Error Estimation for the Particle Filter

Ziyu Liu
Department of Applied Math and Statistics
Johns Hopkins University
Baltimore, MD
Email: zliu82@jhu.edu

James C. Spall
Department of Applied Math and Statistics
Applied Physics Laboratory
Johns Hopkins University
Baltimore, MD
Email: James.Spall@jhuapl.edu

Abstract—**The particle filter is a popular algorithm for solving the state-space problem for its easy implement. Many previous studies have been conducted to study the asymptotical behavior of particle filters. In our previous works, we divided the error of particle filter into two parts. By using Lindeberg's central limit theorem, we showed that one of them is asymptotically normal. However, it's hard to estimate the covariance matrix of it's converged distribution. This paper aims at giving a computable estimator for the covariance matrix.**

Index Terms—**particle filter, error estimation, asymptotically normality, martingale structure**

I. Introduction

In separate work [3], we have established the asymptotic normality for the difference between the particle filter (PF) estimator and the conditional mean for non-linear and/or non-Gaussian discrete state-space model in multivariate cases. In this paper, we show how to estimate the covariance matrix appearing in the asymptotic distribution.

Discrete time state-space model (DSSM) can be expressed as follows:

$$\begin{cases} \boldsymbol{x}_{k+1} = \boldsymbol{f}_k(\boldsymbol{x}_k,\ \boldsymbol{w}_k) \\ \boldsymbol{z}_k = \boldsymbol{h}_k(\boldsymbol{x}_k,\ \boldsymbol{v}_k) \end{cases} \tag{1}$$

where $\boldsymbol{x}_k$ and $\boldsymbol{z}_k$ are the true states and the observations, respectively. We are interested in a general case where $\boldsymbol{f}_k$ and $\boldsymbol{h}_k$ could be non-linear, $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ could be non-Gaussian.

In our previous work [3], we decompose the PF error into two terms as following:

$$\hat{\boldsymbol{x}}_k - \boldsymbol{x}_k = (\hat{\boldsymbol{x}}_k - E[\boldsymbol{x}_k|\boldsymbol{z}_{1:k}]) + (E[\boldsymbol{x}_k|\boldsymbol{z}_{1:k}] - \boldsymbol{x}_k) \tag{2}$$

where $\boldsymbol{z}_{1:k}$ represent $(\boldsymbol{z}_1, \cdots, \boldsymbol{z}_k)$. The first part is the difference between the PF estimate $\hat{\boldsymbol{x}}_k$ and the conditional mean $E[\boldsymbol{x}_k|\boldsymbol{z}_{1:k}]$, and second part is the difference between the conditional mean $E[\boldsymbol{x}_k|\boldsymbol{z}_{1:k}]$ and the true state $\boldsymbol{x}_k$. For the first part of the error decomposition, [1] uses the result of [2] to show that the difference between generic PF estimate and the conditional mean is asymptotically normal in scalar cases as the number of particles gets large. We then established that a similar result holds for the multivariate case in [3], for which we give a proof sketch in the appendix. However, in order to describe the limit distribution of the first term, we have to estimate the covariance matrix of the limit distribution. That is the purpose of this paper.

The PF, proposed in [4], is a popular Bayesian filtering algorithm for its ease of implementation and wide range of application. The principle of PFs (alternatively term sequential Monte Carlo methods) are using iterative algorithms which product a set of weighted simulations (the "particles") in order to approximate the posterior distributions by weighted summations [5]. Many previous studies have been conducted on describing the asymptotic properties of PFs. Liu and Chen [6] stated some form of law of large numbers for some special case PFs, which is PF estimators converge almost surely to interest quantities. A central limit theorem for this algorithm was derived in [7] and later refined in [8]. Later on, a more simple proof method was provided in [9]. Chopin [5] gives a central limit theorem for PFs in order to provide an exact measure of the Monte Carlo error. From Chopin's paper, we can also access the stability of PFs. [10] derived an asymptotic theory of weighted system of particles. It provides both a law of large number and central limit theorem for PFs. Note that the asymptotic normality for the PF is in terms of the number of particles, not the number of time points. In particular, it can be shown that the Kalman filter (and presumably the PF, as well) are generally asymptotically non-normally distributed as the number of time points increase when the noise terms are non-normal (see [11] or [12]).

In this paper, we provide an estimator of the covariance matrix based on results in [1], which discuss the scalar case limit distribution of $\hat{\boldsymbol{x}}_k - E(\boldsymbol{x}_k|\boldsymbol{z}_{1:k})$. The first part will describe the error decomposition and result of [3]. The second part will give the estimator of the covariance matrix of the limit distribution, then, followed by rigorous proof and some discussion. Some details of proof will shown in appendix.

