Automatica 167 (2024) 111740

Contents lists available at ScienceDirect

Automatica

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

Brief paper

Time-varying feedback particle filter☆

Xiuqiong Chen [a], Jiayi Kang [b], Stephen S.-T. Yau [c,b,*]

[a] School of Mathematics, Renmin University of China, Beijing, 100872, PR China
[b] Department of Mathematical Sciences, Tsinghua University, Beijing, 100084, PR China
[c] Beijing Institute of Mathematical Sciences and Applications (BIMSA), Beijing, 101408, PR China

ARTICLE INFO

Article history:
Received 13 February 2023
Received in revised form 2 April 2024
Accepted 20 April 2024
Available online xxxx

Keywords:
Feedback particle filter
Kalman filter
Optimal transportation
Error analysis

ABSTRACT

Feedback particle filter is a novel Monte Carlo algorithm with identically distributed particles evolving under feedback control structure, such that the Kullback–Leibler divergence between the actual posterior of the state and the common posterior of any particle can be minimized. In this work, we consider the time-varying linear systems and explicitly analyze the errors between the optimal solution obtained by Kalman filter and the estimates given by feedback particle filter and the optimal transportation particle filter, respectively. These theoretical analyses are also supported by the numerical simulation, where we compare the performances of particle filter, feedback particle filter, optimal transportation particle filter and Kalman filter.

1. Introduction

Filtering problems arise in a variety of areas such as control, finance, aerospace and so on. The general continuous time-varying filtering system can be modeled as the following stochastic differential equations (SDEs) on the probability space $(\Omega, \mathscr{F}, P)$:

$$\begin{cases} dX_t = f(X_t, t)dt + g(X_t, t)dB_t, \\ dZ_t = h(X_t, t)dt + dW_t, \end{cases} \tag{1}$$

where $t \in [0, T]$, X_t is the n-dimensional state, Z_t is the m-dimensional observation with $Z_0 = 0$, $\{B_t\}_{t\geq 0}$ and $\{W_t\}_{t\geq 0}$ are r- and m-dimensional Brownian motions, respectively, with $\mathbb{E}\left[dB_t dB_t^{\mathrm{T}}\right] = Q_t dt$ and $\mathbb{E}\left[dW_t dW_t^{\mathrm{T}}\right] = R_t dt$. Here, we call (1) "time-varying" system since the drift term f, diffusion term g, observation term h and the variances of the noises have the explicit dependence on time t.

Define the σ-algebra formed by the observations till to time t as $\mathscr{F}_t := \sigma\{Z_s : \ 0 \leq s \leq t\} = \sigma(\bigcup_{0\leq s\leq t} Z_s^{-1}(\mathscr{B}(\mathbb{R}^m)))$, where $\mathscr{B}(\mathbb{R}^m)$ is the Borel set of $\mathbb{R}^m$. The state process X_t is indirectly observed through the observation process Z_t and the goal of the standard filtering problem is to seek the optimal estimate of the state X_t given the observation history $\mathscr{F}_t$. The "optimality" here refers to minimizing the mean squared error between X_t and its estimate, and the optimal estimate is $\mathbb{E}[X_t|\mathscr{F}_t]$ (Jazwinski, 1970). Obviously, this problem can be completely solved if we can get the conditional density function $p(X_t|\mathscr{F}_t)$ of the state X_t given $\mathscr{F}_t$. When system (1) is linear Gaussian, the density $p(X_t|\mathscr{F}_t)$ is Gaussian and the optimal solution can be obtained by the famous Kalman filter (KF) (Kalman & Bucy, 1961).

When system (1) is nonlinear, things get much more complicated to obtain $p(X_t|\mathscr{F}_t)$. In a general sense, one idea is to solve the Duncan–Mortensen–Zakai (DMZ) equation (Zakai, 1969), which is satisfied by the unnormalized conditional density. And this equation can be efficiently solved by the Yau–Yau algorithm (Luo & Yau, 2013; Yau & Yau, 2008).

The other one is based on the Monte Carlo methods. Particle filter (PF) is one of the most famous algorithm (Gordon, Salmond, & Smith, 1993), which uses the empirical distribution formed by a large number of particles to approximate the posterior distribution of the state. However, PF suffers from some disadvantages such as weight degeneracy problem. Recently, an alternative approach of PF, named feedback particle filter (FPF), is proposed in Yang, Mehta, and Meyn (2013). Apparently, the empirical distribution is determined by the number of particles N and the conditional density $p(X_t^i|\mathscr{F}_t)$ of the particles X_t^i, $1 \leq i \leq N$. In FPF, all particles $\{X_t^i\}_{i=1}^N$ are identically distributed and evolve according to the following SDE:

$$d\overline{X}_t = f(\overline{X}_t, t)dt + g(\overline{X}_t, t)d\overline{B}_t + u(\overline{X}_t, t)dt + K(\overline{X}_t, t)dZ_t, \tag{2}$$

where $\overline{B}_t$ is an independent copy of state noise B_t, $p(\overline{X}_0) = p(X_0)$, the gain functions $u(x, t)$ and $K(x, t)$ are obtained by minimizing

☆ This work is supported by National Natural Science Foundation of China (NSFC) grant (12201631, 11961141005) and Tsinghua University Education Foundation fund, PR China (042202008). The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Gianluigi Pillonetto under the direction of Editor Alessandro Chiuso.

* Corresponding author at: Beijing Institute of Mathematical Sciences and Applications (BIMSA), Beijing, 101408, PR China.

E-mail addresses: cxq0828@ruc.edu.cn (X. Chen), kangjy19@mails.tsinghua.edu.cn (J. Kang), yau@uic.edu (S.S.-T. Yau).

https://doi.org/10.1016/j.automatica.2024.111740

the Kullback–Leibler divergence between the actual posterior $p(X_t|\mathscr{F}_t)$ and the posterior $p(\overline{X}_t|\mathscr{F}_t)$ of the particle. And the equations satisfied by the optimal control input $\{u(x,t), K(x,t)\}$ for (2) can be easily obtained following the trivial extension of Yang, Laugesen, Mehta, and Meyn (2016) from time-invariant system to time-varying system.

As for the convergence of FPF, Taghvaei and Mehta analyzed linear FPF in Taghvaei and Mehta (2018b), and analyzed the optimal transport formulation of the linear Gaussian FPF in Taghvaei and Mehta (2018a), respectively. Later, in Taghvaei and Mehta (2020), a detailed error analysis was carried out for the deterministic form of the optimal FPF. Chen et al. analyzed the mean squared error of FPF for time-invariant continuous linear systems in Chen and Yau (2023), and analyzed the L^p error of FPF for nonlinear systems with continuous states and discrete observations in Chen, Luo, Shi, and Yau (2022).

However, all the existing discussions about FPF are limited to time-invariant cases. In the real world, due to the reasons outside and inside the system, parameter changes are inevitable. So strictly speaking, almost all systems belong to the category of time-varying systems. Examples include the time-varying control in unmanned aerial vehicles (Magree & Johnson, 2016) and variable interest rate in life cycle permanent income hypothesis (Tanizaki, 1996).

In this paper, we consider the linear Gaussian time-varying systems. We investigate FPF and also provide another way to formulate the FPF, which optimally transport the particles from prior distribution to posterior distribution. In this new formulation, the noise term in the evolution equation is replaced by a deterministic one and we call this filter optimal transportation particle filter (OTPF). Furthermore, the estimation errors of two algorithms have also been explicitly analyzed.

The contributions of this work can be summarized as follows.

- In linear case, we provide the uniform estimates of the state X_t, sample mean $m_t^{(N)}$ and covariance $P_t^{(N)}$ of FPF, which can be found in Theorem 3.
- The L^p errors of the estimates provided by the FPF and OTPF are explicitly provided. Furthermore, it can be seen that the L^p asymptotic error of OTPF is zero while that of FPF is order $\mathcal{O}(1/\sqrt{N})$ for any $p \geq 1$, where N is the number of particles. And these results are given in Theorems 4 and 5.

