant that does not depend on x, and this constant can be dropped to obtain the recursion for the unnormalized density $q^*(t,x)$:

$$\ln q^*(x,t_n) = \ln q^*(x,t_n^-) + h^\top(x,t_n)R_n^{-1}\left(Y_n - \frac{1}{2}h(x,t_n)\right), \tag{2.9}$$

where

$$p^*(x,t_n) = \frac{q^*(x,t_n)}{\int q^*(x',t_n)\,\mathrm{d}x'},\ p^*\left(x,t_n^-\right) = \frac{q^*(x,t_n^-)}{\int q^*\left(x',t_n^-\right)\mathrm{d}x'}.$$

Let us define two homotopy functions $\zeta_n(x,\lambda)$ and $\rho_n^*(x,\lambda)$ as follows:

$$\begin{aligned}\zeta_n(x,\lambda) &:= \ln q^*\left(x,t_n^-\right) + \lambda h^\top(x,t_n)R_n^{-1}\left(Y_n - \frac{1}{2}h(x,t_n)\right),\\ \rho_n^*(x,\lambda) &:= \frac{\exp\left(\zeta_n(x,\lambda)\right)}{\int \exp\left(\zeta_n\left(x',\lambda\right)\right)\mathrm{d}x'},\end{aligned} \tag{2.10}$$

where $\lambda \in [0,1]$ is the pseudo-time parameter.

By construction, it can be easily checked that, for $\lambda = 0$ and $\lambda = 1$:

$$\begin{aligned}\zeta_n(x,0) = \ln q^*\left(x,t_n^-\right), \quad \zeta_n(x,1) = \ln q^*\left(x,t_n\right)\\ \rho_n^*(x,0) = p^*\left(x,t_n^-\right), \quad \rho_n^*(x,1) = p^*\left(x,t_n\right).\end{aligned} \tag{2.11}$$

And the evolution of $\rho_n^*(x,\lambda)$ is described in the following proposition.

Proposition 2.2 (Proposition 2 [16]). *Consider the normalized density function $\rho_n^*(x,\lambda)$ as defined in* (2.10) *with $\lambda \in [0,1]$. Then its evolution is given by the following partial differential equation: For $\lambda \in [0,1]$*

$$\frac{\partial \rho_n^*}{\partial \lambda}(x,\lambda) = \rho_n^*(x,\lambda)\left[(h-\hat{h})^T R_n^{-1}Y_n - \frac{1}{2}h^\top R_n^{-1}h + \frac{1}{2}\left(\widehat{h^\top R_n^{-1}h}\right)\right], \tag{2.12}$$

where

$$\hat{h} := \int \rho_n^*(x,\lambda)h(x,t_n)dx,\ \left(\widehat{h^\top R_n^{-1}h}\right) := \int \rho_n^*(x,\lambda)h^\top(x,t_n)R_n^{-1}h(x,t_n)dx.$$

Let us denote $\rho_n(x,\lambda)$ the distribution of $S_n^i(\lambda)$ in (2.23). More specifically, we have

$$\rho_n(dx,0) := P(S_n^i(0) \in dx|\mathscr{F}_{t_{n-1}}), \tag{2.13}$$

$$\rho_n(dx,1) := P(S_n^i(1) \in dx|\mathscr{F}_{t_n}). \tag{2.14}$$

And the evolution equation for $\rho_n(x,\lambda)$ is given by the following Kolmogorov's forward equation [7]:

$$\frac{\partial \rho_n}{\partial \lambda}(x,\lambda) = -\nabla\cdot\left(\rho_n \mathcal{K}_n\right)Y_n - \nabla\cdot\left(\rho_n u_n\right). \tag{2.15}$$

Theorem 2.3. *For each fixed $\lambda \in [0,1]$, let η_j be the solution of:*

$$\begin{aligned}&\nabla\cdot\left(\rho_n \nabla \eta_j\right) = -(h-\hat{h})^\top (R_n^{-1})_j \rho_n,\\ &\int_{R^{d_1}} \eta_j(x,\lambda)\rho_n(x,\lambda)dx = 0,\end{aligned} \tag{2.16}$$

for $j = 1, \cdots, d_2$*, where* $(R_n^{-1})_j$ *is the* j*-th column of matrix* R_n^{-1}*. Then the optimal*

$$\mathcal{K}_n = [\nabla\eta_1, \nabla\eta_2, \cdots, \nabla\eta_{d_2}], \tag{2.17}$$

and the optimal u_n *is obtained as*

$$u_n(x,\lambda) = -\frac{1}{2}\mathcal{K}_n(x,\lambda)(h(x,t_n)+\hat{h}) + \frac{1}{2}\Omega_n(x,\lambda), \tag{2.18}$$

where $\Omega = \nabla\varphi$*,* φ *is a scalar function, and it is a solution to*

$$\begin{aligned} &\nabla\cdot(\rho_n\nabla\varphi) = (\bar{\xi}-\xi)\rho_n, \\ &\int_{R^{d_1}} \varphi(x,\lambda)\rho_n(x,\lambda)dx = 0, \end{aligned} \tag{2.19}$$

where $\xi := \sum_{j=1}^{d_2}(\nabla\eta_j)^\top\nabla h_j$ *and* $\bar{\xi} := \hat{h}^\top R^{-1}\hat{h} - \widehat{h^\top R^{-1}h}$*.*

The proof of Theorem 2.3 can be found in Appendix A.1. The connection between the actual density function $p^*(x,t)$ of X_t and the posterior density $p(x,t)$ of $\overline{X}_t$ is illustrated in Fig. 1. They share the same initial values and evolution equation during the prediction step. Therefore, we only need to ensure that they follow the same evolution equation during the updating step. Specifically, (2.12) and (2.15) must be identical. This is precisely what is demonstrated in Theorem 2.3.

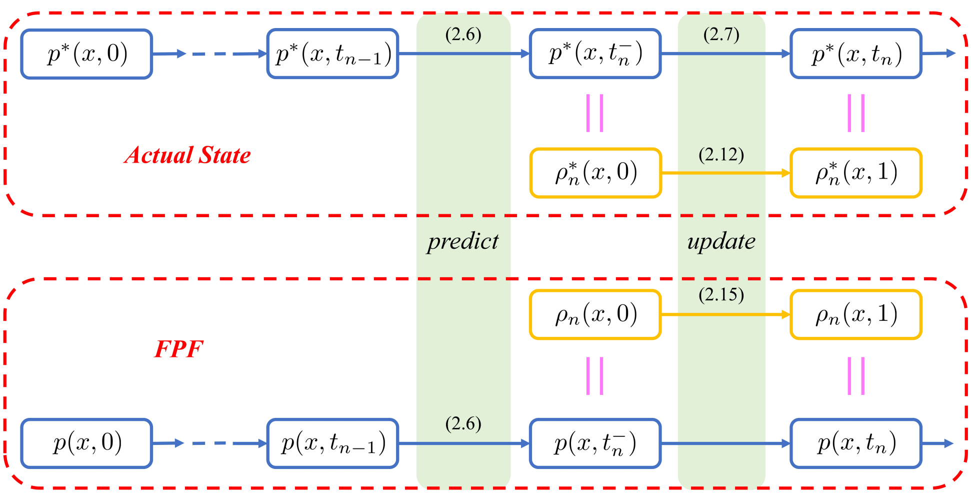

FIG. 1. *The connection between* $p^*(x,t)$*,* $\rho_n^*(x,\lambda), \rho_n(x,\lambda)$ *and* $p(x,t)$*.*

Substituting the optimal $\{\mathcal{K}, u\}$ in (2.17) and (2.18) into (2.3), we have

$$\frac{d\overline{S}_n}{d\lambda}(\lambda) = \mathcal{K}(\overline{S}_n(\lambda),\lambda)\left[Y_n - \frac{h(\overline{S}_n(\lambda))+\hat{h}}{2}\right] + \frac{1}{2}\Omega(\overline{S}_n(\lambda),\lambda). \tag{2.20}$$

We can now conclude that if $\overline{X}_0 \sim p^*(x,0)$ and the particles $\overline{X}_t$ evolve according to (2.1)-(2.3), with (2.3) having the explicit form (2.20), then $p(x,t) = p^*(x,t)$ for all $t \geq 0$.

2.2. Practical FPF. In practical scenarios, the exact solutions $\mathcal{K}_n$ and u_n in (2.3), obtained by solving (2.16) and (2.19), are often unattainable. However, employing numerical techniques allows us to obtain approximations, denoted as $\widetilde{\mathcal{K}}n$ and $\tilde{u}_n$ respectively. Consequently, the evolution equations for particles $\{X_t^i\}_{i=1}^N$ are given as follows, with N representing the number of particles.

1. Prediction: Given N particles $X_{t_{n-1}}^i \in \mathbb{R}^{d_1}$, $i = 1, 2, \cdots, N$ (sampled i.i.d. from $p_X(x, 0)$ at time $t = 0$), these particles evolve according to (1.1) in the time interval $t \in [t_{n-1}, t_n)$:

$$dX_t^i = f(X_t^i, t)dt + g(X_t^i, t)dB_t^i, \tag{2.21}$$

with an initial value of $X_{t_{n-1}}^i$. Here, $X_t^i \in \mathbb{R}^{d_1}$ denotes the state for the i-th particle at time t, and $\{B_t^i\}$ are mutually independent copies of B_t. The left limit is denoted as:

$$X_{t_n^-}^i := \lim_{t \nearrow t_n} X_t^i. \tag{2.22}$$

2. Updating: Define $S_n^i(0) := X_{t_n^-}^i$ for $i = 1, \cdots, N$. Each $S_n^i(\lambda)$ evolves according to the following equation:

$$\frac{dS_n^i}{d\lambda}(\lambda) = \widetilde{\mathcal{K}}_n(S_n^i(\lambda), \lambda)Y_n + \tilde{u}_n(S_n^i(\lambda), \lambda), \tag{2.23}$$

with an initial condition of $S_n^i(0)$ for $i = 1, 2, \cdots, N$, and the pseudo-time $\lambda \in [0, 1]$. The initial condition for the next interval is assigned as $X_{t_n}^i = S_n^i(1)$ for $i = 1, 2, \cdots, N$.

