ON THE STABILITY OF LINEAR FEEDBACK PARTICLE FILTER*

XIUQIONG CHEN† AND STEPHEN S.-T. YAU‡

Abstract. In this paper, we study the stability of feedback particle filter (FPF) for linear filtering systems with Gaussian noises. We first provide some local contraction estimates of the exact linear FPF, whose conditional distribution is exactly the posterior distribution of the state as long as their initial values are equal. Then we study the convergence of the linear FPF formed by N particles, and prove that the mean squared errors between the actual moments (m_t, P_t) and their approximations $(m_t^{(N)}, P_t^{(N)})$ by FPF are of order $\mathcal{O}(1/N)$ and decay exponentially fast as time t goes to infinity.

Key words. Feedback particle filter; Kalman-Bucy filter; Linear system; Convergence

MSC codes. 93C05, 93D20, 93D23, 34A12

1. Introduction.

The feedback particle filter (FPF) proposed by Yang et al. in 2013 [21], whose feedback control law is obtained by minimizing the Kullback-Leibler divergence between the actual posterior and the common posterior of any particle, is motivated by the mean-filed game theory. In the mean-field limit ($N = \infty$), where N is the number of particles, the FPF is exact, i.e., the distribution of the particle and the posterior distribution of the state at any time are equal provided that they are equal at $t = 0$. For linear system, the optimal control law of the exact FPF is determined by the conditional mean and covariance of the state. In real computations, the linear FPF, with approximations of the control law in the evolution equation, is formed by N particles .

It is well known that, for linear system with Gaussian noises, Kalman and Bucy proposed the famous Kalman-Bucy filter (KBF) [11] , which provides the optimal solution. There are many works investigating the stability of KBF such as [2, 6], and the work [4] provides an excellent survey on this problem.

In KBF, we need to solve a differential Riccati equation, which is the computational bottleneck in simulations for high dimensional problems. Therefore, the filters combined with Monte Carlo techniques, such as FPF, are still very promising even for linear filtering problems, if they can avoid solving the differential Riccati equation. Another filter using Monte Carlo idea is the ensemble Kalman filter (EnKF). The EnKF was introduced by Evensen in 1994 [9], and it is often used to solve the high dimensional forecasting and data assimilation problems, which arised in atmosphere sciences [1], whether forecasting [3] and so on. The theoretical analysis of EnKF is very active and we refer the interested readers to the paper [8] and the references therein. Compared with EnKF, the simulation variance of FPF is less. And the comparison of EnKF and FPF can be found in [17].

In this paper, we focus on the FPF for linear filtering problems with Gaussian noises, and we just call it linear FPF. There are two kinds of linear FPF considered

*Submitted to the editors. Stephen S.-T. Yau is the corresponding author.

Funding: This work is supported by National Natural Science Foundation of China (NSFC) grant (12201631, 11961141005), Tsinghua University start-up fund, and Tsinghua University Education Foundation fund (042202008).

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

‡Department of Mathematical Sciences, Tsinghua University, Beijing, 100084, P. R. China, and Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, Beijing 101400, P. R. China. (yau@uic.edu).

in this work. The first one is the exact linear FPF $\overline{X}_t$ defined in (3.7), where the control law is optimal without any approximation. Obviously, $\overline{X}_t$ is the conditional Mckean-Vlasov diffusion process, whose solution cannot be easily obtained since the conditional mean and covariance of the state in the evolution equation (3.7) cannot be obtained. Therefore, we use Monte Carlo method and replace the conditional mean and covariance by their empirical approximations formed by N particles $\{X_t^i\}_{i=1}^N$, and this is the second kind of linear FPF defined in (3.8). In this work, we shall analyze the stability of linear FPF. Actually, there have been some works on this problem. We have provided the convergence of FPF for nonlinear filtering systems with continuous state and discrete observations in [7]. However, we have not discussed the relationship between the estimation error and time t. Besides, the stability of FPF w.r.t. the initial conditions has also not been investigated. Taghvaei and Mehta analyzed the errors of deterministic and stochastic linear FPF in [19] and [20]. Kang et al. gave the error analysis of linear FPF for linear filtering systems with correlated noises [12]. However, they all assumed that the dimension of the state is 1 in [12, 19, 20], while we consider the any dimensional state in this paper.

The contributions of this paper are listed as follows:

- For KBF, we give a local contraction estimate of the conditional mean m_t. We prove that, for m_t and $m_t^\star$, which start from two possible initial values (m_0, P_0) and $(m_0^\star, P_0^\star)$ and satisfy the KBF, the L^p-error between them will decay exponentially fast to 0 w.r.t. the initial error. Besides, the decay rate is determined by the logarithmic norm of $A - P_\infty S$, where P_∞ is the solution of algebraic Riccati equation (4.5), and this result is summarized in Theorem 4.4.
- For exact linear FPF (3.7), which is a conditional Mckean-Vlasov diffusion process, we estimate the p-th Wasserstein distance between the distributions of $\overline{X}_t$ and $\overline{X}_t^\star$, which starts from two initial values. Similarly, we prove that the error will decay exponentially fast to 0 w.r.t. the initial error and the decay rate is determined by the logarithmic norm of $A - P_\infty S/2$. This result is listed in Theorem 4.6.
- For linear FPF (3.8) formed by N particles, where the actual conditional mean m_t and covariance P_t are approximated by sample mean $m_t^{(N)}$ and covariance $P_t^{(N)}$, we analyze the mean squared errors between (m_t, P_t) and $(m_t^{(N)}, P_t^{(N)})$. More explicitly, we prove that the errors are of order $\mathcal{O}(1/N)$ and decay exponentially fast as time t goes to infinity. This result can be found in Theorem 4.9.

The organization of this paper is as follows. In section 2, we shall introduce some preliminary results to be used in the subsequent sections. In section 3, the KBF and FPF for linear Gaussian filtering systems will be introduced. Section 4, which is also the main section, is devoted to present the stability of FPF. In this section, we shall firstly give some contraction estimates of the exact linear FPF and then analyze the convergence of the linear FPF formed by N particles. In the final section, we will present a brief summary of this article and propose an avenue of future works.

2. Preliminary.

In this section, we shall present some preliminary knowledges, including notations and some results which will used in the subsequent contents.

2.1. Notations.

$\mathbb{S}_n$ represents the set of all $n \times n$ real symmetric matrices. $\mathbb{S}_n^+$ is the subset of $\mathbb{S}_n$ where the matrices are positive definite. For two matrix $A, B \in \mathbb{S}_n$, we write $A > B$ if $A - B$ is a positive definite matrix, and $A \geq B$

if $A - B$ is positive semidefinite. For a $(n \times n)$-matrix A, let $\mathrm{Spec}(A)$ be the set of all eigenvalues, and $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the minimal and the maximal eigenvalue of A, respectively.

Let $\|\cdot\|$ represent the Euclidean norm of the vectors on $\mathbb{R}^n$. The spectral norm (or 2-norm) of A is the largest singular value of A, i.e., $\|A\| := \sqrt{\lambda_{\max}(A^\top A)}$. The Frobenius matrix norm of a given $(n_1 \times n_2)$-matrix A is defined by

$$\|A\|_{\mathrm{F}}^2 = \mathrm{Tr}\left(A^\top A\right) \qquad \text{with the trace operator } \mathrm{Tr}(\cdot).$$

More properties of trace can be found in Appendix B.1, which will be frequently used in the following sections. For any $(n \times n)$-matrix A, these exist the following norm equivalence formulae:

$$\|A\|^2 = \lambda_{\max}(A^\top A) \leq \mathrm{Tr}\left(A^\top A\right) = \|A\|_{\mathrm{F}}^2 \leq n \|A\|^2 .$$

We define the logarithmic norm $\mu(A)$ of a $(n \times n)$-square matrix A by

$$\begin{aligned}\mu(A) &:= \inf\left\{\alpha : \forall x \in \mathbb{R}^{n\times 1}, x^\top A x \leqslant \alpha \|x\|^2\right\}\\ &= \lambda_{\max}\left(\left(A + A^\top\right)/2\right)\\ &= \inf\left\{\alpha : \forall\, t \geq 0, \|\exp(At)\| \leq \exp(\alpha t)\right\}.\end{aligned}$$

Besides, using definition, we also have

$$\mu(A+B) \leq \mu(A) + \mu(B). \tag{2.1}$$

It can be proved that [15]

$$\mu(A) \geqslant \varsigma(A) := \max\{\mathrm{Re}(\lambda) : \lambda \in \mathrm{Spec}(A)\}, \tag{2.2}$$

where $\mathrm{Re}(\lambda)$ stands for the real part of the eigenvalues λ. The parameter $\varsigma(A)$ is often called the spectral abscissa of A. We use $\mathcal{N}(m, P)$ to denote the Gaussian distribution with mean m and covariance P.

For $p \geq 1$, the p-th Wasserstein distance between two probability measure ν_1 and ν_2 on $\mathbb{R}^n$ is defined as follows:

$$\mathbb{W}_p(\nu_1, \nu_2) = \inf\left\{\mathbb{E}^{1/p}\left[\|Z_1 - Z_2\|^p\right]\right\},$$

where the infimum is taken over all joint distributions of the random variables Z_1 and Z_2 with marginals ν_1 and ν_2, respectively.

The state transition matrix associated with a smooth flow of $(r \times r)$-matrix $A : u \mapsto A_u$ is denoted by

$$\begin{aligned}&\mathcal{E}_{s,t}(A) = \exp\left[\oint_s^t A_u du\right]\\ &\qquad\Longleftrightarrow \frac{d}{dt}\mathcal{E}_{s,t}(A) = A_t \mathcal{E}_{s,t}(A) \quad \text{and} \quad \partial_s \mathcal{E}_{s,t}(A) = -\mathcal{E}_{s,t}(A) A_s\end{aligned}$$

for any $s \leq t$, with $\mathcal{E}_{s,s} = \mathrm{I}$, the identity matrix. When $s = 0$, we write $\mathcal{E}_t(A) := \mathcal{E}_{0,t}(A)$. Some semigroup estimates of the state transition matrices associated with a sum of drift-type matrices can be found in the following lemma.

LEMMA 2.1. *(Perturbation lemma, [8]). Let $A : u \mapsto A_u$ and $B : u \mapsto B_u$ be some smooth flows of $(n \times n)$-matrices. For any $s \leq t$, we have*

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

In addition, we have

$$\|\mathcal{E}_{s,t}(A+B)\| \leq \alpha_A \exp\left[-\beta_A(t-s) + \alpha_A \int_s^t \|B_u\| \, du\right]$$

as soon as

$$\forall \ 0 \leq s \leq t \quad \|\mathcal{E}_{s,t}(A)\| \leq \alpha_A \exp\left(-\beta_A(t-s)\right),$$

where α_A and β_A are some constants depending on A.

For time homogeneous matrices $A_t = A$, we have $\mathcal{E}_{s,t}(A) = e^{(t-s)A} = \mathcal{E}_{t-s}(A)$, and it is known that [8], when $\varsigma(A) < 0$, for any $\varepsilon \in (0, 1]$ and any $t \geq 0$, we have

$$e^{\varsigma(A)t} \leq \|\mathcal{E}_t(A)\| \leq \kappa(\varepsilon) e^{(1-\varepsilon)\varsigma(A)t}, \tag{2.3}$$

with some constants $\kappa(\varepsilon)$ only depending on the parameter ε.