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 parameter depending on p and q. Let $\|\cdot\|$ represent the Euclidean norm of vectors and the 2-norm of matrices. The Frobenius matrix norm of a given $(n_1 \times n_2)$-matrix A is defined by

$$\|A\|_{\mathrm{F}}^2 = \mathrm{Tr}\left(A^{\mathrm{T}} A\right) \quad \text{with the trace operator } \mathrm{Tr}(\circ).$$

Besides, $\forall\ p \geq 1$, we define the L^p norms $\|\circ\|_{2,p} := \mathbb{E}^{1/p}[\|\circ\|^p]$ and $\|\circ\|_{\mathrm{F},p} := \mathbb{E}^{1/p}[\|\circ\|_{\mathrm{F}}^p]$ for any vectors and matrices satisfying $\mathbb{E}[\|\circ\|^p] < \infty$. For an $(n \times n)$-matrix A, let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the minimal and maximal eigenvalues of A, respectively. We define the logarithmic norm $\mu(A)$ of an $(n \times n)$-square matrix A by

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

One thing to note is that $\mu(\cdot)$ is not a matrix norm since it is not positive-valued. It can be seen that $\left(A + A^{\mathrm{T}}\right)/2$ is negative semidefinite as soon as $\mu(A) < 0$. And we also have

$$\mu(A) \geqslant \max\{\mathrm{Re}(\lambda) : \lambda \text{ is the eigenvalue of } A\}, \tag{3}$$

where $\mathrm{Re}(\lambda)$ stands for the real part of the eigenvalues λ. Let $\mathbb{S}_n$ and $\mathbb{S}_n^+$ represent the sets of all $n \times n$ real symmetric matrices and real positive definite matrices, respectively. Let A, B be $(r \times r)$-matrices, $\circ_{i,j}$ denote the (i,j)-th entry of any matrix $\circ$ with $1 \leq i,j \leq r$, and $\circ_{(i,k),(j,l)}$ denote the $(r(i-1)+k, r(j-1)+l)$-th entry of an $(r^2 \times r^2)$-matrix $\circ$ with $1 \leq i,j,k,l \leq r$. Then we define the tensor product $A \otimes^\sharp B$ which is an $\left(r^2 \times r^2\right)$-matrix with entry computed by

$$(A \otimes^\sharp B)_{(i,k),(j,l)} = A_{i,k} B_{j,l},\ \ \forall\ 1 \leq i,k,j,l \leq r.$$

And we also define the symmetric tensor product $A \otimes_s B$ which is an $\left(r^2 \times r^2\right)$-matrix with entry computed by:

$$4\,(A \otimes_s B)_{(i,j),(k,l)} = A_{i,k}B_{j,l} + A_{i,l}B_{j,k} + A_{j,l}B_{i,k} + A_{j,k}B_{i,l},$$

$\forall\ 1 \leq i,k,j,l \leq r$. The angle bracket $\langle M \rangle$ of an r-column-vector continuous martingale M is the $(r \times r)$-matrix $\langle M \rangle$ such that $MM^{\mathrm{T}} - \langle M \rangle$ is a martingale. More generally, the angle bracket $\langle M \rangle^\sharp$ of an $(r \times r)$-matrix valued continuous martingale M is the $\left(r^2 \times r^2\right)$-matrix $\langle M \rangle^\sharp$ such that $M \otimes^\sharp M^{\mathrm{T}} - \langle M \rangle^\sharp$ is a martingale.

The organization of this paper is as follows. In Section 2, we shall introduce three filtering algorithms, i.e., KF, FPF and OTPF. In Section 3, we shall analyze the errors between the optimal solution provided by KF and the estimates given by FPF and OTPF. In Section 4, we will use a linear numerical example to verify our analyses. The conclusion will be drawn in Section 5.

2. Linear filtering algorithms

In the following sections, we will focus on the linear Gaussian case of the general system (1), i.e., we consider the following linear system:

$$\begin{cases} dX_t = A_t X_t dt + G_t dB_t, \\ dZ_t = H_t X_t dt + dW_t, \end{cases} \tag{4}$$

where the initial state $X_0 \sim \mathcal{N}(m_0, P_0)$ with $P_0 > 0$ is assumed to be Gaussian and also independent of Brownian motion processes $\{B_t\}_{t\geq 0}$ and $\{W_t\}_{t\geq 0}$. Let us denote

$$\widetilde{Q}_t := G_t Q_t G_t^{\mathrm{T}},\ \ S_t := H_t^{\mathrm{T}} R_t^{-1} H_t. \tag{5}$$

We assume that the eigenvalues of $\widetilde{Q}_t$ and S_t are uniformly bounded positive throughout the reminder of this paper.

2.1. Kalman filter

It is well known that the conditional distribution of the state X_t given observation history $\mathscr{F}_t$ for (4) is Gaussian. More precisely, $p(X_t|\mathscr{F}_t) = \mathcal{N}(m_t, P_t)$, where m_t and P_t are the conditional mean and covariance of the state X_t given the observation history $\mathscr{F}_t$, respectively. It is well known that m_t and P_t satisfy the KF (Jazwinski, 1970):

$$dm_t = A_t m_t dt + P_t H_t^{\mathrm{T}} R_t^{-1}\left(dZ_t - H_t m_t dt\right), \tag{6}$$

$$\frac{dP_t}{dt} = \mathrm{Ricc}(P_t), \tag{7}$$

where $\mathrm{Ricc}(\cdot) : \mathbb{S}_n^+ \to \mathbb{S}_n$ is the Riccati drift function defined for any $\Sigma \in \mathbb{S}_n^+$ by

$$\mathrm{Ricc}(\Sigma) := A_t\Sigma + \Sigma A_t^{\mathrm{T}} - \Sigma S_t \Sigma + \widetilde{Q}_t.$$

We make the following assumption w.r.t. system (4).

Assumption 1. The system (4) is uniformly completely observable and uniformly completely controllable (Jazwinski, 1970).

It has been proved that, under Assumption 1, for any $P_0 \in \mathbb{S}_n^+$, the time-varying Riccati flow P_t is well defined and a unique solution exists $\forall\ t \geq 0$. Furthermore, P_t is uniformly bounded, see Bishop and Del Moral (2017) for more details.

2.2. Linear FPF

It can be easily checked that, for linear system (4), the explicit form of FPF (2) with the optimal gain functions is

$$d\overline{X}_t = A_t\overline{X}_t dt + G_t d\overline{B}_t + \overline{P}_t H_t^{\mathrm{T}} R_t^{-1}\left(dZ_t - H_t\frac{\overline{X}_t + \overline{m}_t}{2}\right), \tag{8}$$

which is the conditional Mckean–Vlasov diffusion process, and $\overline{m}_t$ and $\overline{P}_t$ are the conditional mean and covariance of $\overline{X}_t$ given $\mathscr{F}_t$, respectively. For diffusion process (8), similar to Taghvaei and Mehta (2016), we have the following result.

Lemma 1. *Consider KF (6)–(7) and Mckean–Vlasov diffusion process (8). If $\overline{m}_0 = m_0$, $\overline{P}_0 = P_0$, then $\forall\ t \geq 0$, we have $\overline{m}_t = m_t$, $\overline{P}_t = P_t$. Furthermore, if $p(\overline{X}_0) = p(X_0)$, then we have $p(\overline{X}_t|\mathscr{F}_t) = p(X_t|\mathscr{F}_t)$, $\forall\ t \geq 0$.*

In the following contents, we shall write $\overline{m}_t = m_t$, $\overline{P}_t = P_t$ as we assume $p(\overline{X}_0) = p(X_0)$. However, we cannot solve the Mckean–Vlasov SDE (8) because the exact $\overline{P}_t$ and $\overline{m}_t$ cannot be obtained. Instead, we use the following evolution equation for the N particles $\{X_t^i\}_{i=1}^N$:

$$\begin{aligned} dX_t^i =& A_t X_t^i dt + G_t dB_t^i \\ &+ P_t^{(N)} H_t^{\mathrm{T}} R_t^{-1}\left(dZ_t - H_t\frac{X_t^i + m_t^{(N)}}{2}dt\right), \end{aligned} \tag{9}$$

where $m_t^{(N)}$ and $P_t^{(N)}$ are the sample mean and covariance of $\{X_t^i\}_{i=1}^N$, respectively, which are computed by

$$\begin{aligned} \overline{m}_t &\approx m_t^{(N)} := \frac{1}{N}\sum_{i=1}^N X_t^i, \\ \overline{P}_t &\approx P_t^{(N)} := \frac{1}{N-1}\sum_{i=1}^N \left(X_t^i - m_t^{(N)}\right)\left(X_t^i - m_t^{(N)}\right)^{\mathrm{T}}. \end{aligned} \tag{10}$$

The initial particles are generated according to $X_0^i \overset{i.i.d}{\sim} \mathcal{N}(m_0, P_0)$, and $\{B_t^i\}_{i=1}^N$ are N independent copies of B_t. The evolution equations of $m_t^{(N)}$ and $P_t^{(N)}$ are listed in the following lemma.

Lemma 2. *The evolutions of $m_t^{(N)}$ and $P_t^{(N)}$ satisfy*

$$\begin{aligned} dm_t^{(N)} =& A_t m_t^{(N)} dt + \frac{1}{\sqrt{N}} d\overline{M}_t \\ &+ P_t^{(N)} H_t^{\mathrm{T}} R_t^{-1}\left(dZ_t - H_t m_t^{(N)} dt\right), \\ dP_t^{(N)} =& \mathrm{Ricc}(P_t^{(N)})dt + \frac{1}{\sqrt{N-1}} dM_t, \end{aligned} \tag{11}$$

where $\overline{M}_t$ is a vector-valued martingale with $\frac{d}{dt}\langle\overline{M}\rangle_t = \widetilde{Q}_t$, and M is a matrix-valued continuous martingale with $\frac{d}{dt}\langle M\rangle_t^{\sharp} = 4P_t^{(N)} \otimes_s \widetilde{Q}_t$.