II. Problem Statement

A. Particle Filter

Consider the DSSM as a hidden Markov process (HMP) with state $\boldsymbol{x}_k$ and measurement $\boldsymbol{z}_k$ according to $p_k(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$ and $p_k(\boldsymbol{z}_k|\boldsymbol{x}_k)$ respectively. Note that we are following convention where the argument of a p.d.f (probability density function) defines the form of the p.d.f. Suppose our target is to estimate the true state $\boldsymbol{x}_T$ for some terminal time T. PF is a Monte Carlo estimation based on vector of obervations history $\boldsymbol{z}_{1:T} = (\boldsymbol{z}_1, \cdots, \boldsymbol{z}_T)$.

To avoid degeneration problem, where all but few particles have non-zero weight, we do re-sampling at each time of updating. The multinomial re-sampling schemes were mentioned in previous work [3]. Remark: our notations: $\tilde{\boldsymbol{x}}_k^i$, $\boldsymbol{x}_k^i$ are states of the i^{th} particle at time k before and after resampling respectively. Denote vectors record the history of particles by $\tilde{\boldsymbol{x}}_{1:k}^i = (\boldsymbol{x}_{1:k-1}^i, \tilde{\boldsymbol{x}}_k^i)$, $\boldsymbol{x}_{1:k}^i = \tilde{\boldsymbol{x}}_{1:k}^j$ for some j, where $1 \le j \le m$.

Given a proposal distribution $q_k(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$ which dominates the p.d.f $p_k(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$ (it means that for any $\boldsymbol{x}_k$ such that $p_k(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) > 0$, we have $q_k(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) > 0$) and a total number of particles m, which is unchangeable during updating and re-sampling, we can construct particle filter as follows:

Update: sample $\tilde{\boldsymbol{x}}_k^i \sim q_k(\cdot|\boldsymbol{x}_{k-1}^i)$.
Reweight: compute weight of particle i as:
$\alpha_k(\boldsymbol{x}_{1:k}^i) = p_k(\boldsymbol{x}_k^i|\boldsymbol{x}_{k-1})p_k(\boldsymbol{z}_k|\boldsymbol{x}_k^i)/q_k(\boldsymbol{x}_k^i|\boldsymbol{x}_{1:k-1}^i)$.
Normalize the weight $w_k^i = \alpha_k(\boldsymbol{x}_{1:k}^i)/\sum_{j=1}^{m}\alpha_k(\boldsymbol{x}_{1:k}^j)$.
Re-sampling: sample $\boldsymbol{x}_{1:k}^i$ from $\{\tilde{\boldsymbol{x}}_{1:k}^1, \cdots, \tilde{\boldsymbol{x}}_{1:k}^m\}$ with probability $\{w_k^1, \cdots, w_k^m\}$. Denote the number of copies of $\tilde{\boldsymbol{x}}_k^i$ appearing in $\boldsymbol{x}_k^j, j = 1, \cdots, m$ by $\#_k^i$.
The PF estimator for the terminal state is $\hat{\boldsymbol{x}}_T = m^{-1}\sum_{i=1}^m \boldsymbol{x}_T^i w_T^i$.

B. Error Decomposition

In our previous work, we decomposed the error of between PF estimator and true state $\boldsymbol{x}_T$ into two parts, which are uncorrelated:

$$\hat{\boldsymbol{x}}_T - \boldsymbol{x}_T = (\hat{\boldsymbol{x}}_T - E(\boldsymbol{x}_T|\boldsymbol{z}_{1:T})) + (E(\boldsymbol{x}_T|\boldsymbol{z}_{1:T}) - \boldsymbol{x}_T). \quad (3)$$

The motivation to do this is: based on our partially obervations, the best estimator of $\boldsymbol{x}_T$ should be $E(\boldsymbol{x}_T|\boldsymbol{z}_{1:T})$. The PF estimator can not beat the conditional mean, which means $E[(\hat{\boldsymbol{x}}_T - \boldsymbol{x}_T)^2] \ge E[(E(\boldsymbol{x}_T|\boldsymbol{z}_{1:T}) - \boldsymbol{x}_T)^2]$ always holds. However, as far as PF estimator is close enough to $E(\boldsymbol{x}_T|\boldsymbol{z}_{1:T})$, we are satisfied with its performance. Then, $\hat{\boldsymbol{x}}_T - E(\boldsymbol{x}_T|\boldsymbol{z}_{1:T})$, as a random variable, can be proved to have a normal limit distribution.