The quadratic variation $\langle M\rangle$ of a n-column vector continuous martingale M is the $(n \times n)$ matrix $\langle M\rangle$ such that $MM^\top - \langle M\rangle$ is a martingale. Given a real-valued continuous martingale M_t with $M_0 = 0$, for any $p \geqslant 1$ and any time horizon $t \geqslant 0$, we have the following Burkholder-Davis-Gundy inequality [13]:

$$\mathbb{E}\left(\sup_{0\leqslant s\leqslant t} |M_s|^p\right)^{1/p} \leqslant 2\sqrt{2}\sqrt{p}\,\mathbb{E}\left(\langle M\rangle_t^{p/2}\right)^{1/p}. \tag{2.4}$$

2.2. A technical theorem. The following theorem is used to control the moments of Riccati-type stochastic differential equations uniformly w.r.t. the time horizon, and is frequently used in the proofs of the subsequent sections.

THEOREM 2.2 (Lemma 7.1 in [8]). *Let a_t be some stochastic processes adapted to some filtration $\mathcal{F}_t$ and taking values in some measurable state space $(E, \mathtt{E})$. Let ψ be some nonnegative measurable function on $(E, \mathtt{E})$ such that*

$$d\psi\left(a_t\right) = \mathcal{L}_t\psi\left(a_t\right) dt + d\mathcal{M}_t(\psi)$$

with a $\mathcal{F}_t$-martingale $\mathcal{M}_t(\psi)$ and some $\mathcal{F}_t$-adapted process $\mathcal{L}_t\psi\left(a_t\right)$.

I: *Assume that*

$$\mathcal{L}_t\psi\left(a_t\right) \leq 2\gamma\sqrt{\psi\left(a_t\right)} + 3\alpha\psi\left(a_t\right) - \beta\psi\left(a_t\right)^2 + r,$$
$$\frac{d}{dt}\langle\mathcal{M}(\psi)\rangle_t \leq \psi\left(a_t\right)\left(\tau_0 + \tau_1\psi\left(a_t\right) + \tau_2\psi\left(a_t\right)^2\right)$$

for some parameters $\alpha < 0$ and $\gamma, \beta, r, \tau_0, \tau_1, \tau_2 \geq 0$. In this situation, we have the uniform moment estimate

$$\sup_{t\geq 0} \mathbb{E}\left[\psi\left(a_t\right)^p\right] < \infty, \ \forall \ 1 \leq p < 1 + 2\min\left(\beta/\tau_2, |\alpha|/\tau_1\right)$$

with the convention $\beta/0 = \infty = |\alpha|/0$ when $\tau_2 = 0$ or when $\tau_1 = 0$.

II: *Assume that*

$$\mathcal{L}_t\psi(a_t) \le 2\tau_t(a_t)\sqrt{\psi(a_t)} + 2\alpha\psi(a_t) + \beta_t(a_t)$$
$$\frac{d}{dt}\langle \mathcal{M}(\psi)\rangle_t \le \psi(a_t)\gamma_t(a_t)$$

for some $\alpha < 0$ *and some nonnegative functions* $(\tau_t, \beta_t, \gamma_t)$ *s.t.*

$$\delta_{\tau,t}(p) := \mathbb{E}\left[\tau_t(a_t)^p\right]^{\frac{1}{p}} < \infty, \delta_{\beta,t}(p) := \mathbb{E}\left[\beta_t(a_t)^p\right]^{\frac{1}{p}} < \infty \text{ and}$$
$$\delta_{\gamma,t}(p) := \mathbb{E}\left[\gamma_t(a_t)^p\right]^{\frac{1}{p}} < \infty$$

for any $p \ge 1$*. In this situation, we have the estimate*

$$\begin{aligned}\mathbb{E}\left[\psi(a_t)^p\right]^{\frac{1}{p}} \le & e^{\alpha t}\mathbb{E}\left[\psi(a_0)^p\right]^{\frac{1}{p}} \\ & + \int_0^t e^{\alpha(t-s)}\left[\delta_{\tau,s}(2p)^2/|\alpha| + \delta_{\beta,s}(p) + (p-1)\delta_{\gamma,s}(p)/2\right] ds.\end{aligned}$$

Remark 2.3. For case I in Theorem 2.2, according to the proof of Lemma 7.1 in [8], we have the more explicit result:

$$\mathbb{E}\left[\psi(a_t)^p\right]^{\frac{1}{p}} \le g_{p,t} + \exp\left\{2\left(\alpha + \frac{(p-1)}{2}\tau_1\right)t\right\}\left(\mathbb{E}\left[\psi(a_0)^p\right]^{\frac{1}{p}} - g_{p,0}\right),$$

where $g_{p,t}$ is a function satisfying

$$g_{p,t} \ge 0, \forall\ t \ge 0, \text{ and } \sup_{t\ge 0} g_{p,t} < \infty.$$

Remark 2.4. From case II in Theorem 2.2, it can be concluded that, if

$$\sup_{t\ge 0}\max\{\delta_{\tau,t}(p), \delta_{\beta,t}(p), \delta_{\gamma,t}(p)\} \le c(p) < \infty,$$

for some constant $c(p)$ depending on p, then we have

$$\mathbb{E}\left[\psi(a_t)^p\right]^{\frac{1}{p}} \le e^{\alpha t}\mathbb{E}\left[\psi(a_0)^p\right]^{\frac{1}{p}} + \tilde{c}(p)\int_0^t e^{\alpha(t-s)} ds$$

for some $\tilde{c}(p) < \infty$. It follows that

$$\sup_{t\ge 0}\mathbb{E}\left[\psi(a_t)^p\right] < \infty,\ \forall\ p \ge 1.$$

3. Filtering algorithms. The time homogeneous linear-Gaussian filtering model considered here is of the following form:

$$\begin{cases} dX_t = AX_t dt + R_1^{1/2} dB_t, \\ dZ_t = HX_t dt + R_2^{1/2} dW_t, \end{cases} \tag{3.1}$$

where X_t is the n-dimensional state, Z_t is the m-dimensional observation, B_t and W_t are independent standard Brownian motions which are also independent of the initial state X_0, $R_1^{1/2}$ and $R_2^{1/2}$ are invertible symmetric matrices, and $Y_0 = 0$. Define the σ-algebra formed by the observations till to time t as $\mathscr{F}_t := \sigma\{Z_s :\ \ 0 \le s \le t\}$. Then the optimal estimate of X_t based on $\mathscr{F}_t$ in minimum mean squared error sense is $\mathbb{E}[X_t|\mathscr{F}_t]$, i.e., the conditional expectation of the state X_t based on $\mathscr{F}_t$ [10].

3.1. Kalman-Bucy filter. It is well known that, if the initial state X_0 is Gaussian, i.e., $X_0 \sim \mathcal{N}(m_0, P_0)$, then the conditional distribution of the state X_t conditioned on $\mathscr{F}_t$ is a Gaussian distribution with the mean m_t and covariance P_t defined as follows:

$$m_t := \mathbb{E}\left[X_t | \mathscr{F}_t\right], P_t := \text{Cov}\left[X_t | \mathscr{F}_t\right], \tag{3.2}$$

where the conditional covariance $\text{Cov}\left[X_t|\mathscr{F}_t\right] \triangleq \mathbb{E}\left[\left(X_t - m_t\right)\left(X_t - m_t\right)^\top \middle| \mathscr{F}_t\right]$. In addition, the evolution equations of m_t and P_t are given in the Kalman-Bucy filter [10]:

$$dm_t = Am_t dt + P_t H^\top R_2^{-1}\left(dZ_t - Hm_t dt\right), \tag{3.3}$$

$$\frac{dP_t}{dt} = \text{Ricc}(P_t), \tag{3.4}$$

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

$$\text{Ricc}(Q) := AQ + QA^\top - QSQ + R_1 \tag{3.5}$$

with

$$S := H^\top R_2^{-1} H. \tag{3.6}$$

3.2. Feedback particle filter. Although the KBF is optimal for linear Gaussian problem (3.8), FPF may provide a computationally efficient option for filtering problems with very large state dimension n, since we need to solve differential Riccati equation (3.4) in KBF.

The evolution equation of the exact linear FPF is as follows:

$$d\overline{X}_t = A\overline{X}_t dt + R_1^{1/2} d\overline{B}_t + \overline{P}_t H^\top R_2^{-1}\left(dZ_t - H\frac{\overline{X}_t + \overline{m}_t}{2}dt\right), \tag{3.7}$$

where $\overline{B}_t$ is an independent copy of B_t,

$$\overline{m}_t := \mathbb{E}\left[\overline{X}_t \middle| \mathscr{F}_t\right], \text{ and } \overline{P}_t := \text{Cov}\left[\overline{X}_t \middle| \mathscr{F}_t\right].$$

Apparently, $\overline{X}_t$ is a conditional Mckean-Vlasov diffusion process.

For diffusion process (3.7), the following lemma tells us it is exact.

LEMMA 3.1 ([18]). *Consider KF* (3.3)-(3.4) *and Mckean-Vlasov diffusion process* (3.7). *If* $\overline{m}_0 = m_0$, $\overline{P}_0 = P_0$, *then we have*

$$\overline{m}_t = m_t, \text{ and } \overline{P}_t = P_t$$

for $\forall\, t \geq 0$. *Furthermore, if* $p(\overline{X}_0) = p(X_0)$, *then we have*

$$p(\overline{X}_t | \mathscr{F}_t) = p(X_t | \mathscr{F}_t).$$

In simulations, $\overline{m}_t$ and $\overline{P}_t$ are approximated by the sample mean and covariance formed by N particles $\{X_t^i\}_{i=1}^N$, whose evolution equation is as follows:

$$dX_t^i = AX_t^i dt + R_1^{1/2} dB_t^i + P_t^{(N)} H^\top R_2^{-1}\left(dZ_t - H\frac{X_t^i + m_t^{(N)}}{2}\right), \; 1 \leq i \leq N, \tag{3.8}$$

where the sample mean $m_t^{(N)}$ and sample covariance $P_t^{(N)}$ are computed according to

$$
\begin{aligned}
m_t^{(N)} &:= \frac{1}{N}\sum_{i=1}^{N} X_t^i, \\
P_t^{(N)} &:= \frac{1}{N-1}\sum_{i=1}^{N} \left(X_t^i - m_t^{(N)}\right)\left(X_t^i - m_t^{(N)}\right)^\top,
\end{aligned} \tag{3.9}
$$

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

$$
\begin{aligned}
dm_t^{(N)} =& A m_t^{(N)} dt + R_1^{1/2} d\widetilde{B}_t^{(N)} + P_t^{(N)} H^\top R_2^{-1}\left(dZ_t - H m_t^{(N)}\right), \\
dP_t^{(N)} =& \operatorname{Ricc}(P_t^{(N)}) dt + dM_t + dM_t^\top,
\end{aligned} \tag{3.10}
$$

where $\widetilde{B}_t^{(N)} := \frac{1}{N}\sum_{i=1}^N B_t^i$ *and* $dM_t = \frac{1}{N-1}\sum_{i=1}^N R_1^{1/2}\left(dB_t^i\right)(\vartheta_t^i)^\top$ *with* $\vartheta_t^i := X_t^i - m_t^{(N)}$.

The proof can be found in appendix A.1.

4. Stability analysis.

In this part, firstly, the Lipschitz property of the Mckean-Vlasov diffusion process $\overline{X}_t$ will be discussed and the contraction estimate of the conditional mean m_t is also provided. Then, we shall investigate the convergence of the FPF formed by N particles.

We first need to make two assumptions w.r.t. the linear system (3.1).

Assumption 1. *A in system* (3.1) *satisfies*

$$
\mu(A) < 0 \tag{4.1}
$$

From (2.2), it is known that, under this assumption, A is Hurwitz. In other words, Assumption 1 makes sure that the linear system (3.1) is stable.

Assumption 2. *S defined in* (3.6) *is a scalar matrix, i.e.,*

$$
S = \rho(S) I, \ \text{for some } \rho(S) > 0, \tag{4.2}
$$

where $I \in \mathbb{R}^{n\times n}$ is the identity matrix.