The proof can be found in Appendix A.

2.3. Linear OTPF

Actually, the optimal control law $\{u, K\}$ in FPF (2) is not unique (Taghvaei & Mehta, 2016). To find a unique control law, one way to formulate the filtering problem is to use optimal transportation. In this way, the particles following the initial distribution $p(X_0)$ can be optimally transported to particles following the posterior $p(X_t|\mathscr{F}_t)$. Next, we shall review the optimal transportation briefly.

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 a quadratic cost aims to minimize

$$\min_T \mathbb{E}\left[\|T(X) - X\|^2\right] \tag{12}$$

over all measurable maps $T : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$X \sim \mu_X, T(X) \sim \mu_Y.$$

The minimizer T^* is called the optimal transport map between μ_X and μ_Y, if it exists.

Theorem 1 (*Optimal Map Between Gaussians Peyré, Cuturi, et al., 2019*)**.** *For the optimization problem (12), if $\mu_X = \mathcal{N}(m_X, P_X)$ and $\mu_Y = \mathcal{N}(m_Y, P_Y)$ are Gaussian distributions, with $P_X, P_Y > 0$, then the optimal transport map T^* between μ_X and μ_Y is given by*

$$T^*(x) = m_Y + P_Y^{\frac{1}{2}}\left(P_Y^{\frac{1}{2}} P_X P_Y^{\frac{1}{2}}\right)^{-\frac{1}{2}} P_Y^{\frac{1}{2}}(x - m_X).$$

Now we aim to construct a stochastic process $\{\widetilde{X}_t\}$ with evolution equation

$$d\widetilde{X}_t = \tilde{u}(\widetilde{X}_t, t)dt + \widetilde{K}(\widetilde{X}_t, t)dZ_t, \tag{13}$$

and we hope $p(\widetilde{X}_t|\mathscr{F}_t)$ is equal to the posterior density $p(X_t|\mathscr{F}_t)$ of the state X_t for all $t \geq 0$. The evolution of $\{\widetilde{X}_t\}$ is not unique (Taghvaei & Mehta, 2016) and we want to obtain an optimal evolution equation by the following time stepping optimization procedure:

(1) Divide the time interval $[0, T]$ into N_1 segments equally and denote the instants $0 = t_0 < t_1 < \cdots < t_{N_1} = T$.
(2) Let $\widetilde{X}_0 \sim p(X_0)$, i.e., $p(\widetilde{X}_0) = p(X_0)$.
(3) For each time step $[t_k, t_{k+1}]$, evolve $\widetilde{X}_t$ according to
$$\widetilde{X}_{t_{k+1}} = T^*_{t_k,t_{k+1}}(\widetilde{X}_{t_k}),\ \forall\ 0 \leq k \leq N_1 - 1,$$
where $T^*_{t_k,t_{k+1}}$ is the optimal transport map from $p(X_{t_k}|\mathscr{F}_{t_k})$ to $p(X_{t_{k+1}}|\mathscr{F}_{t_{k+1}})$.
(4) Let $N_1 \to \infty$, we obtain a SDE (13) for $\widetilde{X}_t$ with optimal $\{\tilde{u}, \widetilde{K}\}$ which is referred to the OTPF.

The detailed procedures can refer to Taghvaei and Mehta (2016). Following the similar procedures in Proposition 3 of Taghvaei and Mehta (2016), we can get the following conclusion.

Proposition 2. *Under Assumption 1, for the linear Gaussian system (4), the SDE in OTPF is*

$$\begin{aligned} d\widetilde{X}_t =& A_t m_t dt + P_t H_t^{\mathrm{T}} R_t^{-1}(dZ_t - H_t m_t dt) \\ &+ \Theta_t\left(\widetilde{X}_t - m_t\right)dt, \end{aligned} \tag{14}$$

where Θ_t is the solution to

$$\Theta_t P_t + P_t \Theta_t = \mathrm{Ricc}(P_t). \tag{15}$$

If $p(\widetilde{X}_0) = p(X_0)$, then $\forall\ t \geq 0$, we have $p(\widetilde{X}_t|\mathscr{F}_t) = p(X_t|\mathscr{F}_t)$.

It is known that (15) is a Lyapunov equation and it admits a unique solution given $P_t > 0$. Furthermore, Θ_t is symmetric and can be written in the following form:

$$\Theta_t = A_t - \frac{1}{2}P_t S_t + \frac{1}{2}\widetilde{Q}_t P_t^{-1} + \bar{\Theta}_t P_t^{-1}, \tag{16}$$

where $\bar{\Theta}_t$ is an $n \times n$ skew symmetric matrix and is the solution to

$$\begin{aligned} &\bar{\Theta}_t P_t^{-1} + P_t^{-1}\bar{\Theta}_t \\ =& A_t^{\mathrm{T}} - A_t + \frac{1}{2}(P_t S_t - S_t P_t) + \frac{1}{2}\left(P_t^{-1}\widetilde{Q}_t - \widetilde{Q}_t P_t^{-1}\right). \end{aligned}$$

Similar to linear FPF, we cannot obtain the exact P_t and m_t in (14). Instead, we use the following evolution equation for the N particles $\{\widetilde{X}_t^i\}_{i=1}^N$:

$$d\widetilde{X}_t^i = A_t\tilde{m}_t^{(N)}dt + \widetilde{P}_t^{(N)}H_t^{\mathrm{T}}R_t^{-1}\left(dZ_t - H_t\tilde{m}_t^{(N)}dt\right) + \Theta_t^{(N)}\left(\widetilde{X}_t^i - \tilde{m}_t^{(N)}\right)dt, \tag{17}$$

where $\tilde{m}_t^{(N)}$ and $\widetilde{P}_t^{(N)}$ are the sample mean and covariance of $\{\widetilde{X}_t^i\}_{i=1}^N$ similar to (10),

$$\Theta_t^{(N)} := A_t - \frac{1}{2}P_t^{(N)}S_t + \frac{1}{2}\widetilde{Q}_t\left(P_t^{(N)}\right)^{-1} + \bar{\Theta}_t^{(N)}\left(P_t^{(N)}\right)^{-1},$$

and $\bar{\Theta}_t^{(N)}$ is the solution to

$$\begin{aligned}&\bar{\Theta}_t^{(N)}\left(P_t^{(N)}\right)^{-1} + \left(P_t^{(N)}\right)^{-1}\bar{\Theta}_t^{(N)}\\ =&A_t^{\mathrm{T}} - A_t + \frac{1}{2}\left(P_t^{(N)}S_t - S_tP_t^{(N)}\right)\\ &+\frac{1}{2}\left[\left(P_t^{(N)}\right)^{-1}\widetilde{Q}_t - \widetilde{Q}_t\left(P_t^{(N)}\right)^{-1}\right].\end{aligned} \tag{18}$$

The initial particles are generated according to $\widetilde{X}_0^i \overset{i.i.d}{\sim} \mathcal{N}(m_0, P_0)$. Similar to Lemma 2, we can obtain the evolution equations of $\tilde{m}_t^{(N)}$ and $\widetilde{P}_t^{(N)}$.

Lemma 3. *The evolutions of $\tilde{m}_t^{(N)}$ and $\widetilde{P}_t^{(N)}$ satisfy*

$$d\tilde{m}_t^{(N)} = A_t\tilde{m}_t^{(N)}dt + \widetilde{P}_t^{(N)}H_t^{\mathrm{T}}R_t^{-1}\left(dZ_t - H_t\tilde{m}_t^{(N)}dt\right), \tag{19}$$

$$d\widetilde{P}_t^{(N)} = \mathrm{Ricc}(\widetilde{P}_t^{(N)})dt. \tag{20}$$

3. Error analysis

In this section, we shall analyze the estimation errors of the FPF and OTPF for linear system (4).

3.1. Error analysis of linear FPF

First of all, we need to make two assumptions.

Assumption 2. A_t in system (4) satisfies $\sup_{t\geq 0}\mu(A_t) < 0$.

By (3), it is known that, under this assumption, A_t is Hurwitz uniformly w.r.t. time t. In other words, Assumption 2 makes sure that the linear system (4) is stable.

Assumption 3. S_t defined in (5) is a scalar matrix, i.e.,

$$S_t = \rho(S_t)I, \text{ for some scalar } \rho(S_t) > 0, \tag{21}$$

where I is an $(n \times n)$-dimensional identity matrix.