3. Feedback particle filter for liner system. In the subsequent sections of this paper, our focus shifts to the time-varying linear system:

$$dX_t = F_t X_t dt + G_t dB_t, \tag{3.1}$$

$$Y_n = H_n X_{t_n} + W_n, \tag{3.2}$$

which is a special case of system (1.1)-(1.2). Additionally, we introduce the assumption that the initial state X_0 follows a normal distribution $\mathcal{N}(m_0, P_0)$ and is independent of both the state noise $\{B_t\}$ and the observation noise $\{W_n\}$. Define

$$S_n := H_n^\top R_n^{-1} H_n. \tag{3.3}$$

We make the assumption that the 2-norms of S_n, G_t, and Q_t are uniformly bounded.

3.1. Linear FPF. For the linear system (3.1)-(3.2), we can derive the explicit form of the optimal control input in the evolution equation (2.3) for the particles of the FPF during the update step. This is presented in the following proposition:

PROPOSITION 3.1. *Consider the d_1-dimensional linear system* (3.1)-(3.2). *Let us assume that the homotopy density function ρ follows a Gaussian distribution, expressed as:*

$$\rho_n(x, \lambda) = \frac{1}{(2\pi)^{d_1/2} |\Sigma_{n,\lambda}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x - v_{n,\lambda})^\top \Sigma_{n,\lambda}^{-1}(x - v_{n,\lambda})\right], \tag{3.4}$$

where $\upsilon_{n,\lambda}$ represents the mean, $\Sigma_{n,\lambda}$ denotes the covariance matrix, and $|\Sigma_{n,\lambda}| > 0$ stands for the determinant. A solution to the boundary value problem defined by equations (2.17) *and* (2.19) *is given by:*

$$\begin{aligned}&\eta_j(x,\lambda) = (x-\upsilon_{n,\lambda})^\top \Sigma_{n,\lambda} H^\top (R_n^{-1})_j, \quad j=1,\dots,d_2\\ &\Omega(x,\lambda) = (0,\dots,0),\end{aligned}$$

where $(R_n^{-1})_j$ represents the j-th column of the matrix R_n^{-1}. Using $\mathcal{K}_n = [\nabla\eta_1,\dots,\nabla\eta_{d_2}]$, we obtain that $\mathcal{K}_n(x,\lambda) = \Sigma_{n,\lambda} H^\top R_n^{-1}$.

We can directly verify that the solutions in Proposition 3.1 satisfy the equations (2.16)-(2.19) and the proof is omitted.

In this linear Gaussian model, the gain function results in the formulation of a closed-form exact feedback particle filter as follows:

$$\begin{aligned}&t\in[t_{n-1},t_n): \mathrm{d}\overline{X}_t = F_t\overline{X}_t\ \mathrm{d}t + G_t\mathrm{d}\overline{B}_t,\ \overline{S}_n(0) = \overline{X}_{t_n^-},\\ &\qquad t=t_n: \frac{\mathrm{d}\overline{S}_n}{\mathrm{d}\lambda}(\lambda) = \Sigma_{n,\lambda}H_n^\top R_n^{-1}\left[Y_n - \frac{1}{2}H_n\left(\overline{S}_n(\lambda)+\upsilon_{n,\lambda}\right)\right],\ X_{t_n} = \overline{S}_n(1).\end{aligned} \tag{3.5}$$

Define the conditional mean m and covariance matrix P as follows:

$$\begin{aligned}m_t &:= \mathbb{E}[X_t|\mathscr{F}_t], \quad m_{t_n^-} := \mathbb{E}[X_{t_n}|\mathscr{F}_{t_{n-1}}],\\ P_t &:= \mathbb{E}\left[(X_t-m_t)(X_t-m_t)^\top\big|\mathscr{F}_t\right],\\ P_{t_n^-} &:= \mathbb{E}\left[(X_{t_n}-m_{t_n^-})(X_{t_n}-m_{t_n^-})^\top\Big|\mathscr{F}_{t_{n-1}}\right]\end{aligned} \tag{3.6}$$

THEOREM 3.2. *Consider the linear Gaussian filtering system* (3.1)-(3.2) *and the exact FPF* (3.5). *In this case the posterior distributions of X_t and $\overline{X}_t$ are same, which are Gaussian with conditional mean m_t and covariance matrix P_t given by the following equations of evolution. Between observations, these satisfy the differential equations:*

$$t\in[t_{n-1},t_n): \quad \begin{aligned}\mathrm{d}m_t &= F_t m_t\ \mathrm{d}t,\\ \frac{\mathrm{d}P_t}{\mathrm{d}t} &= F_tP_t + P_tF_t^\top + G_tQ_tG_t^\top,\end{aligned} \tag{3.7}$$

At discrete time instants $t=t_n$, these satisfy the iterative equations:

$$t=t_n: \begin{aligned}m_{t_n} &= m_{t_n^-} + \mathrm{K}_n\left(Y_n - H_n m_{t_n^-}\right),\\ P_{t_n} &= P_{t_n^-} - \mathrm{K}_nH_nP_{t_n^-},\end{aligned} \tag{3.8}$$

where the gain function $\mathrm{K}_n := P_{t_n^-}H_n^\top\left[H_nP_{t_n^-}H_n^\top + R_n\right]^{-1}$.

Proof. According to (3.1)-(3.2), the posterior distributions of X_t are Gaussian with first two moments evolving according to (3.7)-(3.8), which is exactly the KF [7].

For exact linear FPF, (3.7) can be obtained from the first equation in (3.5). Now we only need to consider the updating step (3.8). Equivalently, we only need to prove

$$m_{t_n} = \mathbb{E}[\overline{X}_{t_n}|\mathscr{F}_{t_n}], P_{t_n} = \mathrm{Cov}[\overline{X}_{t_n}|\mathscr{F}_{t_n}],\ n=0,1,2,\cdots. \tag{3.9}$$

At the initial instant $t_0 = 0$, (3.9) holds since $p(X_0) = p(\overline{X}_0)$.

By Proposition 3.1, it is known that the control input in the evolution equation (3.5) of $\overline{S}_n$ is optimal, by Theorem 2.3, i.e.,

$$\begin{aligned}\mathbb{P}(X_{t_n^-} \in A|\mathscr{F}_{t_{n-1}}) &= \mathbb{P}(\overline{S}_n(0) \in A|\mathscr{F}_{t_{n-1}}),\\ \mathbb{P}(X_{t_n} \in A|\mathscr{F}_{t_n}) &= \mathbb{P}(\overline{S}_n(1) \in A|\mathscr{F}_{t_n}),\end{aligned} \tag{3.10}$$

which means

$$\begin{aligned}\mathcal{N}(x; m_{t_n^-}, P_{t_n^-}) &= \rho_n(x, 0) = \mathbb{P}(\overline{X}_{t_n} \in A|\mathscr{F}_{t_{n-1}}),\\ \mathcal{N}(x; m_{t_n}, P_{t_n}) &= \rho_n(x, 1) = \mathbb{P}(\overline{X}_{t_n} \in A|\mathscr{F}_{t_n}),\end{aligned} \tag{3.11}$$

i.e.,

$$\begin{aligned}m_{t_n^-} &= \upsilon_{n,0} = \mathbb{E}[\overline{X}_{t_n}|\mathscr{F}_{t_{n-1}}],\ P_{t_n^-} = \Sigma_{n,0} = \mathrm{Cov}[\overline{X}_{t_n}|\mathscr{F}_{t_{n-1}}],\\ m_{t_n} &= \upsilon_{n,1} = \mathbb{E}[\overline{X}_{t_n}|\mathscr{F}_{t_n}],\ P_{t_n} = \Sigma_{n,1} = \mathrm{Cov}[\overline{X}_{t_n}|\mathscr{F}_{t_n}].\end{aligned} \tag{3.12}$$

□

Since the density $\rho_n(x, \lambda)$ is Gaussian, its evolution is equivalent to the evolution of its first two moments, $\upsilon_{n,\lambda}$ and $\Sigma_{n,\lambda}$. This result is provided in Lemma 3.3.

LEMMA 3.3. *The evolution equations for* $\upsilon_{n,\lambda}$ *and* $\Sigma_{n,\lambda}$ *in* (3.4) *are given by*

$$\begin{aligned}\frac{\partial \upsilon_{n,\lambda}}{\partial \lambda} &= \Sigma_{n,\lambda} H_n^\top R_n^{-1} \left[Y_n - H_n \upsilon_{n,\lambda}\right],\\ \frac{\partial \Sigma_{n,\lambda}}{\partial \lambda} &= -\Sigma_{n,\lambda} H_n^\top R_n^{-1} H_n \Sigma_{n,\lambda}.\end{aligned} \tag{3.13}$$

The proof can be found in Appendix A.2.

Combing (3.12) and (3.8), we have the following result.

COROLLARY 3.4. *The solutions of ODEs* (3.13) *at* $\lambda = 1$ *with initial values* $\upsilon_{n,0}$, $\Sigma_{n,0}$ *are*

$$\begin{aligned}\upsilon_{n,1} &= \upsilon_{n,0} + \mathrm{K}_n \left(Y_n - H_n \upsilon_{n,0}\right),\\ \Sigma_{n,1} &= \Sigma_{n,0} - \mathrm{K}_n H_n \Sigma_{n,0},\end{aligned} \tag{3.14}$$

with $\mathrm{K}_n := \Sigma_{n,0} H_n^\top \left[H_n \Sigma_{n,0} H_n^\top + R_n\right]^{-1}$.