C. Limit Distribution

Remark notations from [3]: for terminal time point T, the conditional density function is:

$$p_T(\boldsymbol{x}_{1:T}|\boldsymbol{z}_{1:T}) \propto \prod_{k=1}^{T}[p_k(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})p_k(\boldsymbol{z}_k|\boldsymbol{x}_k)]. \quad (4)$$

The likelihood ratio between conditional distribution and proposal distribution is defined as:

$$L_T(\boldsymbol{x}_{1:T}) = \frac{p_T(\boldsymbol{x}_{1:T}|\boldsymbol{z}_{1:T})}{\prod_{k=1}^{T} q_k(\boldsymbol{x}_k|\boldsymbol{x}_{1:k-1})}. \quad (5)$$

For computational convenience, we also define the following quantities:

$$\bar{\alpha}_k = \frac{1}{m}\sum_{j=1}^{m}\alpha_k(\tilde{\boldsymbol{x}}_{1:k}^j), \quad (6)$$

$$H_k^i = \frac{\bar{\alpha}_1 \cdots \bar{\alpha}_k}{\prod_{l=1}^{k}\alpha_l(\boldsymbol{x}_{1:l}^i)} \quad , \tilde{H}_k^i = \frac{\bar{\alpha}_1 \cdots \bar{\alpha}_k}{\prod_{l=1}^{k}\alpha_l(\tilde{\boldsymbol{x}}_{1:l}^i)}. \quad (7)$$

We can state the conclusion from [3] as follows:
Let the function $\boldsymbol{u}_k(\boldsymbol{x}_{1:k})$ be:

$$\boldsymbol{u}_k(\boldsymbol{x}_{1:k}) = \begin{cases} E[\boldsymbol{x}_T|\boldsymbol{z}_{1:T}], & k = 0 \\ E[\boldsymbol{x}_T L_T(\boldsymbol{x}_{1:T})|\boldsymbol{x}_{1:k}], & 1 \le k \le T \end{cases} \quad (8)$$

$$g_k^*(\boldsymbol{x}_{1:k}) = \frac{E[\prod_{l=1}^{k}\alpha_l(\boldsymbol{x}_{1:l})]}{\prod_{l=1}^{k}\alpha_l(\boldsymbol{x}_{1:l})}, \quad (9)$$

and $\Sigma = \sum_{k=1}^{2T-1}\Sigma_k$, where

$$\begin{cases} \Sigma_{2k-1} = E\{(\boldsymbol{u}_k(\boldsymbol{x}_{1:k})\boldsymbol{u}_k(\boldsymbol{x}_{1:k})^T - \\ \qquad \boldsymbol{u}_{k-1}(\boldsymbol{x}_{1:k-1})\boldsymbol{u}_{k-1}(\boldsymbol{x}_{1:k-1})^T)g_{k-1}^*(\boldsymbol{x}_{1:k-1})\}, \\ \Sigma_{2k} = E\{[\boldsymbol{u}_k(\boldsymbol{x}_{1:k})g_k^*(\boldsymbol{x}_{1:k}) - \boldsymbol{u}_0] \times \\ \qquad [\boldsymbol{u}_k(\boldsymbol{x}_{1:k})g_k^*(\boldsymbol{x}_{1:k}) - \boldsymbol{u}_0]^T/g_k^*(\boldsymbol{x}_{1:k})\}. \end{cases} \quad (10)$$

Given the HMP as (1), for the PF estimator $\hat{\boldsymbol{x}}_T$ obtained by resampling at each step, if $\det(\Sigma_k) < \infty$ for all k, then as $m \to \infty$

$$\sqrt{m}(\hat{\boldsymbol{x}}_T - E(\boldsymbol{x}_T|\boldsymbol{z}_{1:T})) \xrightarrow{\text{dist}} N(0, \Sigma). \quad (11)$$

In [3], this theorem was rigorously proved. We give a proof sketch in the appendix. Since Σ contains conditional mean which is uncomputable, we can not practically use the limit distribution of our target. This unsatisfying result motivates us to construct an estimator of Σ, as shown below.