Before we continue, we need to give the uniform estimates of the real state X_t of the linear system (4), particle state X_t^i in (9), sample mean $m_t^{(N)}$ and sample covariance $P_t^{(N)}$ of linear FPF defined in (11).

Theorem 3. *For all $p \geq 1$, We have the following uniform estimates:*

- *If Assumptions 1 and 2 are satisfied, then*

$$\sup_{t\geq 0}\left\|\mathrm{Tr}\left(P_t^{(N)}\right)\right\|_{2,p} \lesssim_{n,p} C, \tag{22}$$

$$\sup_{t\geq 0}\|X_t\|_{2,p} \lesssim_{n,p} C, \tag{23}$$

$$\sup_{t\geq 0}\|m_t\|_{2,p} \lesssim_{n,p} C; \tag{24}$$

- *If Assumptions 1–3 are satisfied, then*

$$\sup_{t\geq 0}\left\|m_t^{(N)}\right\|_{2,p} \lesssim_{n,p} C, \tag{25}$$

$$\sup_{t\geq 0}\|X_t^i\|_{2,p} \lesssim_{n,p} C, \forall\ 1 \leq i \leq N, \tag{26}$$

where C is a positive constant and n is the dimension of the state.

We postpone the proof to Appendix B to avoid distraction. Now we are ready to give the estimation error of FPF (9).

Theorem 4. *The Frobenius norm of the sample covariance matrix fluctuations satisfies the diffusion equation*

$$\begin{aligned}&d\left\|P_t^{(N)} - P_t\right\|_F^2\\ =&2\,\mathrm{Tr}\left\{\left[A_t + A_t^{\mathrm{T}} - \frac{1}{2}\left(P_t^{(N)} + P_t\right)S_t\right.\right.\\ &\left.\left.-\frac{1}{2}S_t\left(P_t^{(N)} + P_t\right)\right]\left(P_t^{(N)} - P_t\right)^2\right\}dt\\ &+\frac{2}{N-1}\left[\mathrm{Tr}\left(P_t^{(N)}\widetilde{Q}_t\right) + \mathrm{Tr}\left(P_t^{(N)}\right)\mathrm{Tr}\left(\widetilde{Q}_t\right)\right]dt\\ &+\frac{2}{\sqrt{N-1}}d\mathcal{M}_t,\end{aligned} \tag{27}$$

where $\mathcal{M}_t$ is a martingale with $d\langle\mathcal{M}\rangle_t = 4\,\mathrm{Tr}\{P_t^{(N)}(P_t^{(N)} - P_t)\widetilde{Q}_t(P_t^{(N)} - P_t)\}dt$. And the Euclidean norm of the sample mean vector fluctuations satisfies the diffusion equation

$$\begin{aligned}&d\left\|m_t^{(N)} - m_t\right\|^2\\ =&\left(m_t^{(N)} - m_t\right)^{\mathrm{T}}\left[A_t + A_t^{\mathrm{T}} - P_t^{(N)}S_t - S_tP_t^{(N)}\right]\left(m_t^{(N)} - m_t\right)dt\\ &+\mathrm{Tr}\left(\frac{1}{N}\widetilde{Q}_t + \left(P_t^{(N)} - P_t\right)S_t\left(P_t^{(N)} - P_t\right)\right)dt\\ &+2\left(m_t^{(N)} - m_t\right)^{\mathrm{T}}\left(P_t^{(N)} - P_t\right)S_t\left(X_t - m_t\right)dt + d\overline{\mathcal{M}}_t,\end{aligned} \tag{28}$$

where $\overline{\mathcal{M}}_t$ is a martingale with $d\langle\overline{\mathcal{M}}\rangle_t = 4\left(m_t^{(N)} - m_t\right)^{\mathrm{T}} \cdot \left(\frac{1}{N}\widetilde{Q}_t + \left(P_t^{(N)} - P_t\right)S_t\left(P_t^{(N)} - P_t\right)\right)\left(m_t^{(N)} - m_t\right)dt$. Furthermore, if Assumptions 1–3 hold, then $\forall\ p \geq 1$, we have the following error estimates:

$$\left\|P_t^{(N)} - P_t\right\|_{F,p} \lesssim_{n,p} e^{\alpha t/2}\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{N}} \tag{29}$$

$$\left\|m_t^{(N)} - m_t\right\|_{2,p} \lesssim_{n,p} e^{\alpha t/8}\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{N}} \tag{30}$$

for all $N > 1,\ t \geq 0$, where $\alpha := 2\sup_{t\geq 0}\mu(A_t)$.

Proof. Apparently, (29)–(30) hold when $t = 0$ by Theorem B.3 in Chen and Yau (2023).

Step 1: Define the error matrix $\Xi_t := P_t^{(N)} - P_t$. We aim to prove (27).

According to (7) and (11), we have

$$\begin{aligned}d\Xi_t =&\left[A_t - \frac{1}{2}\left(P_t^{(N)} + P_t\right)S_t\right]\Xi_t dt\\ &+\Xi_t\left[A_t - \frac{1}{2}\left(P_t^{(N)} + P_t\right)S_t\right]^{\mathrm{T}}dt + \frac{1}{\sqrt{N-1}}dM_t,\end{aligned}$$

where M_t is a matrix-valued martingale with $\frac{d}{dt}\langle M\rangle_t^\sharp = 4P_t^{(N)}\otimes_s \widetilde{Q}_t$. Using Itô's lemma, we have

$$\begin{aligned}
&d\Xi_t^2\\
=&\Xi_t\left[A_t-\frac{1}{2}\left(P_t^{(N)}+P_t\right)S_t\right]\Xi_t dt\\
&+\Xi_t^2\left[A_t-\frac{1}{2}\left(P_t^{(N)}+P_t\right)S_t\right]^{\mathrm{T}}dt\\
&+\left[A_t-\frac{1}{2}\left(P_t^{(N)}+P_t\right)S_t\right]\Xi_t^2 dt\\
&+\Xi_t\left[A_t-\frac{1}{2}\left(P_t^{(N)}+P_t\right)S_t\right]^{\mathrm{T}}\Xi_t dt\\
&+\frac{1}{N-1}\left[P_t^N\widetilde{Q}_t+\widetilde{Q}_tP_t^N+\mathrm{Tr}\left(\widetilde{Q}_t\right)P_t^N+\mathrm{Tr}\left(P_t^N\right)\widetilde{Q}_t\right]dt\\
&+dM_t\frac{\Xi_t}{\sqrt{N-1}}+\frac{\Xi_t}{\sqrt{N-1}}dM_t,
\end{aligned}$$

from which we obtain (27) and $d\mathcal{M}_t = \mathrm{Tr}\left(\Xi_t dM_t\right) = \sum_{1\le k,l\le n}\Xi_t(l,k)dM_t(k,l)$, with $\circ(k,l)$ denoting the (k,l)-th entry of matrix $\circ$. Therefore

$$\begin{aligned}
\frac{d}{dt}\langle\mathcal{M}\rangle_t =& \sum_{1\le k,l,k',l'\le n}\Xi_t(l,k)\Xi_t(l',k')\Big[P_t^{(N)}(k,k')\widetilde{Q}_t(l,l')\\
&+P_t^{(N)}(k,l')\widetilde{Q}_t(l,k')+P_t^{(N)}(l,l')\widetilde{Q}_t(k,k')\\
&+P_t^{(N)}(l,k')\widetilde{Q}_t(k,l')\Big]\\
=&4\,\mathrm{Tr}\left\{P_t^{(N)}\Xi_t\widetilde{Q}_t\Xi_t\right\}.
\end{aligned}$$

Step 2: We shall prove (29) in this step.
Rewrite (27) as $d\left\|\Xi_t\right\|_{\mathrm{F}}^2 = \mathcal{L}_t\left\|\Xi_t\right\|_{\mathrm{F}}^2 dt + \frac{2}{\sqrt{N-1}}d\mathcal{M}_t$, where

$$\begin{aligned}
\frac{d}{dt}\langle\mathcal{M}\rangle_t \le& 4\mu(\widetilde{Q}_t)\,\mathrm{Tr}\left(P_t^{(N)}\right)\left\|\Xi_t\right\|_{\mathrm{F}}^2,\ \text{and}\\
\mathcal{L}_t\left\|\Xi_t\right\|_{\mathrm{F}}^2 \le& 4\sup_{t\ge0}\mu(A_t)\left\|\Xi_t\right\|_{\mathrm{F}}^2+\frac{4}{N-1}\mathrm{Tr}\left(P_t^{(N)}\right)\mathrm{Tr}\left(\widetilde{Q}_t\right)
\end{aligned}$$

using Assumption 3. Define

$$\begin{aligned}
\alpha :=& 2\sup_{t\ge0}\mu(A_t),\ \beta_t := \frac{4}{N-1}\mathrm{Tr}\left(P_t^{(N)}\right)\mathrm{Tr}\left(\widetilde{Q}_t\right),\\
\gamma_t :=& \frac{16}{N-1}\mu(\widetilde{Q}_t)\,\mathrm{Tr}\left(P_t^{(N)}\right).
\end{aligned}$$

For all $p\ge1$, using (22), we have

$$\begin{aligned}
\sup_{t\ge0}\left\|\beta_t\right\|_{2,p} =& \sup_{t\ge0}\frac{4}{N-1}\mathrm{Tr}\left(\widetilde{Q}_t\right)\left\|\mathrm{Tr}\left(P_t^{(N)}\right)\right\|_{2,p}\lesssim_{n,p}\frac{1}{N},\\
\sup_{t\ge0}\left\|\gamma_t\right\|_{2,p} =& \sup_{t\ge0}\frac{16}{N-1}\mu(\widetilde{Q}_t)\left\|\mathrm{Tr}\left(P_t^{(N)}\right)\right\|_{2,p}\lesssim_{n,p}\frac{1}{N}.
\end{aligned}$$

Then by Lemma 7.1 in Del Moral and Tugaut (2018), we obtain (29).