4. Optimal transportation particle filter for linear system. The optimal control law u, K in FPF (2.3) is not unique, as indicated in [14]. To establish a unique control law, one approach is to formulate the filtering problem utilizing optimal transportation. This methodology enables the optimal transport of particles originating from the initial distribution $p(X_0)$ to particles corresponding to the posterior $p(X_t|\mathscr{F}_t)$. Subsequently, we will provide a concise overview of optimal transportation.

4.1. Optimal transportation. Consider two probability measures μ_X and μ_Y defined on $\mathbb{R}^n$, both possessing finite second moments. The Monge optimal transportation problem with quadratic cost seeks to minimize the expected squared norm of the displacement vector:

$$\min_{\mathcal{T}} \mathbb{E}\left[\|\mathcal{T}(X) - X\|^2\right], \tag{4.1}$$

where the minimization is performed over all measurable maps $\mathcal{T}:\mathbb{R}^n\to\mathbb{R}^n$ satisfying the condition $X\sim\mu_X$ and $\mathcal{T}(X)\sim\mu_Y$. The resulting minimizer $\mathcal{T}^*$ is termed the optimal transport map between μ_X and μ_Y, provided it exists.

In most cases, obtaining the explicit form of $\mathcal{T}$ is challenging. However, given that the posterior distributions of the considered filtering system are Gaussian, our attention can be directed towards the optimal transportation problem between Gaussian distributions. We denote a Gaussian distribution with mean m and covariance P as $\mathcal{N}(m,P)$. The solution to the optimal transportation problem between two Gaussian distributions is outlined in the following Theorem 4.1.

THEOREM 4.1 (Remark 2.31 in [12]). *If $\alpha=\mathcal{N}(m_\alpha,P_\alpha)$ and $\beta=\mathcal{N}(m_\beta,P_\beta)$ are two Gaussians in $\mathbb{R}^n$ with $P_\alpha,\ P_\beta\succ 0$, then one can show that the following map*

$$\mathcal{T}^*: x\to m_\beta+V(x-m_\alpha) \tag{4.2}$$

is the optimal transportation with cost function

$$c(\alpha,\beta):=\|m_\alpha-m_\beta\|_2^2+\|P_\alpha^{\frac{1}{2}}-P_\beta^{\frac{1}{2}}\|_F^2, \tag{4.3}$$

where

$$V=P_\alpha^{-\frac{1}{2}}(P_\alpha^{\frac{1}{2}}P_\beta P_\alpha^{\frac{1}{2}})^{\frac{1}{2}}P_\alpha^{-\frac{1}{2}}. \tag{4.4}$$

4.2. Exact OTPF. Our objective is to construct a particle process $\widetilde{X}_t$ such that $\widetilde{X}_{t_{n-1}}\sim p^*(x,t_{n-1})$ and $\widetilde{X}_{t_n^-}\sim p^*(x,t_n^-)$, $\forall\ n\geq 1$, and establish the optimal transportation from $p^*(x,t_{n-1})$ to $p^*(x,t_{n^-})$, and subsequently to $p^*(x,t_n)$, denoted as follows:

$$\widetilde{X}_{t_{n-1}}\overset{\mathcal{T}^*_{t_{n-1}}}{\longrightarrow}\widetilde{X}_{t_n^-}\overset{\mathcal{T}^*_{t_n^-}}{\longrightarrow}\widetilde{X}_{t_n}. \tag{4.5}$$

Now, we need to study the explicit forms of the optimal transport maps $\mathcal{T}^*_{t_{n-1}}$ from $p^*(x,t_{n-1})$ to $p^*(x,t_n^-)$ and $\mathcal{T}^*_{t_n^-}$ from $p^*(x,t_n^-)$ to $p^*(x,t_n)$.

Before proceeding, we first need to establish a technical lemma.

LEMMA 4.2. *Let P_t be the solution of* (3.7) *with $t\in[t_{n-1},t_n)$, then we have the following relationship:*

$$P_t^{-\frac{1}{2}}\left(P_t^{\frac{1}{2}}P_{t+\Delta t}P_t^{\frac{1}{2}}\right)^{\frac{1}{2}}P_t^{-\frac{1}{2}}=I+\Xi_t\Delta t+O(\Delta t^2), \tag{4.6}$$

where Ξ_t is the solution to the matrix equation

$$\Xi_tP_t+P_t\Xi_t=F_tP_t+P_tF_t^\top+G_tQ_tG_t^\top. \tag{4.7}$$

Let $\Sigma_{n,\lambda}$ be the covariance in (3.4) *with $\lambda\in[0,1]$, then we have the following relationship:*

$$\Sigma_{n,\lambda}^{-\frac{1}{2}}\left(\Sigma_{n,\lambda}^{\frac{1}{2}}\Sigma_{\lambda+\Delta\lambda}\Sigma_{n,\lambda}^{\frac{1}{2}}\right)^{\frac{1}{2}}\Sigma_{n,\lambda}^{-\frac{1}{2}}=I+\widetilde{\Xi}_\lambda\Delta\lambda+O(\Delta\lambda^2), \tag{4.8}$$

where $\widetilde{\Xi}_\lambda$ is the solution to the matrix equation

$$\widetilde{\Xi}_\lambda\Sigma_{n,\lambda}+\Sigma_{n,\lambda}\widetilde{\Xi}_\lambda=-\Sigma_{n,\lambda}H_n^\top R_n^{-1}H_n\Sigma_{n,\lambda}. \tag{4.9}$$

Proof. The proof of (4.7) is similar to that of Proposition 3 in [14]. (4.9) comes from the second equation of (3.13) and the proof of (4.7). ☐

We are now prepared to present the optimal transportation between the posterior density functions $p^*(x,t)$, which form the foundation of the OTPF.

THEOREM 4.3. *Consider the linear Gaussian problem. Let* $\widetilde{X}_0 \sim \mathcal{N}(m_0, P_0)$. *The optimal transportation on* $[t_{n-1}, t_n]$ *is given by the following two steps:*

- *for* $t \in [t_{n-1}, t_n)$*:*

$$d\widetilde{X}_t = F_t m_t dt + \Xi_t \left(\widetilde{X}_t - m_t\right) dt, \tag{4.10}$$

where Ξ_t *is the solution to* (4.7). *The right limit is denoted as*

$$\widetilde{X}_{t_n^-} = \lim_{t \nearrow t_n} \widetilde{X}_t. \tag{4.11}$$

- *for* $t = t_n$*: Let* $\widetilde{S}_n(0) = \widetilde{X}_{t_n^-}$*, then*

$$d\widetilde{S}_n(\lambda) = \Sigma_{n,\lambda} H_n^\top R_n^{-1} \left[Y_n - H_n \upsilon_{n,\lambda}\right] d\lambda + \widetilde{\Xi}_\lambda \left(\widetilde{S}_n(\lambda) - \upsilon_{n,\lambda}\right) d\lambda, \tag{4.12}$$

where $\widetilde{\Xi}$ *is the solution to* (4.9). *Then* $\widetilde{X}_{t_n} = S_n(1)$.

Then we have

$$\widetilde{X}_t \sim \mathcal{N}(m_t, P_t),\ \widetilde{X}_{t_n^-} \sim \mathcal{N}(m_{t_n^-},\ P_{t_n^-}),\ \forall\ t \geq 0,\ n \geq 1.$$

Proof. **Step 1:** We design an optimal transportation from $\mathcal{N}(m_{t_{n-1}}, P_{t_{n-1}})$ to $\mathcal{N}(m_{t_n^-}, P_{t_n^-})$.

We divide the time interval $[t_{n-1}, t_n)$ into N equal parts, i.e., $t_{n-1} = \tau_0 < \tau_1 < \cdots < \tau_N = t_n$ with $\tau_k = t_{n-1} + k\Delta t$, $\Delta t = \dfrac{t_n - t_{n-1}}{N}$ and $0 \leq k \leq N$. Now we derive the optimal transportation $\mathcal{T}_{\tau_k}$ between two Gaussian distributions $\mathcal{N}(m_{\tau_k}, P_{\tau_k})$ and $\mathcal{N}(m_{\tau_{k+1}}, P_{\tau_{k+1}})$. By Theorem 4.1 and Lemma 4.2, it is known that $\mathcal{T}_{\tau_k}$ is

$$\widetilde{X}_{\tau_{k+1}} = m_{\tau_{k+1}} + \left(\widetilde{X}_{\tau_k} - m_{\tau_k}\right) + \Xi_{\tau_k}(\widetilde{X}_{\tau_k} - m_{\tau_k})\Delta t + O(\Delta t^2). \tag{4.13}$$

Let $N \to \infty$, we have

$$d\widetilde{X}_t = dm_t + \Xi_t \left(\widetilde{X}_t - m_t\right) dt, \tag{4.14}$$

where dm_t is given by (3.7).

Step 2: We design an optimal transportation from $\mathcal{N}(m_{t_n^-}, P_{t_n^-})$ to $\mathcal{N}(m_{t_n}, P_{t_n})$.

By (3.12), we know that

$$\mathcal{N}(m_{t_n^-}, P_{t_n^-}) = \mathcal{N}(\upsilon_{n,0}, \Sigma_{n,0}), \mathcal{N}(m_{t_n^-},\ P_{t_n}) = \mathcal{N}(\upsilon_{n,1},\ \Sigma_{n,1}).$$

Therefore, we only need to design an optimal transportation from $\mathcal{N}(\upsilon_{n,0}, \Sigma_{n,0})$ to $\mathcal{N}(\upsilon_{n,1}, \Sigma_{n,1})$ in order to design an optimal transportation from $\mathcal{N}(m_{t_n^-}, P_{t_n^-})$ to $\mathcal{N}(m_{t_n}, P_{t_n})$.