We shall discuss this assumption w.r.t. the observation after we use it in the subsequent contents.

Since all our results are based on the contraction estimates of Riccati semigroups and related fundamental matrices, we need to introduce the stability of KBF first.

4.1. Stability of Kalman-Bucy filter.

We first need to make the standard assumption w.r.t. the linear system (3.1).

Assumption 3. *We assume that for linear Gaussian system* (3.1), $\left(A, R_1^{1/2}\right)$ *is a controllable pair and (A, H) is observable, that is the matrices*

$$
\left[R_1^{1/2}, A\left(R_1^{1/2}\right), \ldots, A^{n-1}R_1^{1/2}\right] \ \text{and} \ \begin{bmatrix} H \\ HA \\ \vdots \\ HA^{n-1} \end{bmatrix} \tag{4.3}
$$

have rank n.

Under Assumption 3, there exist some parameters $v, \varpi_{\pm} > 0$ such that

$$\varpi_{-}\mathrm{I} \leq \int_0^v e^{As} R_1 e^{A^\top s} ds \leq \varpi_{+}\mathrm{I} \quad \text{and}$$
$$\varpi_{-}\mathrm{I} \leq \int_0^v e^{-A^\top s} S e^{-As} ds \leq \varpi_{+}\mathrm{I}.$$

The parameter v is the so-called interval of observability-controllability. Then we have the following famous result for P_t.

THEOREM 4.1 (Bucy [5]). *Under Assumption 3, for any $t \geq s \geq v$, we have the uniform estimates*

$$\sup_{P_0 \in \mathbb{S}_n^+} \left\| \exp\left[\oint_s^t (A - P_u S)\, du \right] \right\| \leq \alpha_v \exp\{-\beta_v (t - s)\}$$

for some parameters $\alpha_v < \infty$ and $\beta_v > 0$. In addition, for any $t \geq 0$ we have

$$\|P_t - P_t^\star\| \leq \alpha_v (P_0, P_0^\star) \exp\{-2\beta_v t\} \|P_0 - P_0^\star\| \tag{4.4}$$

for some constant $\alpha_v (P_0, P_0^\star)$ whose values only depend on $(v, P_0, P_0^\star)$.

Let P_∞ be the solution of the following algebraic Riccati equation:

$$\mathrm{Ricc}(P_\infty) = AP_\infty + P_\infty A^\top - P_\infty S P_\infty + R_1 = 0. \tag{4.5}$$

The existence and uniqueness of $P_\infty \in \mathbb{S}^+$ is ensured by Assumption 3 [14]. And this unique fixed point is called the steady state error covariance matrix. According to Bucy's theorem, if Assumption 3 holds, then P_t converges exponentially fast to P_∞ as $t \uparrow \infty$. In addition, the matrix difference $A - P_\infty S$ is asymptotically stable even when the signal drift matrix A is unstable. Taking advantage of (4.4), we can easily get

$$\sup_{t \geq 0} \|P_t\| \leq \|P_\infty\| + \alpha_v (P_0, P_\infty) \|P_0 - P_\infty\|. \tag{4.6}$$

Combing Theorem 4.1 and Lemma 2.1, we have the following result.

COROLLARY 4.2 (Corollary 5.7, [8]). *Under Assumption 3, for any $\varepsilon \in (0, 1]$, any $P_0 \in \mathbb{S}_n^+$, and any $s \leq t$ we have the exponential semigroup estimates*

$$\left\| \exp\left[\oint_s^t (A - P_u S)\, du \right] \right\| \leq \bar{\kappa}_{\varepsilon,\varsigma} (P_0, v) \exp((1 - \varepsilon)\zeta(A - P_\infty S)(t - s)) \tag{4.7}$$

and

$$\left\| \exp\left[\oint_s^t (A - P_u S)\, du \right] \right\| \leq \bar{\kappa}_\mu (P_0, v) \exp(\mu(A - P_\infty S)(t - s)). \tag{4.8}$$

In the above displayed formulae, the finite constants $\bar{\kappa}_\mu (P_0, v)$ and $\bar{\kappa}_{\varepsilon,\varsigma} (P_0, v)$ defined by

$$\begin{aligned} \log \bar{\kappa}_\mu (P_0, v) &= \kappa(\varepsilon)^{-1} \log [\bar{\kappa}_{\varepsilon,\varsigma} (P_0, v) / \kappa(\varepsilon)] \\ &= \|P_0 - P_\infty\| \|S\| \alpha_v (P_0, P_\infty) / (2\beta_v), \end{aligned} \tag{4.9}$$

with the parameters $(\kappa(\varepsilon), \alpha_v (P_0, P_\infty), \beta_v)$ presented in (2.3) and (4.4).

Let $(P_t, P_t^\star)$ be a couple of solutions of the Riccati equation (3.4) starting at two possibly different values $(P_0, P_0^\star)$, and $(m_t, m_t^\star)$ be a couple of solutions of the equation (3.3) starting at two possibly different values $(m_0, m_0^\star)$. Besides, let $\left(\overline{X}_t, \overline{X}_t^\star\right)$ be a couple of linear FPF diffusions (3.7) starting from two random states $\left(\overline{X}_0, \overline{X}_0^\star\right)$ with covariances matrices $(P_0, P_0^\star)$ and mean vectors $(m_0, m_0^\star)$. We denote by $(\pi_t, \pi_t^\star)$ the distributions of $\left(\overline{X}_t, \overline{X}_t^\star\right)$. Here we choose the same observation process $\{Z_t, t \geq 0\}$ and the same noise $\{\overline{B}_t, t \geq 0\}$.

Since

$$\begin{aligned} &\frac{d}{dt}\left(P_t - P_t^\star\right) = \left(A - P_t^\star S\right)\left(P_t - P_t^\star\right) + \left(P_t - P_t^\star\right)\left(A - P_t S\right)^\top \\ &\Rightarrow \left(P_t - P_t^\star\right) = \exp\left(\oint_s^t \left(A - P_u^\star S\right) du\right)\left(P_s - P_s^\star\right)\left[\exp\left(\oint_s^t \left(A - P_u S\right) du\right)\right]^\top, \end{aligned}$$

we have the following result using Corollary 4.2.

COROLLARY 4.3 (Corollary 5.8, [8]). *Under Assumption 3, for any $\varepsilon \in (0,1]$, and any $t \geq 0$, we have the exponential semigroup estimates*

$$\|P_t - P_t^\star\|_2 \leq \bar\kappa_{\varepsilon,\varsigma}\left(P_0, v\right)\bar\kappa_{\varepsilon,\zeta}\left(P_0^\star, v\right)\exp(2(1-\varepsilon)\varsigma(A - P_\infty S)t)\,\|P_0 - P_0^\star\|_2$$

as well as

$$\|P_t - P_t^\star\|_2 \leq \bar\kappa_\mu\left(P_0, v\right)\bar\kappa_\mu\left(P_0^\star, v\right)\exp(2\mu(A - P_\infty S)t)\,\|P_0 - P_0^\star\|_2 \tag{4.10}$$

with functions $Q \mapsto \bar\kappa_\mu(Q, v)$ and $\bar\kappa_\varsigma(Q, v)$ defined in Corollary 4.2.

4.2. Contraction estimate. The first result is about a quantitative contraction estimate for the conditional mean m_t of the state X_t.

THEOREM 4.4. *We assume Assumption 3 holds, and also assume that $\mu(A - P_\infty S) < 0$ where P_∞ is the solution of algebraic Riccati equation (4.5), then we have the following estimate:*

$$\mathbb{E}\left[\|m_t - m_t^\star\|^p\right]^{\frac{1}{p}} \leqslant \exp\left(\mu\left(A - P_\infty S\right)t\right)\left[C_1\|m_0 - m_0^\star\| + C_2\|P_0 - P_0^\star\|\right] \tag{4.11}$$

for $\forall\ p \geq 1$, where the parameters $C_1 := \bar\kappa_\mu\left(P_0, v\right)$, and $C_2 := \bar\kappa_\mu^2\left(P_0, v\right)\bar\kappa_\mu\left(P_0^\star, v\right)$ $\left(-\|S\|\sup_{t\geq 0}\left[\|X_t - m_t^\star\|^p\right]^{\frac{1}{p}}/\mu\left(A - P_\infty S\right) + 4\sqrt{pn\|S\|}/\sqrt{-2\mu\left(A - P_\infty S\right)}\right)$, with $\bar\kappa_\mu$ defined in (4.9).

Before we give the proof, we need the following lemma, whose proof can be found in Appendix A.2.

LEMMA 4.5. *Under Assumption 3, and also assume that there exists a positive semidefinite fixed point P_∞ of the algebraic Riccati equation (4.5). For any $p \geq 1$,*

- *if $\mu(A - P_\infty S) < 0$, then*

$$\sup_{t\geq 0}\mathbb{E}\left[\|X_t - m_t\|^p\right] < \infty, \tag{4.12}$$

- *if $\mu(A - P_\infty S/2) < 0$, then*

$$\sup_{t\geq 0}\mathbb{E}\left[\left\|X_t - \overline{X}_t\right\|^p\right] < \infty. \tag{4.13}$$

Now we give the proof of Theorem 4.4.

Proof. Define

$$e_t := m_t - m_t^\star, \quad \bar{e}_t := X_t - m_t^\star$$
$$Q_t := P_t - P_t^\star, \quad \alpha := \mu\left(A - P_\infty S\right)$$

According to (3.3), we know

$$\begin{aligned} de_t &= \left(A - P_t S\right) e_t dt + Q_t S \bar{e}_t dt + Q_t H^\top R_2^{-1/2} dW_t \\ &\triangleq \left(A - P_t S\right) e_t dt + Q_t S \bar{e}_t dt + dM_t \end{aligned}$$

with

$$\frac{d}{dt}\langle M \rangle_t = Q_t S Q_t.$$

Let $\Psi_{s,t} := \exp\left(\oint_s^t \left(A - P_u S\right) du\right)$. Then we know that

$$e_t = \Psi_{0,t} e_0 + \int_0^t \Psi_{s,t} Q_s S \bar{e}_s ds + \int_0^t \Psi_{s,t} dM_s. \tag{4.14}$$

Now we shall analyze the three terms of RHS of (4.14) individually.

Step I: By (4.8), we have

$$\mathbb{E}\left[\|\Psi_{0,t} e_0\|^p\right]^{\frac{1}{p}} \leqslant \|\Psi_{0,t}\| \, \mathbb{E}\left[\|e_0\|^p\right]^{\frac{1}{p}} \leqslant C_{11} \exp(\alpha t) \mathbb{E}\left[\|e_0\|^p\right]^{\frac{1}{p}}, \tag{4.15}$$

with $C_{11} := \bar{\kappa}_\mu\left(P_0, v\right)$.

Step II: Using (4.8) and (4.10), we have

$$\begin{aligned} &\mathbb{E}\left[\left\|\int_0^t \Psi_{s,t} Q_s S \bar{e}_s ds\right\|^p\right]^{\frac{1}{p}} \leqslant \int_0^t \|\Psi_{s,t}\| \, \|Q_s\| \, \|S\| \mathbb{E}\left[\|\bar{e}_s\|^p\right]^{1/p} ds \\ \leqslant & C_{21} \int_0^t e^{\alpha(t-s)+2\alpha s} ds \, \|Q_0\| \leqslant -C_{21}/\alpha e^{\alpha t} \|Q_0\|, \end{aligned} \tag{4.16}$$

where we use generalized Minkowski inequality in the first inequality, and $C_{21} := \bar{\kappa}_\mu^2\left(P_0, v\right) \bar{\kappa}_\mu\left(P_0^\star, v\right) \|S\| \sup_{t \geq 0}\left[\|X_t - m_t^\star\|^p\right]^{\frac{1}{p}}$.