Step 3: We aim to prove (28).
Define $e_t := m_t^{(N)} - m_t$. Comparing (6) and (11), we can get

$$\begin{aligned}
de_t =& \left[\left(A_t-P_t^{(N)}S_t\right)e_t+\Xi_tS_t(X_t-m_t)\right]dt\\
&+\frac{1}{\sqrt{N}}d\overline{M}_t+\Xi_tH_t^{\mathrm{T}}R_t^{-1}dW_t,
\end{aligned}$$

where $\overline{M}_t$ is defined in (11). Using Itô's lemma, we obtain (28) with $d\overline{\mathcal{M}}_t := 2e_t^{\mathrm{T}}\left(\frac{1}{\sqrt{N}}d\overline{M}_t+\Xi_tH_t^{\mathrm{T}}R_t^{-1}dW_t\right)$.

Step 4: We shall prove (30) in this step.

Rewrite (28) as $d\left\|e_t\right\|^2 = \mathcal{L}_t\left\|e_t\right\|^2dt + d\overline{\mathcal{M}}_t$, where

$$\begin{aligned}
\frac{d}{dt}\langle\overline{\mathcal{M}}\rangle_t \le& 4\left(\frac{1}{N}\mathrm{Tr}(\widetilde{Q}_t)+\mu(S_t)\left\|\Xi_t\right\|_{\mathrm{F}}^2\right)\left\|e_t\right\|^2 \triangleq \bar{\gamma}_t\left\|e_t\right\|^2,\\
\mathcal{L}_t\left\|e_t\right\|^2 \le& \mu(A_t)\left\|e_t\right\|^2+\frac{1}{N}\mathrm{Tr}\left(\widetilde{Q}_t\right)+\mu(S_t)\left\|\Xi_t\right\|_{\mathrm{F}}^2\\
&+2\left|\mu(A_t)\right|^{-1}\left\|\Xi_t\right\|_{\mathrm{F}}^2\left\|S_t\right\|^2\left(\left\|m_t\right\|^2+\left\|X_t\right\|^2\right)\\
\triangleq& \mu(A_t)\left\|e_t\right\|^2+\bar{\beta}_t
\end{aligned}$$

using Assumption 3. By (22)–(24) and (29), we have

$$\begin{aligned}
\sup_{t\ge0}\left\|\bar{\beta}_t\right\|_{2,p} \le& \sup_{t\ge0}\left\{\mathrm{Tr}\left(\widetilde{Q}_t\right)/N+\mu(S_t)\left\|\Xi_t\right\|_{\mathrm{F},2p}^2\right.\\
&+2\sqrt{2}\left|\mu(A_t)\right|^{-1}\left\|\Xi_t\right\|_{\mathrm{F},4p}^2\left\|S_t\right\|^2\\
&\left.\cdot\left(\left\|m_t\right\|_{2,4p}^2+\left\|X_t\right\|_{2,4p}^2\right)\right\}\lesssim_{n,p}\frac{1}{N},\\
\sup_{t\ge0}\left\|\bar{\gamma}_t\right\|_{2,p} \le& 4\sup_{t\ge0}\left\{\frac{1}{N}\mathrm{Tr}(\widetilde{Q}_t)+\mu(S_t)\left\|\Xi_t\right\|_{\mathrm{F},2p}^2\right\}\lesssim_{n,p}\frac{1}{N}.
\end{aligned}$$

Therefore using Assumption 2 and Lemma 7.1 in Del Moral and Tugaut (2018), we obtain (30). □

3.2. Error analysis of linear OTPF

We first need to give an assumption.

Assumption 4. The initial sample covariance $\widetilde{P}_0^{(N)}$ in OTPF (17) is positive definite almost surely.

Now we can given the estimate of the errors of OTPF.

Theorem 5. *If Assumptions 1 and 4 hold, then $\forall\ p\ge1$, we have the following error estimates for OTPF (17):*

$$\left\|\widetilde{P}_t^{(N)}-P_t\right\|_{F,p}\lesssim_{n,p}\frac{1}{\sqrt{N}}e^{-2\varrho t} \tag{31}$$

$$\left\|\tilde{m}_t^{(N)}-m_t\right\|_{2,p}\lesssim_{n,p}\frac{1}{\sqrt{N}}e^{-\varrho t} \tag{32}$$

for all $N>1$, $t\ge0$, where ϱ is a positive constant parameter depending on the system (4).

Proof. When $t=0$, (31)–(32) hold (Chen & Yau, 2023).
Step I: The state transition matrix associated with a smooth flow of any $(r\times r)$-matrix $U:\tau\mapsto U_\tau$ is denoted by $\mathcal{E}_{s,t}(U)$ s.t. for any $s\le t$,

$$\frac{d}{dt}\mathcal{E}_{s,t}(U)=U_t\mathcal{E}_{s,t}(U)\quad\text{and}\quad\partial_s\mathcal{E}_{s,t}(U)=-\mathcal{E}_{s,t}(U)U_s$$

with $\mathcal{E}_{s,s}=\mathrm{I}$, the identity matrix. Define $\Phi_{s,t}:=\mathcal{E}_{s,t}(A-PS)$ and $\Phi_{s,t}^{(N)}:=\mathcal{E}_{s,t}(A-\widetilde{P}^{(N)}S)$. Since both P_t and $\widetilde{P}_t^{(N)}$ satisfy the Riccati equation by (7) and (20), then according to Corollary 4.9 in Bishop and Del Moral (2017), using Assumptions 1 and 4, it can be known that

$$\left\|\Phi_{s,t}\right\|\lesssim_n e^{-\varrho(t-s)},\quad\left\|\Phi_{s,t}^{(N)}\right\|\lesssim_n e^{-\varrho(t-s)}\ a.s., \tag{33}$$

where ϱ is a positive constant.
Step II: Define $\widetilde{\Xi}_t:=\widetilde{P}_t^{(N)}-P_t$. By (7) and (20), we know

$$d\widetilde{\Xi}_t=(A_t-\widetilde{P}_t^{(N)}S_t)\widetilde{\Xi}_tdt+\widetilde{\Xi}_t(A_t-P_tS_t)^{\mathrm{T}}dt, \tag{34}$$

from which we obtain

$$\widetilde{\Xi}_t=\Phi_{0,t}^{(N)}\widetilde{\Xi}_0\Phi_{0,t}^{\mathrm{T}}. \tag{35}$$

Therefore by (33), we get $\left\|\widetilde{\Xi}_t\right\|_{\mathrm{F}}\lesssim_n e^{-2\varrho t}\left\|\widetilde{\Xi}_0\right\|_{\mathrm{F}}$ a.s., from which we obtain (31).