III. Estimator of Σ

The estimator of Σ will be expressed as:
$\tilde{\Sigma} = m^{-1}\sum_{i=1}^m(\sum_{k=1}^{2T-1}\tilde{M}_t^i)(\sum_{k=1}^{2T-1}\tilde{M}_t^i)^T$, where:

$$\begin{cases} \tilde{M}_{2k-1}^i = \sum_{j:A_{k-1}^j = i}[\boldsymbol{u}_k(\tilde{\boldsymbol{x}}_{1:k}^j) - \boldsymbol{u}_{k-1}(\boldsymbol{x}_{1:k-1}^j)]H_{k-1}^j, \\ \tilde{M}_{2k}^i = \sum_{j:A_{k-1}^j = i}(\#_k^j - mw_k^j)[\boldsymbol{u}_k(\tilde{\boldsymbol{x}}_{1:k}^j)\tilde{H}_k^j - \boldsymbol{u}_0]. \end{cases} \quad (12)$$

We denote the ancestry origin of particle $\boldsymbol{x}_k^i$ by A_k^i to keep track of it and $A_0^i = i$ for all $1 \le i \le m$ by definition. For example, if $\boldsymbol{x}_{1:k+1}^i = (\boldsymbol{x}_{1:k}^j, \tilde{\boldsymbol{x}}_{k+1}^j)$, then $A_k^j = A_{k+1}^i$.

A. Computable form of $\tilde{\Sigma}$

We can expand the espression of $\tilde{\Sigma}$ as follows:

$$\tilde{\Sigma} = m^{-1}\sum_{i=1}^{m}(\sum_{k=1}^{2T-1}\tilde{M}_k^i)(\sum_{k=1}^{2T-1}\tilde{M}_k^i)^T$$
$$= m^{-1}\sum_{i=1}^{m}\Big\{\sum_{j:A_{t-1}^j=i}\boldsymbol{u}_T(\tilde{\boldsymbol{x}}_{1:T}^j)H_{T-1}^j - \boldsymbol{u}_0 - \sum_{k=1}^{T-1}\sum_{j:A_{k-1}^j=i}(\#_k^j - mw_k^j)\boldsymbol{u}_0\Big\}$$
$$\Big\{\sum_{j:A_{t-1}^j=i}\boldsymbol{u}_T(\tilde{\boldsymbol{x}}_{1:T}^j)H_{T-1}^j - \boldsymbol{u}_0 - \sum_{k=1}^{T-1}\sum_{j:A_{k-1}^j=i}(\#_k^j - mw_k^j)\boldsymbol{u}_0\Big\}^T,$$

where $\boldsymbol{u}_T(\tilde{\boldsymbol{x}}_{1:T}^j) = \tilde{\boldsymbol{x}}_T^j L_T(\tilde{\boldsymbol{x}}_{1:T}^j)$. From Lemma 3 in the appendix $L_T(\tilde{\boldsymbol{x}}_{1:T}^j)H_{t-1}^j$ converges to $\alpha_T(\tilde{\boldsymbol{x}}_T^i)/\bar{\alpha}_T$ in probability. The only uncomputable part in this expression is $\boldsymbol{u}_0 = E(\boldsymbol{x}_T|z_{1:T})$. Fortunately, we can replace $\boldsymbol{u}_0$ by the PF estimator $\hat{\boldsymbol{x}}_T$ since $\boldsymbol{u}_0 = \hat{\boldsymbol{x}}_T + O_p(1/\sqrt{m})$ from (11). The result will converge to $\tilde{\Sigma}$ in probability.

B. Proof

We introduce the following sets to clarify the proof:

$$\begin{cases} \mathcal{F}_{2k-1} = \{\tilde{\boldsymbol{x}}_1^i : 1 \le i \le m\}\cup \\ \qquad\{(\boldsymbol{x}_{1:l}^i, \tilde{\boldsymbol{x}}_{1:l+1}^i, A_l^i) : 1 \le l < k, 1 \le i \le m\}, \\ \mathcal{F}_{2k} = \mathcal{F}_{2k-1} \cup \{(\boldsymbol{x}_{1:k}^i, A_k^i) : 1 \le i \le m\}. \end{cases} \tag{13}$$

We prove the result by induction on l to show that: $m^{-1}\sum_{i=1}^{m}(\sum_{k=1}^{l}\tilde{M}_k^i)(\sum_{k=1}^{l}\tilde{M}_k^i)^T \xrightarrow{p} \sum_{k=1}^{l}\Sigma_k$. Recall that $l < \infty$ in the setting here while $m \to \infty$. Hence an induction argument is valid. For $l = 1$, it's a directly result of the weak law of large numbers. This is the base case.