Step III: Define

$$\gamma_t := \int_0^t \Psi_{s,t} dM_s.$$

Then we have

$$\begin{aligned} &\mathbb{E}\left[\|\gamma_t\|^{2p}\right]^{\frac{1}{p}} = \mathbb{E}\left[\left(\sum_{k=1}^n \left(\gamma_t(k)\right)^2\right)^p\right]^{\frac{1}{p}} \\ \leqslant & \sum_{k=1}^n \mathbb{E}\left[\left(\gamma_t(k)\right)^{2p}\right]^{\frac{1}{p}} = \sum_{k=1}^n \mathbb{E}\left[\left(\sum_{l=1}^n \int_0^t \Psi_{s,t}(k,l) dM_s(l)\right)^{2p}\right]^{\frac{1}{p}}, \end{aligned}$$

where $a(k)$ denotes the k-th entry of vector a and $A(k,l)$ denotes the (k,l)-th entry of matrix A.

Using the Burkholder-Davis-Gundy inequality (2.4), we have

$$
\begin{aligned}
&\mathbb{E}\left[\left(\sum_{l=1}^{n}\int_0^t \Psi_{s,t}(k,l)dM_s(l)\right)^{2p}\right]^{\frac{1}{p}} \\
\leqslant & 16p\mathbb{E}\left[\left(\sum_{l,l'=1}^{n}\int_0^t \Psi_{s,t}(k,l)\Psi_{s,t}\left(k,l'\right) d\left\langle M(l), M\left(l'\right)\right\rangle_s\right)^{p}\right]^{\frac{1}{p}} \\
=&16p\sum_{l,l'=1}^{n}\int_0^t \Psi_{s,t}(k,l)\Psi_{s,t}\left(k,l'\right)\left(Q_s S Q_s\right)\left(l,l'\right) ds \\
=&16p\int_0^t \left(\Psi_{s,t}Q_s S Q_s \Psi_{s,t}^{\top}\right)(k,k)ds.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\mathbb{E}\left[\|\gamma_t\|^{2p}\right]^{\frac{1}{p}} \leqslant 16p\int_0^t \operatorname{Tr}\left(\Psi_{s,t}Q_s S Q_s \Psi_{s,t}^{\top}\right) ds \leqslant 16p\int_0^t \|Q_s S Q_s\|\left\|\Psi_{s,t}\right\|_F^2 ds \\
\leqslant & 16pn\|S\|\int_0^t \|Q_s\|^2\left\|\Psi_{s,t}\right\|^2 ds \\
\leqslant & 16pn\|S\|\bar{\kappa}_\mu^4\left(P_0,v\right)\bar{\kappa}_\mu^2\left(P_0^\star,v\right)\int_0^t \exp(\alpha(4s+2(t-s)))ds\,\|Q_0\|^2 .
\end{aligned}
$$

It follows that

$$
\mathbb{E}\left[\|\gamma_t\|^{2p}\right]^{\frac{1}{2p}} \leqslant C_{31}\exp(\alpha t)\|Q_0\|
$$

with $C_{31} := 4\sqrt{pn\|S\|}\bar{\kappa}_\mu^2\left(P_0,v\right)\bar{\kappa}_\mu\left(P_0^\star,v\right)/\sqrt{-2\alpha}$.

And we also have

$$
\mathbb{E}\left[\|\gamma_t\|^{p}\right]^{\frac{1}{p}} \leqslant C_{31}\exp(\alpha t)\|Q_0\| \tag{4.17}
$$

□

since $\mathbb{E}\left[\|\gamma_t\|^{2p-1}\right]^{\frac{1}{2p-1}} \leqslant E\left[\|\gamma_t\|^{2p}\right]^{\frac{1}{2p}}$, for $\forall\ p \geq 1$.

Combining (4.15), (4.16) and (4.17), we obtain the desired result.

The second result provides a estimate of the Wasserstein distance between π_t and $\pi_t^\star$, which are the distributions of $\overline{X}_t$ and $\overline{X}_t^\star$, respectively.

THEOREM 4.6. *When Assumption 1, 2 and 3 hold, for $\forall\ p \geq 1$, we have the following result:*

$$
\begin{aligned}
\mathbb{W}_p(\pi_t,\pi_t^\star) \leq & \exp\left[\mu\left(A - P_\infty S/2\right)t\right] \\
& \cdot\left(C_1\mathbb{W}_p(\pi_0,\pi_0^\star) + C_2\left\|m_0 - m_0^\star\right\| + C_3\left\|P_0 - P_0^\star\right\|\right),
\end{aligned} \tag{4.18}
$$

where

$$
\begin{aligned}
C_1 :=& \exp\left[\alpha_v\left(P_0^\star,P_\infty\right)\|S\|\left\|P_\infty - P_0^\star\right\|/\left(2\beta_v\right)\right], \\
C_2 :=& C_1\left(\|P_\infty\| + \alpha_v\left(P_0^\star,P_\infty\right)\left\|P_0^\star - P_\infty\right\|\right)\|S\|\bar{\kappa}_\mu\left(P_0,v\right)/(2(\mu_2-\mu_1)),
\end{aligned} \tag{4.19}
$$

and

$$
\begin{aligned}
C_3 :=& \bar{\kappa}_\mu\left(P_0,v\right)\bar{\kappa}_\mu\left(P_0^\star,v\right)\Big\{C_1\|S\|/\left(-2\mu_1\right) + C_1\left(\|P_\infty\| + \alpha_v\left(P_0^\star,P_\infty\right)\left\|P_0^\star - P_\infty\right\|\right) \\
& \cdot\|S\|\bar{\kappa}_\mu\left(P_0,v\right)\left(-\|S\|\sup_{t\geq 0}\left[\|X_t - m_t^\star\|^p\right]^{\frac{1}{p}}/\mu_1 + 4\sqrt{pn\|S\|}/\sqrt{-2\mu_1}\right)/\left(-2(\mu_2-\mu_1)\right) \\
& +4\sqrt{pn\|S\|}C_1/\sqrt{-2\mu_2}\Big\},
\end{aligned}
$$

with the parameters $(\bar{\kappa}, \alpha_v(P_0, P_\infty), \beta_v)$ *presented in* (4.9) *and* (4.4), $\mu_1 := \mu(A - P_\infty S)$, *and* $\mu_2 := \mu(A - P_\infty S/2)$.

Since $\left(\overline{X}_t, \overline{X}_t^\star\right)$ are a couple of linear FPF diffusions (3.7) starting from two random states $\left(\overline{X}_0, \overline{X}_0^\star\right)$ with covariances matrices $(P_0, P_0^\star)$ and mean vectors $(m_0, m_0^\star)$, and $(\pi_t, \pi_t^\star)$ are the distributions of $\left(\overline{X}_t, \overline{X}_t^\star\right)$, it is known that the difference of $\overline{X}_t$ and $\overline{X}_t^\star$ are caused by the initial error between $\overline{X}_0$ and $\overline{X}_0^\star$ which are measured by three terms $\mathbb{W}_p(\pi_0, \pi_0^\star)$, $C_2\|m_0 - m_0^\star\|$ and $\|P_0 - P_0^\star\|$. Theorem 4.6 tells us that the error $\mathbb{W}_p(\pi_t, \pi_t^\star)$ between π_t and $\pi_t^\star$ decays exponentially fast w.r.t. time t and the decay rate is $-\mu(A - P_\infty S/2)$.

Before we give the proof, we need to estimate the semigroup $\exp\left[\oint_s^t (A - P_u S/2)\, du\right]$, which is shown in the following Corollary, and the proof is given in Appendix A.3.

Corollary 4.7. *Under Assumption 3, for any* $\varepsilon \in (0,1]$, *any* $P_0 \in \mathbb{S}_n^+$, *and any* $s \le t$, *we have the exponential semigroup estimates*

$$\left\|\exp\left[\oint_s^t (A - P_u S/2)\, du\right]\right\| \le C_1(\varepsilon, P_0, v) \exp\left((1-\varepsilon)\zeta(A - P_\infty S/2)(t-s)\right) \tag{4.20}$$

and

$$\left\|\exp\left[\oint_s^t (A - P_u S/2)\, du\right]\right\| \le C_2(\varepsilon, P_0, v) \exp(\mu(A - P_\infty S/2)(t-s)). \tag{4.21}$$

where $C_1(\varepsilon, P_0, v) := \kappa(\varepsilon)\exp\left[\kappa(\varepsilon)\alpha_v(P_0, P_\infty)\|S\|\|P_\infty - P_0\|/(2\beta_v)\right]$, *and* $C_2(\varepsilon, P_0, v) := \exp\left[\alpha_v(P_0, P_\infty)\|S\|\|P_\infty - P_0\|/(2\beta_v)\right]$, *with the parameters* $(\kappa(\varepsilon), \alpha_v(P_0, P_\infty), \beta_v)$ *presented in* (2.3) *and* (4.4).

Now we can start the proof of Theorem 4.6.

Proof. Firstly, we define

$$\begin{aligned} &e_t := \overline{X}_t - \overline{X}_t^\star, Q_t := P_t - P_t^\star \\ &e_{1,t} := m_t - m_t^\star, e_{2,t} := X_t - \overline{X}_t, e_{3,t} := X_t - m_t, \\ &\mu_1 := \mu(A - P_\infty S), \mu_2 := \mu(A - P_\infty S/2). \end{aligned}$$

Under Assumption 1, 2 and 3, we know that

$$\begin{aligned} \mu_1 &\le \mu(A) + \mu(-\rho(S)P_\infty) < \mu(A) < 0, \\ \mu_2 &\le \mu(A) + \mu(-\rho(S)P_\infty)/2 < \mu(A) < 0, \end{aligned}$$

and

$$\mu_1 = \mu(A - P_\infty S/2 - P_\infty S/2) \leqslant \mu_2 + \mu(-\rho(S)P_\infty)/2 < \mu_2. \tag{4.22}$$

By (3.7), we have

$$\begin{aligned} de_t =& Ae_t dt + Q_t H^\top R_2^{-1}\left(HX_t dt + R_2^{1/2} dW_t - H\frac{\overline{X}_t + m_t}{2} dt\right) \\ &+ P_t^\star H^\top R_2^{-1}(-H(e_t + e_{1,t})/2)\, dt \\ =& \left[\left(A - \frac{1}{2}P_t^\star S\right)e_t + Q_t S(e_{2,t} + e_{3,t})/2 - \frac{1}{2}P_t^\star S e_{1,t}\right] dt + Q_t H^\top R_2^{-1/2} dW_t. \end{aligned}$$

Let $\Phi_{s,t} := \exp\left[\oint_s^t \left(A - P_u^\star S/2\right) du\right], M_t := Q_t H^\top R_2^{-1/2} dW_t$, then we have

$$\frac{d}{dt}\langle M\rangle_t = Q_t S Q_t,$$

and

(4.23)
$$\begin{aligned} e_t =& \Phi_{0,t} e_0 + \frac{1}{2}\int_0^t \Phi_{s,t} Q_s S\left(e_{2,s} + e_{3,s}\right) ds - \frac{1}{2}\int_0^t \Phi_{s,t} P_s^\star S e_{1,s} ds + \int_0^t \Phi_{s,t} dM_s \\ \triangleq& I_1 + \frac{1}{2} I_2 - \frac{1}{2} I_3 + I_4. \end{aligned}$$

For I_1, using Corollary 4.7, we have

$$\mathbb{E}\left[\|I_1\|^p\right]^{\frac{1}{p}} \leqslant \|\Phi_{0,t}\| \,\mathbb{E}\left[\|e_0\|^p\right]^{\frac{1}{p}} \leqslant C_{11} e^{\mu_2 t} \mathbb{E}\left[\|e_0\|^p\right]^{\frac{1}{p}} \tag{4.24}$$

with $C_{11} := \exp\left[\alpha_v\left(P_0^\star, P_\infty\right)\|S\|\left\|P_\infty - P_0^\star\right\| / \left(2\beta_v\right)\right]$.