Similar to step 1 and using the first equation of (3.13), we have

$$\begin{aligned} d\widetilde{S}_n(\lambda) =& d\upsilon_{n,\lambda} + \widetilde{\Xi}_t \left(\widetilde{S}_n(\lambda) - \upsilon_{n,\lambda}\right) d\lambda \\ =& \Sigma_{n,\lambda} H_n^\top R_n^{-1} \left[Y_n - H_n \upsilon_{n,\lambda}\right] d\lambda + \widetilde{\Xi}_\lambda \left(\widetilde{S}_n(\lambda) - \upsilon_{n,\lambda}\right) d\lambda, \end{aligned} \tag{4.15}$$

where $\widetilde{\Xi}_\lambda$ is the solution to (4.9). ☐

4.3. Practical OTPF. In practical applications, the exact conditional means and covariances in (4.10) and (4.12) are unattainable, and we resort to approximations using empirical means and covariances. Thus, we implement the following practical OTPF, represented by particles denoted as $\{\widetilde{X}_t^i\}_{i=1}^N$, where N signifies the number of particles.

1. Prediction: Given N particles $\widetilde{X}_{t_{n-1}}^i \in \mathbb{R}^{d_1}$, $i = 1, 2, \cdots, N$ (sampled i.i.d. from $p_X(x, 0)$ at time $t = 0$), the particles evolve during $t \in [t_{n-1}, t_n)$ according to:

$$d\widetilde{X}_t^i = F_t m_t^{(N)} dt + \Xi_t^{(N)} \left(\widetilde{X}_t^i - m_t^{(N)}\right) dt, \tag{4.16}$$

with initial value $\widetilde{X}_{t_{n-1}}^i$, where $m_t^{(N)}$ denotes the empirical mean calculated by

$$m_t^{(N)} := \frac{1}{N} \sum_{i=1}^N \widetilde{X}_t^i, \tag{4.17}$$

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

$$\Xi_t^{(N)} P_t^{(N)} + P_t^{(N)} \Xi_t^{(N)} = F_t P_t^{(N)} + P_t^{(N)} F_t^\top + G_t Q_t G_t^\top, \tag{4.18}$$

with the empirical covariance given by

$$P_t^{(N)} := \frac{1}{N-1} \sum_{i=1}^N \left(\widetilde{X}_t^i - m_t^{(N)}\right) \left(\widetilde{X}_t^i - m_t^{(N)}\right)^\top. \tag{4.19}$$

The left limit is denoted as:

$$\widetilde{X}_{t_n^-}^i := \lim_{t \nearrow t_n} \widetilde{X}_t^i. \tag{4.20}$$

2. Updating: Let $\widetilde{S}_n^i(0) := \widetilde{X}_{t_n^-}^i, i = 1, \cdots, N$, for $\lambda \in [0, 1]$, $S_n^i(\lambda)$ evolves according to the following equation

$$d\widetilde{S}_n^i(\lambda) = \Sigma_{n,\lambda}^{(N)} H_n^\top R_n^{-1} \left[Y_n - H_n \upsilon_{n,\lambda}^{(N)}\right] d\lambda + \widetilde{\Xi}_\lambda^{(N)} \left(\widetilde{S}_n^i(\lambda) - \upsilon_{n,\lambda}^{(N)}\right) d\lambda, \tag{4.21}$$

where $\upsilon_{n,\lambda}^{(N)}$ is the empirical mean computed by

$$\upsilon_{n,\lambda}^{(N)} := \frac{1}{N} \sum_{i=1}^N S_n^i(\lambda), \tag{4.22}$$

and $\widetilde{\Xi}_\lambda^{(N)}$ is the solution to

$$\widetilde{\Xi}_\lambda^{(N)} \Sigma_{n,\lambda}^{(N)} + \Sigma_{n,\lambda}^{(N)} \widetilde{\Xi}_\lambda^{(N)} = -\Sigma_{n,\lambda}^{(N)} H_n^\top R_n^{-1} H_n \Sigma_{n,\lambda}^{(N)}, \tag{4.23}$$

with the empirical covariance computed by

$$\Sigma_{n,\lambda}^{(N)} := \frac{1}{N-1} \sum_{i=1}^N \left(S_n^i(\lambda) - \upsilon_{n,\lambda}^{(N)}\right) \left(S_n^i(\lambda) - \upsilon_{n,\lambda}^{(N)}\right)^\top. \tag{4.24}$$

The initial condition for the next interval is assigned as $X_{t_n}^i = S_n^i(1)$ for $i = 1, 2, \cdots, N$.

5. Error analysis. In this section, we shall undertake an analysis of the discrepancies between the actual conditional mean and covariance m_t, P_t and their corresponding numerical approximations $m_t^{(N)}, P_t^{(N)}$ obtained through the OTPF approach.

Before conducting the error analysis, we need to make some assumptions about the system.

ASSUMPTION 1. *F_t in (3.1) satisfies $\sup_{t\geq 0}\mu(F_t) < -\varrho$, where ϱ is a positive constant.*

Based on this assumption, it is known that F_t is Hurwitz uniformly with respect to time. Therefore, this assumption ensures that the linear system (3.1) is stable.

ASSUMPTION 2. *S_n defined in (3.3) is a scalar matrix, i.e.,*

$$S_n = \rho(S_n)I, \text{ for some scalar } \rho(S_n) \geq 0, \tag{5.1}$$

where I is an $d_1 \times d_1$-dimensional identity matrix.

The state transition matrix associated with a smooth flow of any $(r \times r)$-matrix $U : \tau \mapsto U_\tau$ is denoted by $\mathcal{E}_{s,t}(U)$ such that for any $s \leq t$,

$$\frac{\partial}{\partial t}\mathcal{E}_{s,t}(U) = U_t\mathcal{E}_{s,t}(U) \quad \text{and} \quad \mathcal{E}_{t,s}(U) := \mathcal{E}_{s,t}(U)^{-1},$$

with $\mathcal{E}_{s,s} = \mathrm{I}$, the identity matrix.

LEMMA 5.1. *Define $\Phi_{s,t} := \mathcal{E}_{s,t}(F)$, $\Psi_{n,s,t} := \mathcal{E}_{s,t}(-\Sigma_n S_n)$, $\Psi_{n,s,t}^{(N)} := \mathcal{E}_{s,t}(-\Sigma_n^{(N)} S_n)$. If Assumption 1 and 2 hold, then*

$$\|\Phi_{s,t}\| \leq e^{-\varrho(t-s)}, \ \|\Psi_{n,0,1}\| \leq 1, \left\|\Psi_{n,0,1}^{(N)}\right\| \leq 1. \tag{5.2}$$

Proof. The first inequality follows from Assumption 1 and Lemma B.1. It can be easily verified that $\Sigma_{n,\lambda}$ and $\Sigma_{n,\lambda}^{(N)}$ are positive semi-definite using (A.11). By Assumption 2, we have $\mu(-\Sigma_{n,\lambda}S_n) < 0$ and $\mu(-\Sigma_{n,\lambda}^{(N)}S_n) < 0$. Using Lemma B.1 again, we obtain the second and third inequalities. ∎

LEMMA 5.2. *For state in (3.1) and conditional moments in (3.4), and for $\forall\, p \geq 1$, we have the following results.*

- *If Assumption 1 holds, then*

$$\|X_t\|_p \leq C, \ \forall\ t \geq 0, \tag{5.3}$$

$$\|\Sigma_{n,\lambda}\|_p \leq C, \ \forall\ n \geq 0, \lambda \in [0,1]. \tag{5.4}$$

- *If Assumption 1 and 2 hold, then*

$$\|\upsilon_{n,\lambda}\|_p \leq C, \ \forall\ n \geq 0, \lambda \in [0,1]. \tag{5.5}$$

Here C is a positive constant depending on p and independent of λ and n.

The proof can be found in Appendix A.3.

LEMMA 5.3. *The evolution equations of $m_t^{(N)}$ and $P_t^{(N)}$ are as follows:*

- *for* $t \in [t_{n-1}, t_n)$*:*

$$
\begin{aligned}
dm_t^{(N)} &= F_t m_t^{(N)} dt, \\
dP_t^{(N)} &= \left[F_t P_t^{(N)} + P_t^{(N)} F_t^\top + G_t Q_t G_t^\top \right] dt.
\end{aligned} \tag{5.6}
$$

The right limit is denoted as

$$
m_{t_n^-}^{(N)} = \lim_{t \nearrow t_n} m_t^{(N)}, \ P_{t_n^-}^{(N)} = \lim_{t \nearrow t_n} P_t^{(N)}. \tag{5.7}
$$

- *for* $t = t_n$*: Let* $v_{n,0}^{(N)} = m_{t_n^-}^{(N)}$, $\Sigma_{n,0}^{(N)} = P_{t_n^-}^{(N)}$, *for* $\lambda \in [0, 1]$, *then*

$$
\begin{aligned}
dv_{n,\lambda}^{(N)} &= \Sigma_{n,\lambda}^{(N)} H_n^\top R_n^{-1} \left[Y_n - H_n v_{n,\lambda}^{(N)} \right] d\lambda, \\
d\Sigma_{n,\lambda}^{(N)} &= -\Sigma_{n,\lambda}^{(N)} H_n^\top R_n^{-1} H_n \Sigma_{n,\lambda}^{(N)} d\lambda.
\end{aligned} \tag{5.8}
$$

Then $m_{t_n}^{(N)} = v_{n,1}^{(N)}$, $P_{t_n}^{(N)} = \Sigma_{n,1}^{(N)}$. *And*

$$
t = t_n : \begin{aligned}
m_{t_n}^{(N)} &= m_{t_n^-}^{(N)} + \mathrm{K}_n^{(N)} \left(Y_n - H_n m_{t_n^-}^{(N)} \right), \\
P_{t_n}^{(N)} &= P_{t_n^-}^{(N)} - \mathrm{K}_n^{(N)} H_n P_{t_n^-}^{(N)},
\end{aligned} \tag{5.9}
$$

where the gain function $\mathrm{K}_n^{(N)} := P_{t_n^-}^{(N)} H_n^\top \left[H_n P_{t_n^-}^{(N)} H_n^\top + R_n \right]^{-1}$.