Step III: Define $\tilde{e}_t := \tilde{m}_t^{(N)} - m_t$. By (6) and (19), we get

$$d\tilde{e}_t = \left(A_t - \widetilde{P}_t^{(N)} S_t\right) \tilde{e}_t dt + \widetilde{\Xi}_t H_t^{\mathrm{T}} R_t^{-1} dI_t, \tag{36}$$

where $I_t := Z_t - \int_0^t H_s m_s ds$ is a martingale with $d\langle I \rangle_t = R_t dt$. From (36), we have

$$\tilde{e}_t = \Phi_{0,t}^{(N)} \tilde{e}_0 + \int_0^t \Phi_{s,t}^{(N)} \widetilde{\Xi}_s H_s^{\mathrm{T}} R_s^{-1} dI_s. \tag{37}$$

Using (33), we have

$$\left\| \Phi_{0,t}^{(N)} \tilde{e}_0 \right\|_{2,p} \le \mathbb{E}^{1/p} \left[\left\| \Phi_{0,t}^{(N)} \right\|^p \left\| \tilde{e}_0 \right\|^p \right] \lesssim_{n,p} \frac{1}{\sqrt{N}} e^{-\varrho t}.$$

Using Burkholder–Davis–Gundy inequality (Rozovsky & Lototsky, 2018), (35), and (33), we get

$$\begin{aligned} &\left\| \int_0^t \Phi_{s,t}^{(N)} \widetilde{\Xi}_s H_s^{\mathrm{T}} R_s^{-1} dI_s \right\|_{2,p} \\ \lesssim_p &\left\| \left| \int_0^t \mathrm{Tr}\left(\Phi_{s,t}^{(N)} \widetilde{\Xi}_s S_s \widetilde{\Xi}_s \left(\Phi_{s,t}^{(N)}\right)^{\mathrm{T}} \right) ds \right|^{1/2} \right\|_{2,p} \\ \lesssim_{n,p} &\left\| \left(\int_0^t e^{-2\varrho t - 2\varrho s} ds \right)^{1/2} \left\| \widetilde{\Xi}_0 \right\|_{\mathrm{F}} \right\|_{2,p} \lesssim_{n,p} \frac{1}{\sqrt{N}} e^{-\varrho t}. \end{aligned}$$

Then we obtain (32). □

Remark 1. Comparing the error bounds in Theorems 4 and 5, it can be found that, asymptotically, as $t \to \infty$, the L^p error of OTPF goes to zero, while the L^p error of FPF is of order $\mathcal{O}(1/\sqrt{N})$. This is because we introduce extra noise $\{B_t^i\}_{i=1}^N$ in FPF (9).

4. Simulation

The example we consider here is a linear Gaussian system with independent noises which is as follows:

$$\begin{cases} dX_t = A_t X_t + dB_t, \\ dZ_t = X_t dt + dW_t, \end{cases} \tag{38}$$

where $X_0 \sim \mathcal{N}(0, I_n)$ with the $n \times n$-dimensional identity matrix I_n, $n = 10$, B_t and W_t are standard Brownian motion processes, and $A_t = [a_{ij}(t)]$ is an $(n \times n)$-matrix with elements as follows:

$$a_{ij}(t) = \begin{cases} 0.1, & \text{if } i + 1 = j, \\ -0.4 + 0.1\cos(t), & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases}$$

We show the performances of three kinds of PF algorithms, which are FPF, OTPF, and PF with different numbers of particles, and KF provides the optimal solution. We choose $t \in [0, 10]$ as the whole time interval, use Euler's method in time discretization with the same time step $\Delta t = 0.01$ and the number of particles N is in $\{10, 20, 50, 100, 500\}$. In the experiments, in order to compare the performances of different methods, we introduce the mean squared error (MSE) based on 100 realizations, which is defined as follows:

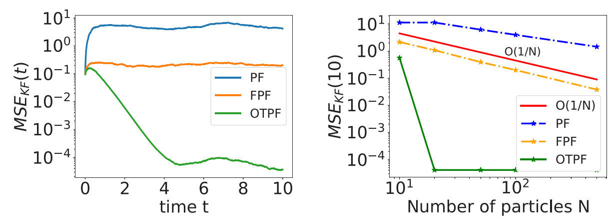
$$\mathrm{MSE} := \frac{1}{100} \sum_{i=1}^{100} \frac{1}{K_1 + 1} \sum_{k=0}^{K_1} \left\| X_{k\cdot\Delta t}^{(i)} - \hat{X}_{k\cdot\Delta t}^{(i)} \right\|, \tag{39}$$

where $X_{k\cdot\Delta t}^{(i)}$ is the real state at discrete time instant $k \cdot \Delta t$ in the i-th experiment and $\hat{X}_{k\cdot\Delta t}^{(i)}$ is the estimation of $X_{k\cdot\Delta t}^{(i)}$, with $0 \le k \le K_1$, where $K_1 = 1000$ is the total time step.

In Table 1, KF gives the optimal results ignoring the numerical errors. It can be seen that OTPF still provides the satisfying result with only 10 particles which is even better than PF with 500 particles. OTPF is almost optimal with about 50 particles, but FPF needs about 500 particles to achieve the same accuracy.

Table 1
The MSE and running time with different particle numbers.

Algorithms	KF	FPF	OTPF	PF	N
MSE	6.5683	8.67313	7.0985	19.6505	10
Time(s)	0.0809	0.1984	0.4311	0.3251	10
MSE	6.5683	7.5596	6.6001	16.2369	20
Time(s)	0.0809	0.3465	0.6109	0.4514	20
MSE	6.5683	7.1080	6.5718	12.8559	50
Time(s)	0.0809	0.69132	0.9928	0.77378	50
MSE	6.5683	6.7807	6.5436	10.9475	100
Time(s)	0.0809	1.1928	1.5532	1.2347	100
MSE	6.5683	6.5756	6.5394	8.5353	500
Time(s)	0.0809	5.2650	6.0152	4.8276	500



(a) $\mathrm{MSE}_{\mathrm{KF}}(t)$ with $N = 100$ (b) $\mathrm{MSE}_{\mathrm{KF}}(t = 10)$

Fig. 1. (a) We fix $N = 100$, vertical axis denotes $\mathrm{MSE}_{\mathrm{KF}}$ which is a function of t, and horizontal axis denotes the time $t \in [0, 10]$. (b) We fix $t = 10$, vertical axis denotes $\mathrm{MSE}_{\mathrm{KF}}(t = 10)$ which is a function of N, and horizontal axis denotes the numbers of particles $N \in \{10, 20, 50, 100, 500\}$.

Next, we shall verify Theorems 4 and 5. We define the MSE w.r.t. the optimal estimate by KF as follows:

$$\mathrm{MSE_{KF}}(t) := \frac{1}{100} \sum_{i=1}^{100} \frac{1}{K_2(t) + 1} \sum_{k=0}^{K_2(t)} \left\| \hat{X}_{k\cdot\Delta t}^{(i)} - \bar{X}_{k\cdot\Delta t}^{(i)} \right\|,$$

where $\bar{X}_{k\cdot\Delta t}^{(i)}$ is the estimate of $X_{k\cdot\Delta t}^{(i)}$ by KF at discrete time instant $k \cdot \Delta t$ in the i-th experiment, $\hat{X}_{k\cdot\Delta t}^{(i)}$ is the estimate of $X_{k\cdot\Delta t}^{(i)}$ by PF algorithms, $K_2(t) = \lfloor t/\Delta t \rfloor$ and $\lfloor \cdot \rfloor$ is the floor function. Apparently, $\mathrm{MSE_{KF}}$ is a function of t and N. We test how $\mathrm{MSE_{KF}}$ varies w.r.t. t and N by three PF algorithms, and the results are displayed in Fig. 1(a)–1(b).

It can be seen that, with fixed N, the $\mathrm{MSE_{KF}}$ of OTPF converges nearly exponentially fast to 0, and with fixed t, the $\mathrm{MSE_{KF}}$ of FPF is of order $\mathcal{O}(1/N)$. These results also verify Theorems 4–5.

5. Conclusion

In this paper, we extended FPF and OTPF to linear Gaussian time-varying cases. Besides, we proved that, the L^p-errors between the optimal estimate and the estimates obtained by linear FPF and OTPF are of order $\mathcal{O}(1/\sqrt{N})$ for any $p \ge 1$. However, the error analysis of FPF for nonlinear systems has not been discussed in this paper and we aim to solve this problem in our future work.

Appendix A. Proof of Lemma 2

Proof. The evolution of $m_t^{(N)}$ can be directly obtained from (9). Define the error process $\zeta_t^i := X_t^i - m_t^{(N)}$, then we can get $d\zeta_t^i = \left(A_t - P_t^{(N)} S_t/2\right) \zeta_t^i + dM_t^i$, where $dM_t^i := G_t \left(dB_t^i - \frac{1}{N} \sum_{j=1}^N dB_t^j \right)$.