For I_2, Using generalized Minkowski inequality, we have

$$\begin{aligned} \mathbb{E}\left[\|I_2\|^p\right]^{\frac{1}{p}} &\leqslant \int_0^t \|\Phi_{s,t}\| \|Q_s\| \|S\| \left(\mathbb{E}\left[\|e_{2,s}\|^p\right]^{\frac{1}{p}} + \mathbb{E}\left[\|e_{3,s}\|^p\right]^{\frac{1}{p}}\right) ds \\ &\leqslant C_{21}\int_0^t e^{\mu_2(t-s)} e^{2\mu_1 s}\|Q_0\|\|S\| ds \\ &\leq C_{21}\|Q_0\|\|S\| e^{\mu_2 t} / \left(-\mu_2\right), \end{aligned} \tag{4.25}$$

where the second inequality is due to Lemma 4.5, Corollary 4.3 and Corollary 4.7, the third inequality comes from (4.22), and $C_{21} := C_{11}\bar{\kappa}_\mu\left(P_0, v\right)\bar{\kappa}_\mu\left(P_0^\star, v\right)\sup_{t\geq 0}\Big(\mathbb{E}\left[\|e_{2,s}\|^p\right]^{\frac{1}{p}}$ $+\mathbb{E}\left[\|e_{3,s}\|^p\right]^{\frac{1}{p}}\Big)$.

Similarly, for I_3, we have

$$\begin{aligned} \mathbb{E}\left[\|I_3\|^p\right]^{\frac{1}{p}} &\leqslant \int_0^t \|\Phi_{s,t}\| \|P_s^\star\| \|S\| \mathbb{E}\left[\|e_{1,s}\|^p\right]^{\frac{1}{p}} ds \\ &\leqslant C_{31}\int_0^t e^{\mu_2(t-s)} e^{\mu_1 s} ds\left(C_{32}\mathbb{E}\left[\|e_{1,0}\|^p\right]^{\frac{1}{p}} + C_{33}\|Q_0\|\right) \\ &\leqslant C_{31} e^{\mu_2 t}\left(C_{32}\mathbb{E}\left[\|e_{1,0}\|^p\right]^{\frac{1}{p}} + C_{33}\|Q_0\|\right) / \left(\mu_2 - \mu_1\right), \end{aligned} \tag{4.26}$$

where we use (4.6), Corollary 4.7 and Theorem 4.4 in the second inequality,

$$C_{31} := C_{11}\left(\|P_\infty\| + \alpha_v\left(P_0^\star, P_\infty\right)\|P_0^\star - P_\infty\|\right)\|S\|, C_{32} := \bar{\kappa}_\mu\left(P_0, v\right)$$

and

$$C_{33} := \bar{\kappa}_\mu^2\left(P_0, v\right)\bar{\kappa}_\mu\left(P_0^\star, v\right)\left(-\|S\|\sup_{t\geq 0}\left[\|X_t - m_t^\star\|^p\right]^{\frac{1}{p}} / \mu_1 + 4\sqrt{pn\|S\|}/\sqrt{-2\mu_1}\right).$$

For I_4, arguing as in the step III of the proof of Theorem 4.4, we can obtain that

$$\mathbb{E}\left[\|I_4\|^p\right]^{\frac{1}{p}} \leqslant C_{41} e^{\mu_2 t}\|Q_0\|, \tag{4.27}$$

where $C_{41} := 4\sqrt{pn\|S\|} C_{11}\bar{\kappa}_\mu\left(P_0, v\right)\bar{\kappa}_\mu\left(P_0^\star, v\right)/\sqrt{-2\mu_2}$.

Putting (4.24)-(4.27) into (4.23), we arrive at the desired result. ∎

Remark 4.8. From the proof, it an be seen that, in Theorem 4.6, Assumption 1 and 2 can be replaced by the weaker conditions $\mu\left(A - P_\infty S/2\right) < 0$, $\mu\left(A - P_\infty S\right) < 0$ and $\mu\left(A - P_\infty S\right) < \mu\left(A - P_\infty S/2\right)$.

4.3. Convergence of FPF (3.8) **.** In this part, we shall analyze the convergences of empirical mean and covariance in FPF (3.8), which are shown in the following theorem.

THEOREM 4.9. *If Assumption 1-3 hold, then for any* ϱ_1 *with*

$$\varrho_1 \in \left(4\mu(A) - 2\rho(S) \inf_{t\geq 0} \lambda_{\min}\left(P_t\right), 0\right), \tag{4.28}$$

there exists positive integer N_0*, such that for* $\forall\ N \geq N_0$*,* $\forall\ t \geq 0$*, we have that*

$$\begin{aligned} \mathbb{E}\left[\left\|m_t^{(N)} - m_t\right\|^2\right] \leq& \frac{c_3}{N} \exp\left\{\varrho_2 t\right\} + \frac{c_4}{N-1}, \\ \mathbb{E}\left[\left\|P_t^{(N)} - P_t\right\|_F^2\right] \leq& \frac{c_1}{N} \exp\left\{\varrho_1 t\right\} + \frac{c_2}{-\varrho_1(N-1)}, \end{aligned} \tag{4.29}$$

where $\varrho_2 = 2\mu(A)$*,* c_1, c_2, c_3 *and* c_4 *are some positive parameters independent of* N*.*

Before we give the proof, we list one result first and put its proof into Appendix A.4.

LEMMA 4.10. *Let* M_t *be defined in Lemma* 3.2*, then we have*

$$\begin{aligned} &\operatorname{Tr}\left\{d\left(M_t + M_t^\top\right) d\left(M_t + M_t^\top\right)\right\} \\ =&\frac{2}{N-1}\left[\operatorname{Tr}(R_1 P_t^{(N)}) + \operatorname{Tr}(R_1)\operatorname{Tr}(P_t^{(N)})\right] dt. \end{aligned} \tag{4.30}$$

Now we are ready for the proof of Theorem 4.9.

Proof: **Step I:** We first analyze the error between the $P_t^{(N)}$ formed by FPF and the actual conditional covariance P_t. Define the error matrix

$$\Xi_t := P_t^{(N)} - P_t,$$

then according to (3.4) and (3.10), we have the following evolution equation for Ξ_t:

$$\begin{aligned} d\Xi_t =& \left(\operatorname{Ricc}(P_t^{(N)}) - \operatorname{Ricc}(P_t)\right) dt + dM_t + dM_t^\top \\ =& \left[A - \frac{1}{2}\left(P_t^{(N)} + P_t\right) S\right] \Xi_t dt + \Xi_t \left[A - \frac{1}{2}\left(P_t^{(N)} + P_t\right) S\right]^\top dt + dM_t + dM_t^\top. \end{aligned} \tag{4.31}$$

Using Itô's lemma [13], we can easily obtain

$$d\left(\Xi_t^2\right) = I_1 dt + I_2 + I_3, \tag{4.32}$$

with

$$\begin{aligned} I_1 :=& \left[A - \frac{1}{2}\left(P_t^{(N)} + P_t\right) S\right] \Xi_t^2 + \Xi_t \left[A - \frac{1}{2}\left(P_t^{(N)} + P_t\right) S\right]^\top \Xi_t \\ &+ \Xi_t^2 \left[A - \frac{1}{2}\left(P_t^{(N)} + P_t\right) S\right]^\top + \Xi_t \left[A - \frac{1}{2}\left(P_t^{(N)} + P_t\right) S\right] \Xi_t, \end{aligned} \tag{4.33}$$

and

$$\begin{aligned} I_2 :=& \Xi_t d\left(M_t + M_t^\top\right) + d\left(M_t + M_t^\top\right) \Xi_t, \\ I_3 :=& d\left(M_t + M_t^\top\right) d\left(M_t + M_t^\top\right). \end{aligned} \tag{4.34}$$

Now we analyze these three terms separately.

Using Assumption 2 and Lemma B.2, we have

$$\begin{aligned}\mathbb{E}\left[\operatorname{Tr}(I_1)\right] =&2\mathbb{E}\left[\operatorname{Tr}\left\{\left[A + A^\top - \left(P_t^{(N)} + P_t\right) S\right] \Xi_t^2\right\}\right]\\ \le&2\lambda_{\max}\left(A + A^\top - \rho(S)P_t^{(N)} - \rho(S)P_t\right)\mathbb{E}\left[\operatorname{Tr}\left(\Xi_t^2\right)\right]\\ \le&2\left[2\mu(A) - \rho(S)\lambda_{\min}\left(P_t\right)\right]\mathbb{E}\left[\operatorname{Tr}\left(\Xi_t^2\right)\right],\end{aligned} \tag{4.35}$$

where the last inequality comes from the result that

$$\begin{aligned}&\lambda_{\max}\left(A + A^\top - \rho(S)P_t^{(N)} - \rho(S)P_t\right)\\ \le&\lambda_{\max}\left(A + A^\top - \rho(S)P_t^{(N)}\right) + \rho(S)\lambda_{\max}\left(-P_t\right)\\ \le&2\mu(A) - \rho(S)\lambda_{\min}\left(P_t\right).\end{aligned}$$

It can easily obtained that

$$\mathbb{E}[\operatorname{Tr}(I_2)] = 0 \tag{4.36}$$

since M_t is a martingale. As for I_3, using Lemma 4.10, we can have

$$\begin{aligned}&\mathbb{E}\left[\operatorname{Tr}\left\{d\left(M_t + M_t^\top\right)d\left(M_t + M_t^\top\right)\right\}\right]\\ =&\frac{2}{N-1}\mathbb{E}\left[\operatorname{Tr}(R_1P_t^{(N)}) + \operatorname{Tr}(R_1)\operatorname{Tr}(P_t^{(N)})\right]dt\\ =&\frac{2}{N-1}\mathbb{E}\left[\operatorname{Tr}(R_1P_t) + \operatorname{Tr}(R_1\Xi_t) + \operatorname{Tr}(R_1)\operatorname{Tr}(P_t) + \operatorname{Tr}(R_1)\operatorname{Tr}(\Xi_t)\right]dt\\ \le&\frac{2(n+1)\mu(P_t)\operatorname{Tr}(R_1) + \operatorname{Tr}(R_1^2) + n\operatorname{Tr}^2(R_1)}{N-1}dt\\ &+\frac{2}{N-1}\mathbb{E}\left[\operatorname{Tr}\left(\Xi_t^2\right)\right]dt,\end{aligned} \tag{4.37}$$

where the last inequality is due to the facts that

$$\begin{aligned}\operatorname{Tr}(R_1P_t) \le&\mu\left(P_t\right)\operatorname{Tr}(R_1),\\ \operatorname{Tr}(R_1)\operatorname{Tr}(P_t) \le&n\mu\left(P_t\right)\operatorname{Tr}(R_1),\end{aligned}$$

and

$$\begin{aligned}\operatorname{Tr}(R_1\Xi_t) + \operatorname{Tr}(R_1)\operatorname{Tr}(I\cdot\Xi_t) \le&\sqrt{\operatorname{Tr}(R_1^2)}\sqrt{\operatorname{Tr}(\Xi_t^2)} + \operatorname{Tr}(R_1)\sqrt{n}\sqrt{\operatorname{Tr}(\Xi_t^2)}\\ \le&\frac{\operatorname{Tr}(R_1^2) + \operatorname{Tr}\left(\Xi_t^2\right)}{2} + \frac{n\operatorname{Tr}^2(R_1) + \operatorname{Tr}\left(\Xi_t^2\right)}{2}.\end{aligned}$$