The proof can be found in Appendix A.4.

The connection between the conditional moments (m_t, P_t) of the actual state and their approximations $(m_t^{(N)}, P_t^{(N)})$ obtained using OTPF is displayed in Fig. 2, which corresponds to Fig. 1. Instead of evolving discretely in the updating step according to (3.8) and (5.9), we provide an equivalent continuous approach according to (3.13) and (5.8). This enables us to construct OTPF.

THEOREM 5.4. *If Assumption 1 and 2 hold, then for* $\forall\ p \geq 1,\ n \geq 0$, *we have*

$$
\left\| m_{t_n}^{(N)} - m_{t_n} \right\|_p \lesssim_p \frac{1}{\sqrt{N}} e^{-\varrho t_n}, \tag{5.10}
$$

$$
\left\| P_{t_n}^{(N)} - P_{t_n} \right\|_p \lesssim_p \frac{1}{\sqrt{N}} e^{-2\varrho t_n}. \tag{5.11}
$$

Proof. **Step 1:** Consider the difference matrix $\Theta_t := P_t^{(N)} - P_t$. For $t \in [t_{n-1}, t_n)$, using (3.7) and (5.6), we derive the evolution equation:

$$
d\Theta_t / dt = F_t \Theta_t + \Theta_t F_t^\top, \tag{5.12}
$$

which yields the transition equation:

$$
\Theta_{t_n^-} = \Phi_{t_{n-1}, t_n} \Theta_{t_{n-1}} \Phi_{t_{n-1}, t_n}^\top. \tag{5.13}
$$

Define $\Pi_{n,\lambda} := \Sigma_{n,\lambda}^{(N)} - \Sigma_{n,\lambda}$. Initially, $\Pi_{n,0} = \Theta_{t_n^-}$. With (3.13) and (5.8), we obtain the differential equation:

$$
\begin{aligned}
d\Pi_{n,\lambda} / d\lambda =& \Sigma_{n,\lambda} S_n \Sigma_{n,\lambda} - \Sigma_{n,\lambda}^{(N)} S_n \Sigma_{n,\lambda}^{(N)} \\
=& -\Pi_{n,\lambda} S_n \Sigma_{n,\lambda} - \Sigma_{n,\lambda}^{(N)} S_n \Pi_{n,\lambda},
\end{aligned} \tag{5.14}
$$

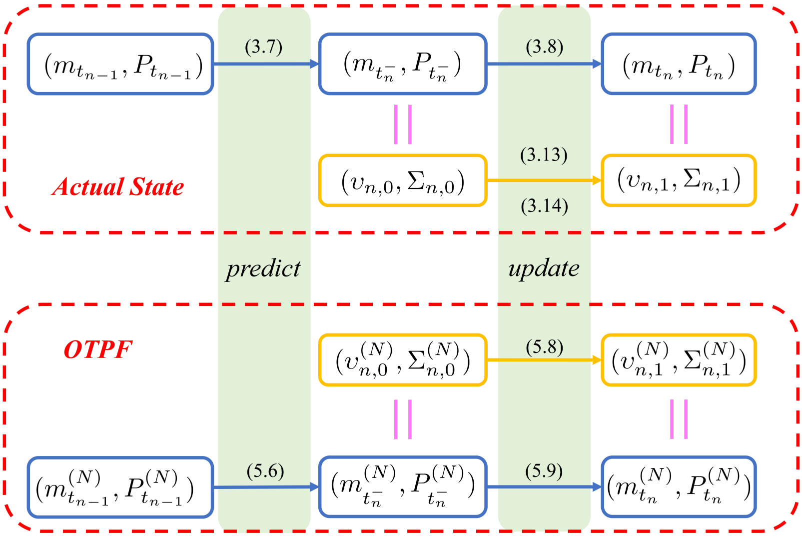

FIG. 2. *The connection between the conditional moments* (m_t, P_t) *of the actual state and their approximations* $(m_t^{(N)}, P_t^{(N)})$ *obtained using OTPF.*

leading to the solution:

$$\Pi_{n,\lambda} = \Psi_{n,0,\lambda}^{(N)} \Pi_{n,0} \Psi_{n,0,\lambda}^{\top}. \tag{5.15}$$

Subsequently, we have

$$\Theta_{t_n} = \Pi_{n,1} = \Psi_{n,0,1}^{(N)} \Phi_{t_{n-1},t_n} \Theta_{t_{n-1}} \Phi_{t_{n-1},t_n}^{\top} \Psi_{n,0,1}^{\top}. \tag{5.16}$$

and by utilizing (5.2), we arrive at the bound:

$$\left\|\Theta_{t_n^-}\right\|_p \leq e^{-2\varrho t_n} \left\|\Theta_0\right\|_p, \ \left\|\Theta_{t_n}\right\|_p \leq e^{-2\varrho t_n} \left\|\Theta_0\right\|_p, \ \left\|\Pi_{n,\lambda}\right\|_p \leq e^{-2\varrho t_n} \left\|\Theta_0\right\|_p. \tag{5.17}$$

According to Theorem B.3, we ascertain that:

$$\left\|\Theta_0\right\|_p \lesssim_p \frac{1}{\sqrt{N}}. \tag{5.18}$$

Consequently, we deduce (5.11).

Step 2: Let $e_t := m_t^{(N)} - m_t$. For $t \in [t_{n-1}, t_n)$, using equations (3.7) and (5.6), we have

$$de_t/dt = F_t e_t, \tag{5.19}$$

from which we obtain

$$e_{t_n^-} = \Phi_{t_{n-1},t_n} e_{t_{n-1}}. \tag{5.20}$$

Introduce $\theta_{n,\lambda} := \upsilon^{(N)}_{n,\lambda} - \upsilon_{n,\lambda}$, with $\theta_{n,0} = e_{t_n^-}$. Utilizing (3.13) and (5.8), we have

$$\begin{aligned} d\theta_{n,\lambda}/d\lambda =& \Sigma^{(N)}_{n,\lambda} H_n^\top R_n^{-1} \left[Y_n - H_n \upsilon^{(N)}_{n,\lambda} \right] - \Sigma_{n,\lambda} H_n^\top R_n^{-1} \left[Y_n - H_n \upsilon_{n,\lambda} \right] \\ =& \Sigma^{(N)}_{n,\lambda} H_n^\top R_n^{-1} \left[Y_n - H_n \upsilon^{(N)}_{n,\lambda} \right] - \Sigma^{(N)}_{n,\lambda} H_n^\top R_n^{-1} \left[Y_n - H_n \upsilon_{n,\lambda} \right] \\ &+ \Sigma^{(N)}_{n,\lambda} H_n^\top R_n^{-1} \left[Y_n - H_n \upsilon_{n,\lambda} \right] - \Sigma_{n,\lambda} H_n^\top R_n^{-1} \left[Y_n - H_n \upsilon_{n,\lambda} \right] \\ =& -\Sigma^{(N)}_{n,\lambda} S_n \theta_{n,\lambda} + \Pi_{n,\lambda} H_n^\top R_n^{-1} \left[Y_n - H_n \upsilon_{n,\lambda} \right], \end{aligned} \tag{5.21}$$

which leads to the solution:

$$\theta_{n,1} = \Psi^{(N)}_{n,0,1} \theta_{n,0} + \int_0^1 \Psi^{(N)}_{n,\lambda,1} \Pi_{n,\lambda} H_n^\top R_n^{-1} \left[Y_n - H_n \upsilon_{n,\lambda} \right] d\lambda. \tag{5.22}$$

Subsequently, the norm of e_{t_n} can be bounded as follows:

$$\begin{aligned} \|e_{t_n}\|_p &= \|\theta_{n,1}\|_p \\ &\le \|\theta_{n,0}\|_p + \int_0^1 \left\| \Pi_{n,\lambda} H_n^\top R_n^{-1} \left[Y_n - H_n \upsilon_{n,\lambda} \right] \right\|_p d\lambda \\ &\le \|\theta_{n,0}\|_p + C \int_0^1 \|\Pi_{n,\lambda}\|_{2p} \|Y_n - H_n \upsilon_{n,\lambda}\|_{2p} d\lambda \\ &\le \|\theta_{n,0}\|_p + \frac{C_p}{\sqrt{N}} e^{-2\varrho t_n} \left(\|Y_n\|_{2p} + \|H_n\|_{2p} \sup_{\lambda \in [0,1]} \|\upsilon_{n,\lambda}\|_{2p} \right) \\ &\le e^{-\varrho \Delta t} \left\| e_{t_{n-1}} \right\|_p + \frac{C_p}{\sqrt{N}} e^{-2\varrho t_n} \end{aligned} \tag{5.23}$$

utilizing (5.2), Hölder's inequality, (5.17), (A.9) and (5.5). Here C is a positive constant and C_p is a positive constant depending on p. From (5.23) we have

$$\begin{aligned} \|e_{t_n}\|_p &\le e^{-\varrho \Delta t} \left\| e_{t_{n-1}} \right\|_p + \frac{C_p}{\sqrt{N}} e^{-2\varrho t_n} \\ &\le e^{-\varrho \Delta t} \left(e^{-\varrho \Delta t} \left\| e_{t_{n-2}} \right\|_p + \frac{C_p}{\sqrt{N}} e^{-2\varrho t_{n-1}} \right) + \frac{C_p}{\sqrt{N}} e^{-2\varrho t_n} \\ &= e^{-\varrho 2\Delta t} \left\| e_{t_{n-2}} \right\|_p + \frac{C_p}{\sqrt{N}} e^{-2\varrho t_n} \left(1 + e^{\varrho \Delta t} \right) \\ &\le e^{-\varrho 3\Delta t} \left\| e_{t_{n-3}} \right\|_p + \frac{C_p}{\sqrt{N}} e^{-2\varrho t_n} \left(1 + e^{\varrho \Delta t} + e^{\varrho 2\Delta t} \right) \\ &\le \cdots \\ &\le e^{-\varrho t_n} \|e_0\|_p + \frac{C_p}{\sqrt{N}} e^{-2\varrho t_n} \left(1 + e^{\varrho \Delta t} + e^{\varrho 2\Delta t} + \cdots + e^{\varrho (n-1)\Delta t} \right) \\ &\lesssim_p \frac{1}{\sqrt{N}} e^{-\varrho t_n} \end{aligned} \tag{5.24}$$