Using Itô's lemma, we have

$$
\begin{aligned}
& d\left[\zeta_t^i\left(\zeta_t^i\right)^{\mathrm{T}}\right] \\
=& \left(A_t-\frac{P_t^{(N)} S_t}{2}\right) \zeta_t^i\left(\zeta_t^i\right)^{\mathrm{T}} d t+\zeta_t^i\left(\zeta_t^i\right)^{\mathrm{T}}\left(A_t-\frac{P_t^{(N)} S_t}{2}\right)^{\mathrm{T}} d t \\
& +\left(1-\frac{1}{N}\right) \widetilde{Q}_t d t+d M_t^i\left(\zeta_t^i\right)^{\mathrm{T}}+\zeta_t^i\left(d M_t^i\right)^{\mathrm{T}} .
\end{aligned}
$$

It follows that $dP_t^{(N)} = \mathrm{Ricc}(P_t^{(N)})dt + \frac{1}{\sqrt{N-1}}dM_t$, where $dM_t := \frac{1}{\sqrt{N-1}}\sum_{i=1}^N \left[dM_t^i\left(\zeta_t^i\right)^{\mathrm{T}} + \zeta_t^i\left(dM_t^i\right)^{\mathrm{T}}\right]$. Let $A(k,l)$ denote the (k,l)-th entry of any matrix A and $a(k)$ denote the k-th entry of any vector a. It can be easily known that $dM_t(k,l) = \frac{1}{\sqrt{N-1}}\sum_{i=1}^N \left[dM_t^i(k)\zeta_t^i(l) + \zeta_t^i(k)dM_t^i(l)\right]$. Therefore

$$
\begin{aligned}
& (N-1)\frac{d}{dt}\langle M(k,l), M(k',l')\rangle_t \\
=& \left(1-\frac{1}{N}\right)\sum_{1\le i\le N}\left[\zeta_t^i(k)\zeta_t^i(k')\widetilde{Q}_t(l,l') + \zeta_t^i(k)\zeta_t^i(l')\widetilde{Q}_t(l,k')\right. \\
& \left. +\zeta_t^i(l)\zeta_t^i(l')\widetilde{Q}_t(k,k') + \zeta_t^i(l)\zeta_t^i(k')\widetilde{Q}_t(k,l')\right] \\
& -\frac{1}{N}\sum_{1\le i\ne i'\le N}\left[\zeta_t^i(k)\zeta_t^{i'}(k')\widetilde{Q}_t(l,l') + \zeta_t^i(k)\zeta_t^{i'}(l')\widetilde{Q}_t(l,k')\right. \\
& \left. +\zeta_t^i(l)\zeta_t^{i'}(l')\widetilde{Q}_t(k,k') + \zeta_t^i(l)\zeta_t^{i'}(k')\widetilde{Q}_t(k,l')\right].
\end{aligned}
$$

Since $\sum_{i=1}^N \zeta_t^i = 0$, we know that, $\forall\ 1 \le k,k' \le n$, $\left(\sum_{1\le i\le N}\zeta_t^i(k)\right)\left(\sum_{1\le i'\le N}\zeta_t^{i'}(k')\right) = 0$, from which we can conclude $\frac{1}{N-1}\sum_{1\le i\ne i'\le N}\zeta_t^i(k)\zeta_t^{i'}(k') = -\frac{1}{N-1}\sum_{1\le i\le N}\zeta_t^i(k)\zeta_t^i(k') = -P_t^{(N)}(k,k')$. Then we have

$$
\begin{aligned}
& \frac{d}{dt}\langle M(k,l), M(k',l')\rangle_t \\
=& \left(1-\frac{1}{N}\right)\left[P_t^{(N)}(k,k')\widetilde{Q}_t(l,l') + P_t^{(N)}(k,l')\widetilde{Q}_t(l,k')\right. \\
& \left. +P_t^{(N)}(l,l')\widetilde{Q}_t(k,k') + P_t^{(N)}(l,k')\widetilde{Q}_t(k,l')\right] \\
& +\frac{1}{N}\left[P_t^{(N)}(k,k')\widetilde{Q}_t(l,l') + P_t^{(N)}(k,l')\widetilde{Q}_t(l,k')\right. \\
& \left. +P_t^{(N)}(l,l')\widetilde{Q}_t(k,k') + P_t^{(N)}(l,k')\widetilde{Q}_t(k,l')\right] \\
=& P_t^{(N)}(k,k')\widetilde{Q}_t(l,l') + P_t^{(N)}(k,l')\widetilde{Q}_t(l,k') \\
& + P_t^{(N)}(l,l')\widetilde{Q}_t(k,k') + P_t^{(N)}(l,k')\widetilde{Q}_t(k,l') \\
=& 4\left(P_t^{(N)} \otimes_s \widetilde{Q}_t\right)_{(k,l)(k',l')},
\end{aligned}
$$

which is the desired result. □

Appendix B. Proof of Theorem 3

Proof. The proof is divided into four steps.

Step 1: We shall prove (22). By Lemma 2, we can have

$$
d\,\mathrm{Tr}\left(P_t^{(N)}\right) = \mathcal{L}_t\,\mathrm{Tr}\left(P_t^{(N)}\right)dt + \frac{1}{\sqrt{N-1}}d\mathcal{M}_{1,t},
$$

where $\mathcal{M}_{1,t}$ is a martingale with $\frac{d}{dt}\langle\mathcal{M}_1\rangle_t = 2\,\mathrm{Tr}\left(P_t^{(N)}\widetilde{Q}_t\right) + 2\,\mathrm{Tr}\left(P_t^{(N)}\right)\mathrm{Tr}\left(\widetilde{Q}_t\right) \lesssim_n \sup_{t\ge 0} 4\mu\left(\widetilde{Q}_t\right)\mathrm{Tr}\left(P_t^{(N)}\right)$, and

$$
\begin{aligned}
& \mathcal{L}_t\,\mathrm{Tr}\left(P_t^{(N)}\right) \\
:=& \mathrm{Tr}\left(\left(A_t + A_t^{\mathrm{T}}\right)P_t^{(N)}\right) - \mathrm{Tr}\left(S_t\left(P_t^{(N)}\right)^2\right) + \mathrm{Tr}\left(\widetilde{Q}_t\right) \\
\le& 2\sup_{t\ge 0}\mu\left(A_t\right)\mathrm{Tr}\left(P_t^{(N)}\right) + \sup_{t\ge 0}\mathrm{Tr}\left(\widetilde{Q}_t\right).
\end{aligned}
$$

Then, by Assumption 2, Theorem B.3 in Chen and Yau (2023) and Lemma 7.1 in Del Moral and Tugaut (2018), we obtain (22).

Step 2: We shall prove (23)–(24). By (4) and Itô's lemma, we know $d\left\|X_t\right\|^2 = \mathcal{L}_t\left\|X_t\right\|^2 dt + d\mathcal{M}_{2,t}$, where $\mathcal{M}_{2,t}$ is a martingale with $\frac{d}{dt}\langle\mathcal{M}_2\rangle_t = 4X_t^{\mathrm{T}}\widetilde{Q}_tX_t \le \sup_{t\ge 0}4\mu\left(\widetilde{Q}_t\right)\left\|X_t\right\|^2$, and

$$
\begin{aligned}
\mathcal{L}_t\left\|X_t\right\|^2 :=& X_t^{\mathrm{T}}\left(A_t + A_t^{\mathrm{T}}\right)X_t + \mathrm{Tr}\left(\widetilde{Q}_t\right) \\
\le& 2\sup_{t\ge 0}\mu\left(A_t\right)\left\|X_t\right\|^2 + \sup_{t\ge 0}\mathrm{Tr}\left(\widetilde{Q}_t\right),
\end{aligned}
$$

Hence, by Assumption 2 and Lemma 7.1 in Del Moral and Tugaut (2018), we obtain (23). Then (24) holds observing that $\mathbb{E}\left[\left\|m_t\right\|^p\right] \le \mathbb{E}\left[\mathbb{E}\left[\left\|X_t\right\|^p|\mathscr{F}_t\right]\right] = \mathbb{E}\left[\left\|X_t\right\|^p\right]$.

Step 3: We shall prove (25). By Lemma 2, (4), and Itô's lemma, we have $d\left\|m_t^{(N)}\right\|^2 = \mathcal{L}_t\left\|m_t^{(N)}\right\|^2 dt + d\mathcal{M}_{3,t}$, where $\mathcal{M}_{3,t}$ is a martingale. Similarly, we have

$$
\begin{aligned}
\frac{d}{dt}\langle\mathcal{M}_3\rangle_t =& 4\left(m_t^{(N)}\right)^{\mathrm{T}}\left(\frac{1}{N}\widetilde{Q}_t + P_t^{(N)}S_tP_t^{(N)}\right)m_t^{(N)} \\
\le& \left(\frac{1}{N}\mu\left(\widetilde{Q}_t\right) + \mu\left(S_t\right)\mathrm{Tr}^2\left(P_t^{(N)}\right)\right)\left\|m_t^{(N)}\right\|^2 \\
\triangleq& \gamma_t\left\|m_t^{(N)}\right\|^2
\end{aligned}
$$

and

$$
\begin{aligned}
& \mathcal{L}_t\left\|m_t^{(N)}\right\|^2 \\
\le& \sup_{t\ge 0}\mu\left(A_t\right)\left\|m_t^{(N)}\right\|^2 + \left|\mu(A_t)\right|^{-1}\mathrm{Tr}^2\left(P_t^{(N)}\right)\left\|S_t\right\|^2\left\|X_t\right\|^2 \\
& + \frac{1}{N}\mathrm{Tr}\left(\widetilde{Q}_t\right) + n^2\mu\left(S_t\right)\mathrm{Tr}^2\left(P_t^{(N)}\right) \\
\triangleq& 2\alpha\left\|m_t^{(N)}\right\|^2 + \beta_t
\end{aligned}
$$

using Assumption 3 and the inequality

$$
2a^{\mathrm{T}}b \le \left|\mu(A_t)\right|\left\|a\right\|^2 + \left|\mu(A_t)\right|^{-1}\left\|b\right\|^2. \tag{B.1}
$$