Put (4.35), (4.36) and (4.37) into (4.32), we can have

$$d\mathbb{E}\left[\|\Xi_t\|_{\mathrm{F}}^2\right] = d\mathbb{E}\left[\operatorname{Tr}\left(\Xi_t^2\right)\right] \le a_t\mathbb{E}\left[\|\Xi_t\|_{\mathrm{F}}^2\right]dt + \frac{b_t}{N-1}dt, \tag{4.38}$$

with

$$\begin{aligned}a_t :=&4\mu(A) - 2\rho(S)\lambda_{\min}\left(P_t\right) + \frac{2}{N-1},\\ b_t :=&2(n+1)\mu(P_t)\operatorname{Tr}(R_1) + \operatorname{Tr}(R_1^2) + n\operatorname{Tr}^2(R_1).\end{aligned}$$

It follows that, for any ϱ_1 satisfying (4.28), there exists a positive integer N_0, such that for $\forall\ N \geq N_0$, we have

$$\sup_{t\geq 0} a_t \leq \varrho_1 < 0.$$

Therefore, by Theorem B.3 and Grönwall's inequality, we can get that

$$\begin{aligned}\mathbb{E}\left[\|\Xi_t\|_{\mathrm{F}}^2\right] \leq & \mathbb{E}\left[\|\Xi_0\|_{\mathrm{F}}^2\right]\exp\left\{\int_0^t a_u du\right\} + \frac{1}{N-1}\int_0^t b_s \exp\left\{\int_s^t a_u du\right\} ds \\ \leq & \mathbb{E}\left[\|\Xi_0\|_{\mathrm{F}}^2\right]\exp\left\{\varrho_1 t\right\} + \frac{1}{N-1}\int_0^t b_s \exp\left\{\varrho_1 (t-s)\right\} ds \\ \leq & \mathbb{E}\left[\|\Xi_0\|_{\mathrm{F}}^2\right]\exp\left\{\varrho_1 t\right\} + \frac{\sup_{t\geq 0} b_t}{-\varrho_1 (N-1)}.\end{aligned} \tag{4.39}$$

Then we can easily obtain the second inequality in (4.29).

Step II: Next we analyze the error between the sample mean $m_t^{(N)}$ formed by FPF and the actual conditional mean m_t. Similarly, we define the error

$$e_t := m_t^{(N)} - m_t.$$

Using (3.3) and (3.10), we can have

$$de_t = \left(A - P_t^{(N)} S\right) e_t dt + \Xi_t H^\top R_2^{-1}\left(dZ_t - Hm_t dt\right) + R_1^{1/2} dB_t^{(N)}. \tag{4.40}$$

By Itô's lemma, we can get

$$\begin{aligned}& d\left(e_t e_t^\top\right) \\ =& \left(A - P_t^{(N)} S\right) e_t e_t^\top dt + e_t e_t^\top \left(A - P_t^{(N)} S\right)^\top dt + \Xi_t H^\top R_2^{-1}\left(dZ_t - Hm_t dt\right) e_t^\top \\ & + e_t \left(dZ_t - Hm_t dt\right)^\top R_2^{-1} H \Xi_t + R_1^{1/2} dB_t^{(N)} e_t^\top + e_t \left(dB_t^{(N)}\right)^\top R_1^{1/2} \\ & + \Xi_t H^\top R_2^{-1} R_2^{1/2} dW_t dW_t^\top R_2^{1/2} R_2^{-1} H \Xi_t + R_1^{1/2} dB_t^{(N)} \left(dB_t^{(N)}\right)^\top R_1^{1/2},\end{aligned} \tag{4.41}$$

from which we can conclude that

$$\begin{aligned}& d\mathbb{E}\left[\|e_t\|^2\right] = d\mathbb{E}\left[\mathrm{Tr}\left(e_t e_t^\top\right)\right] \\ =& \mathbb{E}\left[\mathrm{Tr}\left\{\left(A + A^\top - P_t^{(N)} S - S P_t^{(N)}\right) e_t e_t^\top\right\}\right] dt + \mathbb{E}\left[\mathrm{Tr}\left\{\Xi_t S \Xi_t\right\}\right] dt + \frac{1}{N}\mathrm{Tr}\left\{R_1\right\} dt \\ \leq & 2\mu(A)\mathbb{E}\left[\|e_t\|^2\right] dt + \rho(S)\mathbb{E}\left[\|\Xi_t\|_{\mathrm{F}}^2\right] dt + \frac{1}{N}\mathrm{Tr}\left\{R_1\right\} dt,\end{aligned} \tag{4.42}$$

since $Z_t - \int_0^t Hm_s ds$ is a martingale. Similar to (4.39), we have

$$\mathbb{E}\left[\|e_t\|^2\right] \leq \mathbb{E}\left[\|e_0\|^2\right]\exp\left\{2t\mu(A)\right\} + \frac{\rho(S)\sup_{t\geq 0}\mathbb{E}\left[\|\Xi_t\|_{\mathrm{F}}^2\right] + R_1/N}{-2\mu(A)},$$

and then we can obtain the first inequality of (4.29). □

Remark 4.11. From (4.35) in the proof, it can be seen that, we need to estimate the eigenvalue of the matrix $A + A^\top - \left(P_t^{(N)} + P_t\right) S$, which varies with N. And it is hard to make assumptions w.r.t. $A + A^\top - \left(P_t^{(N)} + P_t\right) S$. Using Assumption 2, we only need to estimate $\mu(A)$. Then take advantage of the stability assumption $\mu(A) < 0$, i.e., Assumption 1, we can easily obtain the desired result.

5. Conclusion. In this paper, we analyze the stability of linear FPF for time-invariant system (3.8). We give some local contraction estimates of the conditional mean and the exact linear FPF process w.r.t. the initial values. In addition, for linear FPF formed by N particles, we analyze the errors between actual moments (m_t, P_t) and their empirical approximations $(m_t^{(N)}, P_t^{(N)})$. However, all our discussions are restricted to the linear case. Therefore, how to analyze the long behavior of the general FPF is an important future work.

Appendix A. Proofs.

A.1. Proof of Lemma 3.2.

Proof. The evolution of $m_t^{(N)}$ can be obtained directly from (3.8). For the error process ϑ_t^i , we can get

$$d\vartheta_t^i = \left(A - \frac{P_t^{(N)} S}{2}\right) \vartheta_t^i + R_1^{1/2} \left(dB_t^i - d\widetilde{B}_t^{(N)}\right), \tag{A.1}$$

where S is defined in (3.6). Using Itô's lemma, we have

$$\begin{aligned}
dP_t^{(N)} =& \left(A - \frac{P_t^{(N)} S}{2}\right) P_t^{(N)} dt + P_t^{(N)} \left(A - \frac{P_t^{(N)} S}{2}\right)^\top dt \\
&+ \frac{1}{N-1} \sum_{i=1}^N \vartheta_t^i \left(dB_t^i - d\widetilde{B}_t^{(N)}\right)^\top R_1^{1/2} \\
&+ \frac{1}{N-1} \sum_{i=1}^N R_1^{1/2} \left(dB_t^i - d\widetilde{B}_t^{(N)}\right) (\vartheta_t^i)^\top \\
&+ \frac{1}{N-1} \sum_{i=1}^N R_1^{1/2} \left(dB_t^i - d\widetilde{B}_t^{(N)}\right) \left(dB_t^i - d\widetilde{B}_t^{(N)}\right)^\top R_1^{1/2} \\
=& \left(AP_t^{(N)} + P_t^{(N)} A^\top - P_t^{(N)} S P_t^{(N)}\right) dt \\
&+ \frac{1}{N-1} \sum_{i=1}^N \vartheta_t^i \left(dB_t^i\right)^\top R_1^{1/2} + \frac{1}{N-1} \sum_{i=1}^N R_1^{1/2} \left(dB_t^i\right) (\vartheta_t^i)^\top + R_1 dt \\
=& \operatorname{Ricc}(P_t^{(N)}) dt + dM_t + dM_t^\top,
\end{aligned}$$

since $\sum_{i=1}^N \vartheta_t^i = 0$. ◻

A.2. Proof of Lemma 4.5.

Proof. We shall prove the two inequalities by two steps.
Step I: We shall prove the first inequality (4.12). Define $e_t := X_t - m_t$, then

according to (3.1) and (3.3), we have

$$\begin{aligned} de_t =& Ae_t dt + R_1^{1/2} dB_t - P_t H^\top R_2^{-1} \left(He_t dt + R_2^{1/2} dW_t \right) \\ =& \left(A - P_t S\right) e_t dt + R_1^{1/2} dB_t - P_t H^\top R_2^{-1/2} dW_t. \end{aligned}$$

Using Itô's lemma, one has

$$d \left\| e_t \right\|^2 = \mathcal{L}_t \left\| e_t \right\|^2 dt + dM_{1,t},$$

where

$$\begin{aligned} \mathcal{L}_t \left\| e_t \right\|^2 :=& e_t^\top \left(A - P_t S + A^\top - S P_t \right) e_t + \operatorname{Tr}\left(R_1 + P_t S P_t \right) \\ \leqslant& 2\mu \left(A - P_t S \right) \left\| e_t \right\|^2 + \operatorname{Tr}\left(R_1 + P_t S P_t \right), \end{aligned}$$

and

$$dM_{1,t} := 2 e_t^\top R_1^{1/2} dB_t - 2 e_t^\top P_t H^\top R_2^{-1/2} dW_t$$

with

$$\frac{d}{dt} \left\langle M_{1,\cdot} \right\rangle_t = 4 e_t^\top R_1 e_t + 4 e_t^\top P_t S P_t e_t \leqslant 4 \left\| e_t \right\|^2 \operatorname{Tr}\left(R_1 + P_t S P_t \right).$$

According to Theorem 4.1, for any $u \in (0,1]$, there exists some time horizon $\tau_u \geqslant 0$, such that, for any $t \geqslant \tau_u$, we have

$$\sup_{t \geqslant \tau_u} \mu \left(A - P_t S \right) \leqslant (1-u) \mu \left(A - P_\infty S \right)$$

since

$$\begin{aligned} \mu \left(A - P_t S \right) &= \mu \left(A - P_\infty S + \left(P_\infty - P_t \right) S \right) \\ &\leqslant \mu(A - P_\infty S) + \mu \left(\left(P_\infty - P_t \right) S \right) \\ &\leqslant \mu(A - P_\infty S) + \mu \left(P_\infty - P_t \right) \operatorname{Tr}(S) \\ &\leqslant \mu(A - P_\infty S) + \sqrt{n} \left\| P_\infty - P_t \right\| \operatorname{Tr}(S) \end{aligned}$$

by (2.1) and (B.3). Therefore, for $\forall\ t \geqslant \tau_u$, we have

$$\mathcal{L}_t \left\| e_t \right\|^2 \leq 2(1-u) \mu(A - P_\infty S) \left\| e_t \right\|^2 + c(u)$$

and

$$\frac{d}{dt} \left\langle M_{1,\cdot} \right\rangle_t \leqslant c(u) \left\| e_t \right\|^2$$

where

$$c(u) := 4 \sup_{t \geqslant 0} \left(\operatorname{Tr}\left(R_1 \right) + n \mu(S) \left\| P_t \right\|^2 \right)$$

Then the desired result (4.12) can be concluded using Theorem 2.2.