using Theorem B.3. □

6. Experiments. This section investigates the effectiveness of three distinct particle techniques: FPF, OTPF, and the standard PF, considering various particle counts. The KF is employed as the benchmark for optimal performance. The analysis encompasses the entire temporal duration from $t = 0$ to $t = 10$ seconds, utilizing Euler's method for temporal discretization with a consistent step size of $\Delta t = 0.01s$. The particle number, N, varies within the set $\{20, 50, 100, 200, 500, 1000\}$. To effectively assess and compare the different strategies, we employ the mean squared error (MSE) metric, calculated over 100 independent trials, defined as follows:

$$\mathrm{MSE}(t) := \frac{1}{100}\sum_{i=1}^{100}\frac{1}{N_t+1}\sum_{n=0}^{N_t}\left(\bar{X}_t^{(i)} - \hat{X}_t^{(\mathrm{KF},i)}\right)^2, \tag{6.1}$$

where $\bar{X}_t^{(i)}$ is the estimate of X_t by particle algorithms, $\hat{X}_t^{(\mathrm{KF},i)}$ is the estimate of X_t by the KF in the i-th experiment, and $N_t = \lfloor t/\Delta t \rfloor$ and $\lfloor\cdot\rfloor$ is the floor function. The Mean of Time (MT) is defined as the average running time over 100 independent trials.

6.1. Time-invariant case. In the time-invariant case, we consider the continuous filtering system with discrete observation described by the following stochastic differential equations:

$$\begin{cases} dX_t = AX_t dt + \sigma_V dV_t, \\ Y_n = X_{t_n} + \sigma_W W_n, \end{cases} \tag{6.2}$$

where $A = [a_{ij}]_{i,j=1}^{10}$ is a 10×10 matrix with entries defined as follows:

$$a_{ij} = \begin{cases} 0.1, & \text{if } |i-j| = 1, \\ -0.5, & \text{if } i = j, \\ 0, & \text{otherwise}, \end{cases}$$

$\sigma_V = 1$, $\sigma_W = 0.5$, $\{V_t\}$ is the standard Brownian motion process, $\{W_n\}$ is the standard Gaussian white noise, and $\{V_t\}$ and $\{W_n\}$ are mutually independent. Discrete observations are available at times $t_n \in \{0.5, 1, \cdots, 10\}$.

Evidently, the mean squared error (MSE(t)) exhibits dependency on both t and N. We examine the fluctuations of MSE(t) concerning t and N through the implementation of three distinct PF algorithms. The outcomes are illustrated in Fig 3 to 4, and the detailed results can be found in Table 1.

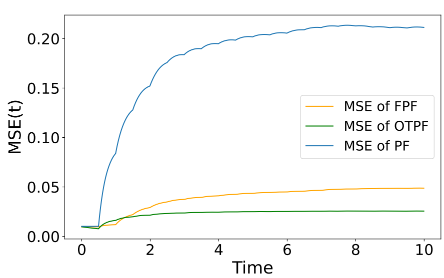

FIG. 3. *MSE*(t) *with* $N = 100$

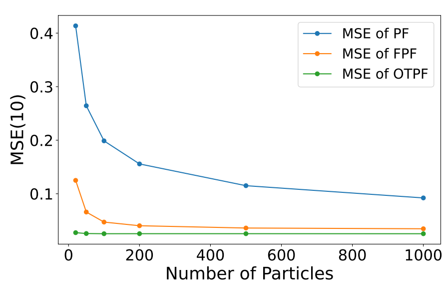

FIG. 4. *MSE*(10) *via different* N

Algorithm	N	MSE(10)	MT
KF	20	-	0.041382
PF	20	0.397661	0.210371
FPF	20	0.125943	0.125573
OTPF	20	0.027279	0.276602
KF	50	-	0.041382
PF	50	0.274631	0.236267
FPF	50	0.064574	0.147266
OTPF	50	0.025905	0.287698
KF	100	-	0.041382
PF	100	0.210356	0.232881
FPF	100	0.048279	0.155715
OTPF	100	0.026051	0.290129
KF	200	-	0.041382
PF	200	0.164222	0.279779
FPF	200	0.041263	0.200644
OTPF	200	0.026036	0.326475
KF	500	-	0.041382
PF	50	0.108608	0.396964
FPF	500	0.035015	0.340288
OTPF	500	0.024993	0.426351
KF	1000	-	0.041382
PF	1000	0.094468	0.598110
FPF	1000	0.035176	0.550917
OTPF	1000	0.025310	0.571500

TABLE 1
The MSE and MT for time-invariant case

6.2. Time-varying case. In the time-varying case, we consider the following system:

$$\begin{cases} dX_t = A(t)X_t dt + \sigma_V dV_t, \\ \ \ Y_n = X_{t_n} + \sigma_W W_n, \end{cases} \tag{6.3}$$

where $A(t) = [a_{ij}(t)]_{i,j=1}^{10}$ is a 10×10 matrix with entries defined as follows:

$$a_{ij} = \begin{cases} 0.1 \cdot \cos(t), & \text{if } |i - j| = 1, \\ -0.2, & \text{if } i = j, \\ 0, & \text{otherwise}, \end{cases}$$

$H = 1$, $\sigma_V = 1$, $\sigma_W = 0.5$, $\{V_t\}$ is the standard Brownian motion process, $\{W_n\}$ is the standard Gaussian white noise, and $\{V_t\}$ and $\{W_n\}$ are mutually independent. Discrete observations are available at time $t_n \in \{0.5, 1, \ldots, 10\}$.The result is displayed in Fig 5 to 6 and Table 2.

6.3. Numerical conclusion. From the results of two experiments, the following conclusions can be drawn:

Algorithm	N	MSE(10)	MT
KF	20	-	0.047040
PF	20	0.583433	0.210670
FPF	20	0.110731	0.119701
OTPF	20	0.028567	0.262941
KF	50	-	0.047040
PF	50	0.379496	0.236933
FPF	50	0.046666	0.143890
OTPF	50	0.024123	0.294905
KF	100	-	0.047040
PF	100	0.263231	0.237359
FPF	100	0.029449	0.151382
OTPF	100	0.022787	0.280011
KF	200	-	0.047040
PF	200	0.181108	0.277406
FPF	200	0.021380	0.191556
OTPF	200	0.022440	0.310503
KF	500	-	0.047040
PF	50	0.126109	0.428971
FPF	500	0.017475	0.359097
OTPF	500	0.023934	0.358677
KF	1000	-	0.047040
PF	1000	0.107000	0.614912
FPF	1000	0.016312	0.562995
OTPF	1000	0.022938	0.581559

TABLE 2
The MSE and MT for time-invariant case

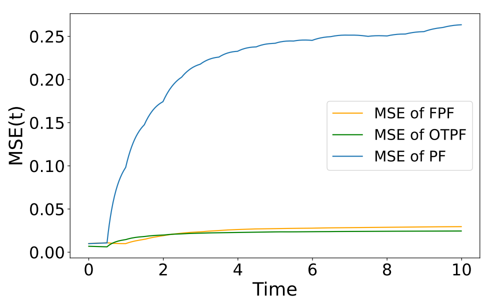

FIG. 5. *MSE*(t) *with* $N = 100$

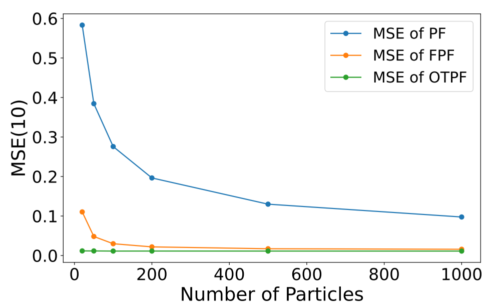

FIG. 6. *MSE*(10) *via different* N

1. In terms of temporal stability, the performance ranking is as follows: OTPF demonstrates the highest stability, followed by FPF, with PF exhibiting the least stability.
2. Considering the influence of particle number, OTPF displays commendable performance even with a limited number of particles. In contrast, PF exhibits substantial improvement with an increasing number of particles but falls short

of approaching the performance achieved by OTPF. With a rising particle count, FPF converges towards the performance level exhibited by OTPF.

3. Regarding computational time, OTPF exhibits slightly higher computational complexity than FPF and PF at lower particle counts. However, as the particle count increases, the computational complexity of OTPF does not escalate as rapidly as that of FPF and PF.

7. Conclusion. In this paper, we have constructed an OTPF for linear time-varying continuous discrete system. Compared with FPF, OTPF has less variance. Additionally, we have evaluated the L^p-errors, quantifying the disparities between the actual posterior moments (m_t, P_t) and their empirical approximations $(m_t^{(N)}, P_t^{(N)})$ obtained through OTPF. It is noteworthy that the explicit optimal transportation between non-Gaussian distributions is often cannot be obtained. The consideration of extending OTPF to accommodate general nonlinear filtering systems represents an avenue for future research efforts.

Appendix A. Proofs.