It can be computed that, $\forall\ p \ge 1$,

$$
\begin{aligned}
\sup_{t\ge 0}\left\|\gamma_t\right\|_{2,p} \le& \sup_{t\ge 0}\left(\frac{1}{N}\mu\left(\widetilde{Q}_t\right) + \mu\left(S_t\right)\left\|\mathrm{Tr}\left(P_t^{(N)}\right)\right\|_{2,2p}^2\right) \\
\lesssim_{n,p}& \frac{1}{N}, \\
\sup_{t\ge 0}\left\|\beta_t\right\|_{2,p} \le& \sup_{t\ge 0}\left(\frac{1}{N}\mathrm{Tr}\left(\widetilde{Q}_t\right) + n^2\mu\left(S_t\right)\left\|\mathrm{Tr}\left(P_t^{(N)}\right)\right\|_{2,2p}^2\right. \\
& \left.\left|\mu(A_t)\right|^{-1}\left\|S_t\right\|^2\left\|\mathrm{Tr}\left(P_t^{(N)}\right)\right\|_{2,4p}^2\left\|X_t\right\|_{2,4p}^2\right) \\
\lesssim_{n,p}& \frac{1}{N},
\end{aligned}
$$

using Hölder's inequality and (22)–(24). Then by Assumption 2 and Lemma 7.1 in Del Moral and Tugaut (2018), we obtain (25).

Step 4: We shall prove (26). By (9) and (4), we know that

$$
\begin{aligned}
dX_t^i =& \left[\left(A_t - \frac{P_t^{(N)}S_t}{2}\right)X_t^i + P_t^{(N)}S_tX_t - \frac{P_t^{(N)}S_t}{2}m_t^{(N)}\right]dt \\
& + G_tdB_t^i + P_t^{(N)}H_t^{\mathrm{T}}R_t^{-1}dW_t.
\end{aligned}
$$

Then by Itô's lemma, we have $d\left\|X_t^i\right\|^2 = \mathcal{L}_t\left\|X_t^i\right\|^2 dt + d\mathcal{M}_{4,t}$, where $\mathcal{M}_{4,t}$ is a martingale. Similarly, we have

$$\begin{aligned}\frac{d}{dt}\langle\mathcal{M}_4\rangle_t =& 4\left(X_t^i\right)^{\mathrm{T}}\left(\widetilde{\mathrm{Q}}_t + P_t^{(N)}S_tP_t^{(N)}\right)X_t^i\\ \le& 4\left[\mu\left(\widetilde{\mathrm{Q}}_t\right) + \mu\left(S_t\right)\mathrm{Tr}^2\left(P_t^{(N)}\right)\right]\left\|X_t^i\right\|^2 \text{ and}\\ \mathcal{L}_t\left\|X_t^i\right\|^2 \le& \mu\left(A_t\right)\left\|X_t^i\right\|^2 + \mathrm{Tr}\left(\widetilde{\mathrm{Q}}_t\right) + \mu\left(S_t\right)\mathrm{Tr}^2\left(P_t^{(N)}\right)\\ &+ 2\left|\mu(A_t)\right|^{-1}\mathrm{Tr}^2\left(P_t^{(N)}\right)\left\|S_t\right\|^2\left(\left\|X_t\right\|^2 + \left\|m_t^{(N)}\right\|^2/4\right).\end{aligned}$$

using Assumption 3 and the inequality (B.1). Then following the similar procedure as in Step 3, we obtain (26) by Assumption 2 and Lemma 7.1 in Del Moral and Tugaut (2018). □

References

Bishop, Adrian N., & Del Moral, Pierre (2017). On the stability of Kalman–Bucy diffusion processes. *SIAM Journal on Control and Optimization*, *55*(6), 4015–4047.

Chen, Xiuqiong, Luo, Xue, Shi, Ji, & Yau, Stephen S.-T. (2022). General convergence result for continuous-discrete feedback particle filter. *International Journal of Control*, *95*(11), 2972–2986.

Chen, Xiuqiong, & Yau, Stephen S.-T. (2023). On the stability of linear feedback particle filter. *Asian Journal of Mathematics*, *27*(1), 95–120.

Del Moral, Pierre, & Tugaut, Julian (2018). On the stability and the uniform propagation of chaos properties of ensemble Kalman–Bucy filters. *The Annals of Applied Probability*, *28*(2), 790–850.

Gordon, N. J., Salmond, D. J., & Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F*, *140*(2), 107–113.

Jazwinski, Andrew H. (1970). *Stochastic processes and filtering theory*. New York and London: Academic Press.

Kalman, Rudolph E., & Bucy, Richard S. (1961). New results in linear filtering and prediction theory. *Journal of Basic Engineering*, *83*(1), 95–108.

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

Magree, Daniel, & Johnson, Eric N. (2016). Factored extended Kalman filter for monocular vision-aided inertial navigation. *Journal of Aerospace Information Systems*, *13*(12), 475–490.

Peyré, Gabriel, Cuturi, Marco, et al. (2019). Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, *11*(5–6), 355–607.

Rozovsky, Boris L., & Lototsky, Sergey V. (2018). *Stochastic evolution systems: linear theory and applications to non-linear filtering*: *vol. 89*, Springer.

Taghvaei, Amirhossein, & Mehta, Prashant G. (2016). An optimal transport formulation of the linear feedback particle filter. In *2016 American control conference* ACC, (pp. 3614–3619). IEEE.

Taghvaei, Amirhossein, & Mehta, Prashant G. (2018a). Error analysis for the linear feedback particle filter. In *2018 annual American control conference* ACC, (pp. 4261–4266). IEEE.

Taghvaei, Amirhossein, & Mehta, Prashant G. (2018b). Error analysis of the stochastic linear feedback particle filter. In *2018 IEEE conference on decision and control* CDC, (pp. 7194–7199). IEEE.

Taghvaei, Amirhossein, & Mehta, Prashant G. (2020). An optimal transport formulation of the ensemble Kalman filter. *IEEE Transactions on Automatic Control*, *66*(7), 3052–3067.

Tanizaki, Hisashi (1996). *Nonlinear filters: estimation and applications*: *vol. 400*, Springer Science & Business Media.

Yang, Tao, Laugesen, Richard S, Mehta, Prashant G, & Meyn, Sean P (2016). Multivariable feedback particle filter. *Automatica*, *71*, 10–23.

Yang, Tao, Mehta, Prashant G., & Meyn, Sean P. (2013). Feedback particle filter. *IEEE Transactions on Automatic Control*, *58*(10), 2465–2480.

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

Zakai, Moshe (1969). On the optimal filtering of diffusion processes. *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, *11*(3), 230–243.

Xiuqiong Chen received the B.S. degree in the School of Mathematical Sciences, Beihang University, Beijing, China, in 2014, and the Ph.D. degree in applied mathematics from the Department of Mathematical Sciences, Tsinghua University, Beijing, China in 2019. After her graduation, she was a Postdoctoral Scholar with Yau Mathematical Sciences Center, Tsinghua University, Beijing, China, from 2019 to 2021. She joined in Renmin University of China, Beijing, China, since 2021. She is currently an Assistant Professor with School of Mathematics, Renmin University of China. Her research interests include nonlinear filtering and deep learning.

Jiayi Kang received the B.S. degree from the college of mathematics, Sichuan University, Chengdu, China, in 2019. He is currently pursuing the Ph. D. degree in Applied Mathematics from the Department of Mathematical Sciences, Tsinghua University, Beijing, China. His research interests include machine learning, nonlinear filtering and bioinformatics.

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

He was a Member of the Institute of Advanced Study at Princeton from 1976–1977 and 1981–1982, and a Benjamin Pierce Assistant Professor at Harvard University during 1977–1980. During 1980–2011, he joined the Department of Mathematics, Statistics and Computer Science (MSCS), University of Illinois at Chicago (UIC). During 2005–2011, he became a joint Professor with the Department of Electrical and Computer Engineering at the MSCS, UIC. During 2011–2022, he joined Tsinghua University, Beijing, China, where he is a full-time professor in the Department of Mathematical Sciences. After that, he joined Beijing Institute of Mathematical Sciences and Applications (BIMSA), Beijing, China. His current research interests include nonlinear filtering, bioinformatics, complex algebraic geometry, CR geometry and singularities theory.

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