Step II: We now prove the second inequality (4.13). Define $\bar{e}_t := X_t - \overline{X}_t$. According to (3.1) and (3.7), we have

$$\begin{aligned} d\bar{e}_t =& A\bar{e}_t + R_1^{1/2} \left(dB_t - d\overline{B}_t \right) - P_t H^\top R_2^{-1} \left(H \left(\bar{e}_t + e_t \right)/2 + R_2^{1/2} dW_t \right) \\ =& \left(A - P_t S/2 \right) \bar{e}_t - P_t S e_t/2 + R_1^{1/2} \left(dB_t - d\overline{B}_t \right) - P_t H^\top R_2^{-1/2} dW_t \end{aligned}$$

By Itô's lemma, we have

$$d \left\| \bar{e}_t \right\|^2 = \mathcal{L}_t \left\| \bar{e}_t \right\|^2 dt + dM_{2,t}$$

with

$$\begin{aligned}\mathcal{L}_t \|\bar{e}_t\|^2 :=& \bar{e}_t^\top \left(A - P_t S/2 + A^\top - SP_t/2\right) \bar{e}_t - \bar{e}_t^\top P_t S e_t + \operatorname{Tr}\left(2R_1 + P_t S P_t\right) \\ \leq& 2\mu\left(A - P_t S/2\right) \|\bar{e}_t\|^2 - \mu\left(A - P_\infty S/2\right) \|\bar{e}_t\|^2 \\ &+ \left|\mu\left(A - P_\infty S/2\right)\right|^{-1} \|P_t\|^2 \|S\|^2 \|e_t\|^2 /4 + 2\operatorname{Tr}\left(R_1\right) + n\mu(S) \|P_t\|^2 \\ \leq& \left(2\mu\left(A - P_t S/2\right) - \mu\left(A - P_\infty S/2\right)\right) \|\bar{e}_t\|^2 + C_1 \|e_t\|^2 + C_2\end{aligned}$$

where $C_1 := \sup_{t\geq 0} \left\{\left|\mu\left(A - P_\infty S/2\right)\right|^{-1} \|P_t\|^2 \|S\|^2/4\right\}$, $C_2 := \sup_{t\geq 0} \{2\operatorname{Tr}\left(R_1\right) + n\mu(S) \|P_t\|^2\Big\}$, and we use the inequality $|a^\top b| \leq \epsilon \|a\|^2 + \|b\|^2 /\epsilon, \forall\ \epsilon > 0$ in the first inequality. Besides, we also have

$$\frac{d}{dt} \langle M_{2,\cdot}\rangle_t = 4\bar{e}_t^\top \left(2R_1 + P_t S P_t\right) \bar{e}_t \leq C_3 \|\bar{e}_t\|^2,$$

where $C_3 := \sup_{t\geq 0} \left\{4\left(2\mu(R_1) + n\mu(S) \|P_t\|^2\right)\right\}$.

Using Remark 2.4 of Theorem 2.2 and the similar procedure in Step I, we can obtain (4.13). ∎

A.3. Proof of Lemma 4.7.

Proof. Let $\bar{A} := A - P_\infty S/2,\ \bar{B}_u := \left(P_\infty - P_u\right) S/2$. Then we have

$$A - P_u S/2 = \bar{A} + \bar{B}_u$$

and

$$\exp\left[\oint_s^t \left(A - P_u S/2\right) du\right] = \mathcal{E}_{s,t}(\bar{A} + \bar{B}).$$

By (2.3), we get

$$\left\|\mathcal{E}_{s,t}(\bar{A})\right\| = \left\|\mathcal{E}_{t-s}(\bar{A})\right\| \leqslant \kappa(\varepsilon) e^{(1-\varepsilon)\zeta(\bar{A})(t-s)},$$

from which and Lemma 2.1, we obtain

$$\|\mathcal{E}_{s,t}(\bar{A} + \bar{B})\| \leqslant \kappa(\varepsilon) \exp\left[(1-\varepsilon)\zeta(\bar{A})(t-s) + \kappa(\varepsilon) \int_s^t \left\|\bar{B}_u\right\| du\right].$$

Since

$$\begin{aligned}\int_s^t \left\|\bar{B}_u\right\| du =& \frac{1}{2} \int_s^t \left\|\left(P_\infty - P_u\right) S\right\| du \\ \leq& \frac{1}{2} \|S\| \int_s^t \|P_\infty - P_u\| du \\ \leq& \frac{1}{2} \|S\| \alpha_v\left(P_0, P_\infty\right) \|P_\infty - P_0\| \int_s^t \exp\left(-2\beta_v u\right) du \\ \leq& \alpha_v\left(P_0, P_\infty\right) \|S\| \|P_\infty - P_0\| / \left(2\beta_v\right)\end{aligned}$$

using Theorem 2, we can obtain (4.20). Following the similar procedure, we can also get (4.21).

A.4. Proof of Lemma 4.10.

Proof. Since $dM_t = \dfrac{1}{N-1}\sum_{i=1}^{N} R_1^{1/2}\left(dB_t^i\right)(\vartheta_t^i)^\top$, we have

$$\begin{aligned}
dM_t dM_t =& \frac{1}{(N-1)^2}\sum_{i,j=1}^{N} R_1^{1/2}\left(dB_t^i\right)(\vartheta_t^i)^\top R_1^{1/2}\left(dB_t^j\right)(\vartheta_t^j)^\top \\
=& \frac{1}{(N-1)^2}\sum_{i,j=1}^{N} R_1^{1/2}\left(dB_t^i\right)\left(dB_t^j\right)^\top R_1^{1/2}\vartheta_t^i(\vartheta_t^j)^\top \\
=& \frac{1}{(N-1)^2} R_1 \sum_{i}^{N} \vartheta_t^i(\vartheta_t^i)^\top dt \\
=& \frac{1}{N-1} R_1 P_t^{(N)} dt
\end{aligned}$$

and

$$\begin{aligned}
dM_t dM_t^\top =& \frac{1}{(N-1)^2}\sum_{i,j=1}^{N} R_1^{1/2}\left(dB_t^i\right)(\vartheta_t^i)^\top \vartheta_t^j\left(dB_t^j\right)^\top R_1^{1/2} \\
=& \frac{1}{(N-1)^2}\sum_{i,j=1}^{N} R_1^{1/2}\left(dB_t^i\right)\left(dB_t^j\right)^\top R_1^{1/2}(\vartheta_t^j)^\top \vartheta_t^i \\
=& \frac{1}{(N-1)^2} R_1 \sum_{i}^{N} (\vartheta_t^j)^\top \vartheta_t^i dt \\
=& \frac{1}{N-1} R_1 \operatorname{Tr}\left(P_t^{(N)}\right) dt.
\end{aligned}$$

It follows that

$$\begin{aligned}
\operatorname{Tr}\left\{d\left(M_t + M_t^\top\right) d\left(M_t + M_t^\top\right)\right\} =& 2\operatorname{Tr}\left\{dM_t dM_t + dM_t dM_t^\top\right\} \\
=& \frac{2}{N-1}\left[\operatorname{Tr}(R_1 P_t^{(N)}) + \operatorname{Tr}(R_1)\operatorname{Tr}(P_t^{(N)})\right] dt.
\end{aligned}$$ ∎

Appendix B. Some results.

B.1. Trace of a square matrix. We list some results about matrix trace here: for any $A, B \in \mathbb{R}^{n\times n}$, we have

$$\begin{aligned}
\operatorname{Tr}(A) &= \sum_{i=1}^{n} A(i,i), \\
\operatorname{Tr}(A) &= \operatorname{Tr}(A^\top), \\
\operatorname{Tr}(A) &= \operatorname{Tr}(PAP^{-1}), \\
\operatorname{Tr}(AB) &= \operatorname{Tr}(BA), \\
\|A\|_{\mathrm{F}}^2 &= \operatorname{Tr}(AA^\top),
\end{aligned} \tag{B.1}$$

where $A(i,i)$ is the (i,i)-th entry of matrix A, and $P \in \mathbb{R}^{n\times n}$ is any invertible matrix.

Lemma B.1. *For any $A, B \in \mathbb{R}^{n\times n}$, we have*

$$\left|\operatorname{Tr}(AB^\top)\right|^2 \le \operatorname{Tr}(AA^\top)\operatorname{Tr}(BB^\top). \tag{B.2}$$

Proof. Define $\langle A, B\rangle := \mathrm{Tr}(AB^\top)$. It can be easily checked that such defined $\langle\cdot,\cdot\rangle$ is a inner product on the Euclidean space $\mathbb{R}^{n\times n}$. Then by Cauchy-Schwarz inequality, we have

$\square$

$$|\langle A, B\rangle|^2 \le \langle A, A\rangle \cdot \langle B, B\rangle,$$

which is the desired result.

LEMMA B.2. *For any $A, B \in \mathbb{S}_n$, if B is also positive semidefinite, then we have*

$$\lambda_{\min}(A)\,\mathrm{Tr}(B) \le Tr(AB) \le \lambda_{\max}(A)\,\mathrm{Tr}(B). \tag{B.3}$$

and

$$\frac{1}{n}\left|\mathrm{Tr}(A)\right|^2 \le \mathrm{Tr}(A^2) \le \left|\mathrm{Tr}(A)\right|^2. \tag{B.4}$$

Proof. Since $A \in \mathbb{S}_n$, there exists an orthogonal matrix P_a such that

$$A = P_a\Lambda_a P_a^{-1},$$

with $\Lambda_a := \mathrm{diag}\{\lambda_1, \cdots, \lambda_n\}$ which is the diagonal matrix composed by all eigenvalues of A. Then we can get

$$\begin{aligned}\mathrm{Tr}(AB) &= \mathrm{Tr}(P_a\Lambda_a P_a^{-1}B)\\ &= \mathrm{Tr}(\Lambda_a P_a^{-1}BP_a)\\ &= \sum_{i=1}^{n}\lambda_i\left(P_a^{-1}BP_a\right)(i,i).\end{aligned}$$

Since B is positive semidefinite, we know that there exists P_b, such that $B = P_b^\top P_b$. Therefore

$$P_a^{-1}BP_a = (P_bP_a)^\top(P_bP_a),$$

and from which we can conclude

$$\left[P_a^{-1}BP_a\right]_{i,i} \ge 0.$$

It follows that

$$\lambda_{\min}(A)\,\mathrm{Tr}(P_a^{-1}BP_a) \le \mathrm{Tr}(AB) \le \lambda_{\max}(A)\,\mathrm{Tr}(P_a^{-1}BP_a).$$

Then we obtain the first desired result using the property $\mathrm{Tr}(B) = \mathrm{Tr}(P_a^{-1}BP_a)$.

On the one hand, when $B = I$ and by Lemma B.1, we have

$$\left|\mathrm{Tr}(A)\right|^2 \le n\,\mathrm{Tr}(A^2).$$

On the other hand,

$$\mathrm{Tr}(A^2) \le \lambda_{\max}(A)\,\mathrm{Tr}(A) \le \left|\mathrm{Tr}(A)\right|^2.$$

Then we obtain the second result. $\square$

B.2. Estimates of the initial errors $m_0^{(N)} - m_0$ and $P_0^{(N)} - P_0$. In this part, we shall consider how close are the sample mean and covariance matrix to the actual mean and covariance matrix of Gaussian distribution.

THEOREM B.3. *Let the n-dimensional random vectors $X_i \overset{i.i.d}{\sim} \mathcal{N}(m, P), i = 1, 2, \ldots, N$. Define*

$$m^{(N)} := \frac{1}{N} \sum_{i=1}^{N} X_i,$$

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

Then for $\forall\ p \geq 1$, we have

$$\mathbb{E}\left[\|m^{(N)} - m\|^p\right]^{\frac{1}{p}} \leq C_{n,p} \frac{1}{\sqrt{N}}. \tag{B.5}$$

and

$$\mathbb{E}\left[\|P^{(N)} - P\|_F^p\right]^{\frac{1}{p}} \leq \bar{C}_{n,p} \frac{1}{\sqrt{N}}, \tag{B.6}$$

where $C_{n,p}$ and $\bar{C}_{n,p}$ are some parameters depending on n and p.