A.1. Proof of Theorem 2.3. We need to prove that $p(x,t) = p^*(x,t)$ for any $t \geq 0$. It can be easily proved that $p(x,t)$ also follows the same evolution equation (2.6) as $p^*(x,t)$ during $[t_{n-1}, t_n)$ in the prediction step. Therefore, we focus on demonstrating that they share the same evolution equation during the updating step, given identical initial values at $t = 0$. Equivalently, we only need to prove that $\rho_n(x, \lambda)$ and $\rho_n^*(x, \lambda)$ have the same evolution equation during $\lambda \in [0, 1]$.

By employing (2.16) and (2.17), we obtain

$$-\nabla \cdot (\rho \mathcal{K}) = (h - \hat{h})^\top R^{-1} \rho, \tag{A.1}$$

Subsequently, using (2.18) and (2.19), we find:

$$\begin{aligned} -\nabla \cdot (\rho u) =& \frac{1}{2} \nabla \cdot \left(\rho \mathcal{K}(h + \hat{h}) \right) - \frac{1}{2} \nabla \cdot (\rho \Omega) \\ =& -\frac{1}{2}(h - \hat{h})^\top R^{-1}(h + \hat{h})\rho + \frac{1}{2} \sum_{j=1}^{d_2} \rho (\nabla \eta_j)^\top \nabla h_j + \frac{1}{2}\xi\rho - \frac{1}{2}\bar{\xi}\rho \\ =& -\frac{1}{2} h^\top R^{-1} h + \frac{1}{2}(\widehat{h^\top R^{-1} h}). \end{aligned} \tag{A.2}$$

Substituting (A.1) and (A.2) into (2.15), we obtain:

$$\frac{\partial \rho_n}{\partial \lambda} = \rho_n \left[(h - \hat{h})^\top R^{-1} Y_n - \frac{1}{2} h^\top R^{-1} h + \frac{1}{2}(\widehat{h^\top R^{-1} h}) \right]. \tag{A.3}$$

This equation aligns precisely with the evolution equation (2.12) for ρ^*.

A.2. Proof of Lemma 3.3. On the one hand, by (3.4), we can obtain

$$\begin{aligned} \frac{\partial \rho_n}{\partial \lambda} =& -\frac{1}{2}\rho_n \operatorname{Tr}\left(\Sigma_{n,\lambda}^{-1} \frac{\partial \Sigma_{n,\lambda}}{\partial \lambda} \right) + \rho_n \left\{ \frac{1}{2} \left(\frac{\partial \upsilon_{n,\lambda}}{\partial \lambda} \right)^\top \Sigma_{n,\lambda}^{-1} (x - \upsilon_{n,\lambda}) \right. \\ & \left. + \frac{1}{2} (x - \upsilon_{n,\lambda})^\top \Sigma_{n,\lambda}^{-1} \frac{\partial \Sigma_{n,\lambda}}{\partial \lambda} \Sigma_{n,\lambda}^{-1} (x - \upsilon_{n,\lambda}) + \frac{1}{2} (x - \upsilon_{n,\lambda})^\top \Sigma_{n,\lambda}^{-1} \left(\frac{\partial \upsilon_{n,\lambda}}{\partial \lambda} \right) \right\}. \end{aligned} \tag{A.4}$$

On the other hand, since $\rho_n = \rho_n^*$, using (2.12), we have

$$\frac{\partial \rho_n}{\partial \lambda} = \rho_n \left[(x - \upsilon_{n,\lambda})^\top H_n^\top R_n^{-1} Y_n - \frac{1}{2} x^\top H_n^\top R_n^{-1} H_n x + \frac{1}{2} \big(\widehat{x^\top H^\top R_n^{-1} H x} \big) \right]. \tag{A.5}$$

Since the right hand sides of (A.4) and (A.5) are equal for $\forall\, x$, by comparing the coefficients of the first and quadratic terms of x, we can obtain

$$\begin{cases} \left(\dfrac{\partial \upsilon_{n,\lambda}}{\partial \lambda}\right)^{\top} \Sigma_{n,\lambda}^{-1} - \upsilon_{n,\lambda}^{\top} \Sigma_{n,\lambda}^{-1} \dfrac{\partial \Sigma_{n,\lambda}}{\partial \lambda} \Sigma_{n,\lambda}^{-1} = Y_n^{\top} R_n^{-1} H_n, \\ \dfrac{1}{2} \Sigma_{n,\lambda}^{-1} \dfrac{\partial \Sigma_{n,\lambda}}{\partial \lambda} \Sigma_{n,\lambda}^{-1} = -\dfrac{1}{2} H_n^{\top} R_n^{-1} H_n, \end{cases} \tag{A.6}$$

from which we get (3.13).

A.3. Proof of Lemma 5.2. Step 1: By (3.1), we have

$$X_t = \Phi_{0,t} X_0 + \int_0^t \Phi_{s,t} G_s dB_s. \tag{A.7}$$

Then

$$\begin{aligned} \|X_t\|_p &\le \|\Phi_{0,t} X_0\|_p + \left\| \int_0^t \Phi_{s,t} G_s dB_s \right\|_p \\ &\lesssim_p \|\Phi_{0,t}\| \, \|X_0\|_p + \left\| \left[\int_0^t \mathrm{Tr}\left(\Phi_{s,t} G_s Q_s G_s^{\top} \Phi_{s,t}^{\top} \right) ds \right]^{1/2} \right\|_p \\ &\lesssim_{p,d_1} \|\Phi_{0,t}\| \, \|X_0\|_p + \left[\int_0^t \left(\|\Phi_{s,t}\|^2 \|G_s\|^2 \|Q_s\| \right) ds \right]^{1/2} \\ &\lesssim_{p,d_1} e^{-\varrho t} + \left[\int_0^t \left(e^{-2\varrho(t-s)} \right) ds \right]^{1/2} \\ &\le C, \end{aligned} \tag{A.8}$$

where we use Burkholder–Davis–Gundy inequality [13], (5.2) and the fact that X_0 is Gaussian. Now we obtain (5.3).

It follows that

$$\|Y_n\|_p \le \|H_n\| \, \|X_{t_n}\|_p + \|W_n\|_p \le C, \tag{A.9}$$

where we use the fact that W_n is Gaussian and C is a positive constant independent of n.

Step 2: By (3.7) and (3.8), it is known that P_t is deterministic since P_0 is deterministic, it follows that $\|P_t\|_p = \|P_t\|$. The updating step of P_t can be rewritten as

$$P_{t_n} = [I - K_n H_n] \, P_{t_n^-} \, [I - K_n H_n]^{\top} + K_n R_n K_n^{\top}, \tag{A.10}$$

from which and (3.7), it is known that

$$P_{t_n} \ge 0, \; P_{t_{n+1}^-} \ge 0, \; \forall\, n \ge 0, \tag{A.11}$$

since $P_0 \ge 0$. Now by (3.7) and (3.8), we have

$$P_{t_n^-} = \Phi_{t_{n-1},t} P_{t_{n-1}} \Phi_{t_{n-1},t_n}^{\top} + \int_{t_{n-1}}^{t_n} \Phi_{s,t_n} G_s Q_s G_s^{\top} \Phi_{s,t_n}^{\top} ds, \; P_{t_n} \le P_{t_n^-}. \tag{A.12}$$

We define a new matrix $\widetilde{P}_t$ with $\widetilde{P}_0 = P_0$ as follows:

$$\frac{\mathrm{d}\widetilde{P}_t}{\mathrm{d}t} = F_t\widetilde{P}_t + \widetilde{P}_tF_t^\top + G_tQ_tG_t^\top, \ \forall\, t \geq 0.$$

Then $\forall\, n \geq 1$, we have

$$\widetilde{P}_{t_n^-} = \Phi_{t_{n-1},t}\widetilde{P}_{t_{n-1}}\Phi_{t_{n-1},t_n}^\top + \int_{t_{n-1}}^{t_n} \Phi_{s,t_n}G_sQ_sG_s^\top\Phi_{s,t_n}^\top ds, \ \widetilde{P}_{t_n} = \widetilde{P}_{t_n^-}. \tag{A.13}$$

It follows that $\widetilde{P}_{t_n}$ is positive semidefinite,

$$\widetilde{P}_{t_n} \geq P_{t_n^-} \geq P_{t_n}. \tag{A.14}$$

and using Assumption 1,

$$\begin{aligned} \left\|\widetilde{P}_{t_n}\right\| &\leq \left\|\Phi_{0,t_n}P_0\Phi_{0,t_n}^\top\right\| + \int_0^{t_n} \left\|\Phi_{s,t_n}G_sQ_sG_s^\top\Phi_{s,t_n}^\top\right\| ds \\ &\leq \left\|\Phi_{0,t_n}\right\|^2\left\|P_0\right\| + \int_0^t \left\|\Phi_{s,t_n}\right\|^2\left\|G_sQ_sG_s^\top\right\| ds \\ &\lesssim_{d_1} e^{-2\varrho t_n} + \int_0^{t_n} e^{-2\varrho(t_n-s)}ds \\ &\lesssim_{d_1} e^{-2\varrho t_n} + 1/(2\varrho), \end{aligned} \tag{A.15}$$

from which and Lemma B.2, we know that

$$C \geq \left\|\widetilde{P}_{t_n}\right\| \geq \left\|P_{t_n^-}\right\| \geq \left\|P_{t_n}\right\|, \tag{A.16}$$

where C is a constant independent of n.

By (3.13), it is known

$$P_{t_n} = \Sigma_{n,1} \leq \Sigma_{n,\lambda} \leq \Sigma_{n,0} = P_{t_n^-}, \ \lambda \in [0,1]. \tag{A.17}$$

Therefore $\Sigma_{n,\lambda}$ is positive semidefinite. Then we obtain (5.4) using (A.16) and Lemma B.2.

Step 3: Using Jensen's inequality and (5.3), we have

$$\mathbb{E}\left[\left\|m_{t_n^-}\right\|^p\right] = \mathbb{E}\left[\left\|\mathbb{E}[X_{t_n}|\mathscr{F}_{t_{n-1}}]\right\|^p\right] \leq \mathbb{E}\left[\mathbb{E}[\left\|X_{t_n}\right\|^p|\mathscr{F}_{t_{n-1}}]\right] = \mathbb{E}[\left\|X_{t_n}\right\|^p] \leq C. \tag{A.18}$$

By (3.13), we have

$$\upsilon_{n,\lambda} = \Psi_{n,0,\lambda}\upsilon_{n,0} + \int_0^\lambda \Psi_{n,s,\lambda}\Sigma_{n,s}H_n^\top R_n^{-1}Y_n ds. \tag{A.19}$$

It follows that

$$\begin{aligned} \left\|\upsilon_{n,\lambda}\right\|_p &\leq \left\|\Psi_{n,0,\lambda}\right\|\left\|\upsilon_{n,0}\right\|_p + \int_0^\lambda \left\|\Psi_{n,s,\lambda}\right\|\left\|\Sigma_{n,s}\right\|\left\|H_nR_n^{-1}\right\|\left\|Y_n\right\|_p ds \\ &\lesssim \left\|m_{t_n^-}\right\|_p + \left\|Y_n\right\|_p\int_0^\lambda \left\|\Sigma_{n,s}\right\| ds \\ &\lesssim C, \end{aligned} \tag{A.20}$$

using (5.2), (5.4) and (A.9).

A.4. Proof of Lemma 5.3. The first equation in (5.6) can be obtained directly (4.16) and (4.17). As for the second equation, we first define

$$e_t^i := \widetilde{X}_t^i - m_t^{(N)}.$$

Then we have $de_t^i = \Xi_t^{(N)} e_t^i dt$ by (4.16) and the first equation in (5.6). It follows that

$$d\left[e_t^i (e_t^i)^\top\right]/dt = \Xi_t^{(N)} e_t^i (e_t^i)^\top + e_t^i (e_t^i)^\top \Xi_t^{(N)}, \tag{A.21}$$

from which we have

$$\begin{aligned} dP_t^{(N)}/dt =& \Xi_t^{(N)} P_t^{(N)} + P_t^{(N)} \Xi_t^{(N)} \\ =& F(t) P_t^{(N)} + P_t^{(N)} F^\top(t) + G(t) Q(t) G^\top(t), \end{aligned} \tag{A.22}$$

where the second equality comes from (4.18).

Similarly, we can prove (5.8). And (5.9) comes from (5.8) and Corollary 3.4.

Appendix B. Some known results.

LEMMA B.1 ([1]). *Let $A : u \mapsto A_u$ and $B : u \mapsto B_u$ be the smooth flows of $(r \times r)$ matrices. For any $s \leqslant t$ we have*

$$\|\mathcal{E}_{s,t}(A+B)\| \leqslant \exp\left(\int_s^t \mu\left(A_u\right) du + \int_s^t \|B_u\|\, du\right).$$

LEMMA B.2. *For any two $r \times r$ positive semidefinite matrices A, B, if $A \geq B$, then $\|A\| \geq \|B\|$.*

Proof. The 2-norm of a positive semidefinite matrix A is expressed as:

$$\|A\| = \sqrt{\lambda_{\max}\left(A^\top A\right)} = \lambda_{\max}(A). \tag{B.1}$$

Hence, our objective is to prove $\lambda_{\max}(A) \geq \lambda_{\max}(B)$.

To initiate the proof, we establish the equivalence:

$$\lambda_{\max}(A) = \max_{\|\alpha\|=1} \alpha^\top A \alpha, \tag{B.2}$$

valid for any positive semidefinite matrix A. On one hand, for any vector $\alpha \in \mathbb{R}^r$ with $\|\alpha\| = 1$, we have $\alpha^\top A \alpha \leq \lambda_{\max}(A)$, a result established in prior literature [20]. On the other hand, there exists an eigenvector β such that $A\beta = \lambda_{\max}(A)\beta$, and consequently, $\beta_0^\top A \beta_0 = \lambda_{\max}(A)$ holds with $\beta_0 = \beta / \|\beta\|$. This leads to the derivation of (B.2).

For any vector $\alpha \in \mathbb{R}^r$ with $\|\alpha\| = 1$, we proceed as follows:

$$\alpha^\top A \alpha = \alpha^\top (A - B + B)\alpha = \alpha^\top (A - B)\alpha + \alpha^\top B \alpha \geq \alpha^\top B \alpha,$$

utilizing the fact that $A \geq B$. Consequently, by combining (B.1) and (B.2), the desired inequality $\lambda_{\max}(A) \geq \lambda_{\max}(B)$ is established. ∎

THEOREM B.3. *Consider n-dimensional random vectors* $X_i \overset{i.i.d}{\sim} \mathcal{N}(m, P)$ *for* $i = 1, 2, \ldots, N$. *Define the sample mean and covariance matrix as follows:*

$$m^{(N)} := \frac{1}{N} \sum_{i=1}^{N} X_i,$$
$$P^{(N)} := \frac{1}{N-1} \sum_{i=1}^{N} \left(X_i - m^{(N)} \right) \left(X_i - m^{(N)} \right)^{\top}.$$

For any $p \geq 1$, *the following inequalities hold:*

$$\|m^{(N)} - m\|_p \leq C_{n,p} \frac{1}{\sqrt{N}}. \tag{B.3}$$

and

$$\|P^{(N)} - P\|_p \leq \bar{C}_{n,p} \frac{1}{\sqrt{N}}, \tag{B.4}$$

where $C_{n,p}$ *and* $\bar{C}_{n,p}$ *denote some positive parameters dependent on* n *and* p.

Proof. Let $\|\circ\|_{\mathrm{F}}$ represent the Frobenius norm of matrix $\circ$. Using norm inequality $\|A\| \leq \|A\|_{\mathrm{F}}$ and

$$\mathbb{E}\left[\|m^{(N)} - m\|^p\right]^{\frac{1}{p}} \leq C_{n,p} \frac{1}{\sqrt{N}}$$
$$\mathbb{E}\left[\|P^{(N)} - P\|_{\mathrm{F}}^p\right]^{\frac{1}{p}} \leq \bar{C}_{n,p} \frac{1}{\sqrt{N}}$$

by Corollary B.5 in [3], we can obtain the desired result. ∎

REFERENCES

[1] A. N. BISHOP AND P. DEL MORAL, *On the stability of Kalman–Bucy diffusion processes*, SIAM Journal on Control and Optimization, 55 (2017), pp. 4015–4047.

[2] X. CHEN, X. LUO, J. SHI, AND S. S.-T. YAU, *General convergence result for continuous-discrete feedback particle filter*, International Journal of Control, 95 (2022), pp. 2972–2986.

[3] X. CHEN AND S. S.-T. YAU, *On the stability of linear feedback particle filter*, Asian Journal of Mathematics, 27 (2023), pp. 95–120.

[4] Z. CHEN, *Bayesian filtering: from Kalman filters to particle filters, and beyond*, Statistics: A Journal of Theoretical and Applied Statistics, 182 (2003), pp. 1–69.

[5] G. EVENSEN, *Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics*, Journal of Geophysical Research: Oceans, 99 (1994), pp. 10143–10162.

[6] N. J. GORDON, D. J. SALMOND, AND A. F. M. SMITH, *Novel approach to nonlinear/non-Gaussian Bayesian state estimation*, Radar and Signal Processing, IEE Proceedings F, 140 (1993), pp. 107–113.

[7] A. H. JAZWINSKI, *Stochastic processes and filtering theory*, Academic Press, New York and London, 1970.

[8] R. E. KALMAN, *A new approach to linear filtering and prediction problems*, Journal of Basic Engineering, 82 (1960), pp. 35–45.

[9] G. Y. KULIKOV AND M. V. KULIKOVA, *Accurate continuous–discrete unscented Kalman filtering for estimation of nonlinear continuous-time stochastic models in radar tracking*, Signal Processing, 139 (2017), pp. 25–35.

[10] X. LUO AND S. S.-T. YAU, *Complete real time solution of the general nonlinear filtering problem without memory*, IEEE Transactions on Automatic Control, 58 (2013), pp. 2563–2578.

[11] A. I. Perov and I. D. Kostrub, *On the spectral abscissa and the logarithmic norm*, Mathematical Notes, 101 (2017), pp. 677–687.
[12] G. Peyré, M. Cuturi, et al., *Computational optimal transport: with applications to data science*, Foundations and Trends® in Machine Learning, 11 (2019), pp. 355–607.
[13] B. L. Rozovsky and S. V. Lototsky, *Stochastic evolution systems: linear theory and applications to non-linear filtering*, vol. 89, Springer, 2018.
[14] A. Taghvaei and P. G. Mehta, *An optimal transport formulation of the linear feedback particle filter*, in 2016 American Control Conference (ACC), IEEE, 2016, pp. 3614–3619.
[15] A. Taghvaei and P. G. Mehta, *Error analysis for the linear feedback particle filter*, in 2018 Annual American Control Conference (ACC), IEEE, 2018, pp. 4261–4266.
[16] T. Yang, H. A. Blom, and P. G. Mehta, *The continuous-discrete time feedback particle filter*, in 2014 American Control Conference, IEEE, 2014, pp. 648–653.
[17] T. Yang, R. S. Laugesen, P. G. Mehta, and S. P. Meyn, *Multivariable feedback particle filter*, Automatica, 71 (2016), pp. 10–23.
[18] T. Yang, P. G. Mehta, and S. P. Meyn, *Feedback particle filter*, IEEE Transactions on Automatic Control, 58 (2013), pp. 2465–2480.
[19] S.-T. Yau and S. S.-T. Yau, *Real time solution of the nonlinear filtering problem without memory II*, SIAM Journal on Control and Optimization, 47 (2008), pp. 163–195.
[20] F. Zhang, *Matrix theory: basic results and techniques*, Springer Science & Business Media, 2011.