Before we prove it, we need to state two results.

LEMMA B.4. *(Rosenthal type inequality [16]) Let $p > 0$, and let $\{\xi_i,\ i = 1, \cdots, N\}$ be conditionally independent random variables given $\sigma-$algebra $\mathcal{G}$ such that $\mathbb{E}[\xi_i|\mathcal{G}] = 0$ and $\mathbb{E}\left[|\xi_i|^p|\,\mathcal{G}\right] < \infty$. Then*

$$\mathbb{E}\left[\left|\sum_{i=1}^{N} \xi_i\right|^p \middle| \mathcal{G}\right] \leq C_p \left[\sum_{i=1}^{N} \mathbb{E}(|\xi_i|^p|\mathcal{G}) + \left(\sum_{i=1}^{N} \mathbb{E}(|\xi_i|^2|\mathcal{G})\right)^{p/2}\right], \tag{B.7}$$

where C_p is a constant that depends only on p. This inequality holds in the almost sure sense.

Using Lemma B.4, we have the following result for the estimate of the moment of the random vectors.

COROLLARY B.5. *Let $p \geq 1$, and let $\{\xi_i,\ i = 1, \cdots, N\}$ be independent identically distributed n-dimensional random vectors with $\mathbb{E}[\xi_i] = 0$ and $\mathbb{E}\left[\|\xi_i\|^{2p}\right] < \infty$. Then*

$$\mathbb{E}\left[\left\|\sum_{i=1}^{N} \xi_i\right\|^{2p}\right] \leq C_{n,p} N^p, \tag{B.8}$$

where $C_{n,p}$ is a finite parameter depending on n and p.

Proof. Let $\xi_{i,j}$ be the j-entry of vector ξ_i. It is obvious that for $\forall\ 1 \leq j \leq n$, $\{\xi_{i,j},\ i = 1, \cdots, N\}$ are independent identically distributed random variables with $\mathbb{E}[\xi_{i,j}] = 0$ and $\mathbb{E}\left[|\xi_{i,j}|^{2p}\right] < \infty$. Then using Jensen's inequality and Lemma B.4, we have

$$\mathbb{E}\left[\left\|\sum_{i=1}^{N}\xi_i\right\|^{2p}\right] = \mathbb{E}\left[\left(\sum_{j=1}^{n}\left(\sum_{i=1}^{N}\xi_{i,j}\right)^2\right)^p\right]$$

$$\leqslant n^{p-1}\sum_{j=1}^{n}\mathbb{E}\left[\left(\sum_{i=1}^{N}\xi_{i,j}\right)^{2p}\right]$$

$$\leqslant n^{p-1}C_p\sum_{j=1}^{n}\left[\sum_{i=1}^{N}\mathbb{E}\left[|\xi_{i,j}|^{2p}\right] + \left(\sum_{i=1}^{N}\mathbb{E}\left[|\xi_{i,j}|^2\right]\right)^p\right]$$

$$\leqslant n^{p-1}C_p\sum_{j=1}^{n}\left(N\mathbb{E}\left[|\xi_{i,j}|^{2p}\right] + N^p\mathbb{E}\left[|\xi_{i,j}|^{2p}\right]\right)$$

$$\leqslant C_{n,p}N^p.$$

□

Now we start the proof of Theorem B.3.

Proof. **Step 1:** We first prove (B.5).

Using Corollary B.5 and $X_i - m \overset{i.i.d}{\sim} \mathcal{N}(0_{n\times 1}, P), i = 1, 2, \ldots, N$, we have

$$\mathbb{E}\left[\|m^{(N)} - m\|^{2p}\right]^{\frac{1}{2p}} = \mathbb{E}\left[\left\|\frac{1}{N}\sum_{i=1}^{N}(X_i - m)\right\|^{2p}\right]^{\frac{1}{2p}}$$

$$= \mathbb{E}\left[\left\|\sum_{i=1}^{N}(X_i - m)\right\|^{2p}\right]^{\frac{1}{2p}}$$

$$\leq C_{n,p}\frac{1}{\sqrt{N}}.$$

Hence we obtain (B.5) by the inequality

$$\mathbb{E}\left[\|m^{(N)} - m\|^{2p-1}\right]^{\frac{1}{2p-1}} \leq \mathbb{E}\left[\|m^{(N)} - m\|^{2p}\right]^{\frac{1}{2p}}.$$

Step 2: Now we prove (B.6). Define $\tilde{e}_i = X_i - m^{(N)},\ 1 \leqslant i \leqslant N$, then we have $E\left[\tilde{e}_i\right] = 0$ and

$$\mathbb{E}\left[\tilde{e}_i\tilde{e}_j^\top\right] = \begin{cases} \left(1 - \frac{1}{N}\right)P & i = j, \\ -\frac{1}{N}P & i \neq j. \end{cases}$$

Define $\tilde{e} = (\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_N) \in \mathbb{R}^{n\times N}$, we have

$$\text{vec}(\tilde{e}) \sim \mathcal{N}\left(0_{nN\times 1}, B_N \otimes P\right),$$

where $\text{vec}(A)$ denotes the vectorization of the matrix A, $0_{n\times 1}$ denotes $n \times 1$ zero vector, and

$$B_N = \begin{bmatrix} 1 - \frac{1}{N} & -\frac{1}{N} & \cdots & -\frac{1}{N} \\ -\frac{1}{N} & 1 - \frac{1}{N} & \cdots & -\frac{1}{N} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{N} & -\frac{1}{N} & \cdots & 1 - \frac{1}{N} \end{bmatrix}.$$

It can be easily computed that

$$\mathrm{Spec}\,(B_N) = \{1, \cdots, 1, 0\}.$$

Since B_N is symmetric, there exists an orthogonal matrix A, s.t.

$$A^{-1} B_N A = \begin{bmatrix} I_{N-1} & \\ & 0 \end{bmatrix}.$$

Define

$$e = (e_1, \ldots, e_N) := \tilde{e} A,$$

and we have

$$\mathrm{vec}(e) = \left(A^\top \otimes I_n\right) \mathrm{vec}(\tilde{e}).$$

Therefore

$$\mathrm{vec}(e) \sim \mathcal{N}\left(0_{nN\times 1}, \begin{bmatrix} I_{N-1} \otimes P & \\ & 0 \end{bmatrix}\right)$$

since

$$\begin{aligned} \mathrm{Cov}(\mathrm{vec}(e)) &= \left(A^\top \otimes I_n\right)(B_N \otimes P)(A \otimes I_n) \\ &= \left(A^\top B_N A\right) \otimes P \\ &= \begin{bmatrix} I_{N-1} \otimes P & \\ & 0 \end{bmatrix}. \end{aligned}$$

It can be concluded that

$$e_i \sim \mathcal{N}(0_{n\times 1}, P), 1 \leqslant i \leqslant N-1, e_N = 0,$$

$\{e_1, \ldots, e_{N-1}\}$ are independent, and

$$P^{(N)} = \frac{1}{N-1} \tilde{e}\tilde{e}^\top = \frac{1}{N-1} e A^{-1} A e^\top = \frac{1}{N-1} e e^\top = \frac{1}{N-1} \sum_{i=1}^{N-1} e_i e_i^\top.$$

Let $Y_i = \mathrm{vec}\left(e_i e_i^\top - P\right),\ 1 \leqslant i \leqslant N-1$. It can be easily checked that $\{Y_1, \ldots, Y_{N-1}\}$ are independent and identically distributed, with $\mathbb{E}[Y_i] = 0$ and $\mathbb{E}\left[\|Y_i\|^{2p}\right] < \infty$. Then according to Corollary B.5, we obtain

$$\begin{aligned} \mathbb{E}\left[\|P^{(N)} - P\|_{\mathrm{F}}^{2p}\right]^{\frac{1}{2p}} &= \mathbb{E}\left[\|\mathrm{vec}(P^{(N)} - P)\|^{2p}\right]^{\frac{1}{2p}} \\ &= \mathbb{E}\left[\left\|\frac{1}{N-1}\sum_{i=1}^{N-1} Y_i\right\|^{2p}\right]^{\frac{1}{2p}} \\ &= \mathbb{E}\left[\left\|\sum_{i=1}^{N-1} Y_i\right\|^{2p}\right]^{\frac{1}{2p}} \\ &\leq C_{n,p}\frac{1}{\sqrt{N}}, \end{aligned}$$

from which we obtain the desired result. ∎

Acknowledgments. We would like to acknowledge the editor and anonymous reviewers for their helpful comments and constructive suggestions.

REFERENCES

[1] J. Allen, M. Eknes, and G. Evensen, *An ensemble Kalman filter with a complex marine ecosystem model: hindcasting phytoplankton in the cretan sea*, in Annales Geophysicae, vol. 21, Copernicus GmbH, 2003, pp. 399–411.

[2] B. D. Anderson, *Stability properties of Kalman-Bucy filters*, Journal of the Franklin Institute, 291 (1971), pp. 137–144.

[3] J. L. Anderson, *An ensemble adjustment Kalman filter for data assimilation*, Monthly weather review, 129 (2001), pp. 2884–2903.

[4] 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.

[5] R. Bucy, *Nonlinear filtering theory*, IEEE Transactions on Automatic Control, 10 (1965), pp. 198–198.

[6] R. S. Bucy, *Global theory of the Riccati equation*, Journal of computer and system sciences, 1 (1967), pp. 349–361.

[7] X. Chen, X. Luo, J. Shi, and S. S.-T. Yau, *General convergence result for continuous-discrete feedback particle filter*, International Journal of Control, (2021), pp. 1–15, https://doi.org/10.1080/00207179.2021.1948105, https://doi.org/10.1080/00207179.2021.1948105.

[8] P. Del Moral and J. Tugaut, *On the stability and the uniform propagation of chaos properties of ensemble Kalman–Bucy filters*, The Annals of Applied Probability, 28 (2018), pp. 790–850.

[9] 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.

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

[11] R. E. Kalman and R. S. Bucy, *New results in linear filtering and prediction theory*, Journal of Basic Engineering, 83 (1961), pp. 95–108.

[12] J. Kang, X. Chen, Y. Tao, and S. S.-T. Yau, *Optimal transportation particle filter for linear filtering systems with correlated noises*, IEEE Transactions on Aerospace and Electronic Systems, (2022), pp. 1–15, https://doi.org/10.1109/TAES.2022.3166863, https://ieeexplore.ieee.org/document/9756375.

[13] I. Karatzas and S. Shreve, *Brownian motion and stochastic calculus*, vol. 113, Springer Science & Business Media, 2012.

[14] P. Lancaster and L. Rodman, *Algebraic Riccati equations*, Clarendon press, 1995.

[15] A. I. Perov and I. D. Kostrub, *On the spectral abscissa and the logarithmic norm*, Mathematical Notes, 101 (2017), pp. 677–687.

[16] H. P. Rosenthal, *On the subspaces of L^p $(p > 2)$ spanned by sequences of independent random variables*, Israel Journal of Mathematics, 8 (1970), pp. 273–303.

[17] A. Taghvaei, J. De Wiljes, P. G. Mehta, and S. Reich, *Kalman filter and its modern extensions for the continuous-time nonlinear filtering problem*, Journal of Dynamic Systems, Measurement, and Control, 140 (2018).

[18] 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.

[19] 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.

[20] A. Taghvaei and P. G. Mehta, *Error analysis of the stochastic linear feedback particle filter*, in 2018 IEEE Conference on Decision and Control (CDC), IEEE, 2018, pp. 7194–7199.

[21] T. Yang, P. G. Mehta, and S. P. Meyn, *Feedback particle filter*, IEEE Transactions on Automatic Control, 58 (2013), pp. 2465–2480.