For $l > 1$, assume $m^{-1}\sum_{i=1}^{m}(\sum_{k=1}^{l}\tilde{M}_k^i)(\sum_{k=1}^{l}\tilde{M}_k^i)^T \xrightarrow{p} \sum_{k=1}^{l}\Sigma_k$, now consider:

$$m^{-1}\sum_{i=1}^{m}(\sum_{k=1}^{l+1}\tilde{M}_k^i)(\sum_{k=1}^{l+1}\tilde{M}_k^i)^T = \underbrace{m^{-1}\sum_{i=1}^{m}(\sum_{k=1}^{l}\tilde{M}_k^i)(\sum_{k=1}^{l}\tilde{M}_k^i)^T}_{(1)} +$$
$$\underbrace{m^{-1}\sum_{i=1}^{m}\tilde{M}_{l+1}^i(\tilde{M}_{l+1}^i)^T}_{(2)} +$$
$$\underbrace{m^{-1}(\sum_{i=1}^{m}\tilde{M}_{l+1}^i(\sum_{k=1}^{l}\tilde{M}_k^i)^T + \sum_{i=1}^{m}(\sum_{k=1}^{l}\tilde{M}_k^i)(\tilde{M}_{l+1}^i)^T)}_{(3)}.$$

By induction, we know that the (1) will converge in probability to $\sum_{k=1}^{l}\Sigma_k$, so, we need to discuss part (2) and (3). Let us denote $\gamma_k^i = [\boldsymbol{u}_k(\tilde{\boldsymbol{x}}_{1:k}^i) - \boldsymbol{u}_{k-1}(\boldsymbol{x}_{1:k-1}^i)]H_{k-1}^i$, $\eta_k^i = \sum_{j:A_{k-1}^j=i}(\#_k^j - mw_k^j)[\boldsymbol{u}_k(\tilde{\boldsymbol{x}}_{1:k}^j)\tilde{H}_k^j - \boldsymbol{u}_0]$. We consider two cases, l is even and l is odd.

Case 1. If l is even, we can find some t such that $l = 2t$.

$$m^{-1}\sum_{i=1}^{m}\tilde{M}_{l+1}^i(\tilde{M}_{l+1}^i)^T = m^{-1}\sum_{i=1}^{m}(\sum_{j:A_t^i=i}\gamma_{t+1}^j)(\sum_{j:A_t^i=i}\gamma_{t+1}^j)^T$$
$$= \underbrace{m^{-1}\sum_{j=1}^{m}\gamma_{t+1}^j(\gamma_{t+1}^j)^T}_{(a)} +$$
$$\underbrace{m^{-1}\sum_{j=1}^{m}\sum_{g:A_t^g=j,g\neq h}\sum_{h:A_t^h=j}\gamma_{t+1}^g(\gamma_{t+1}^h)^T}_{(b)}.$$

Note that $\gamma_{t+1}^j = M_l^i$ (16), therefore, (a) is identical to $m^{-1}\sum_{i=1}^{m}M_{l+1}^iM_{l+1}^i$ which will converge in probability to Σ_{l+1} by (17). We only need to prove that (b) will converge to the 0 matrix.

To prove that (b) converges to the 0 matrix in probability, we only need to prove that each entry of (b) will converge to 0 in probability. To avoid redundant notation, we only consider the entry on the first row and column of (b), which is equal to

$$m^{-1}\sum_{j=1}^{m}\sum_{g:A_t^g=j,g\neq h}\sum_{h:A_t^h=j}\gamma_{t+1}^g(1)\gamma_{t+1}^h(1). \tag{14}$$

We now prove that conditional on $\mathcal{F}_l$, (14) converges to 0 in probability. Then, using characteristic functions, we can prove the unconditional case. Note that $\gamma_{t+1}^j = M_l^i$ which is mean 0 conditional on $\mathcal{F}_l$, γ_{t+1}^i, γ_{t+1}^j are independent conditional on $\mathcal{F}_l$ for any $i \neq j$, we have:

$$\mathrm{Var}(m^{-1}\sum_{j=1}^{m}\sum_{g:A_t^g=j,g\neq h}\sum_{h:A_t^h=j}\gamma_{t+1}^g(1)\gamma_{t+1}^h(1)|\mathcal{F}_l)$$
$$= m^{-2}\sum_{j=1}^{m}\sum_{g:A_t^g=j,g\neq h}\sum_{h:A_t^h=j}E((\gamma_{t+1}^g(1))^2|\mathcal{F}_l)E(\gamma_{t+1}^h(1))^2|\mathcal{F}_l)$$
$$\le (m^{-1}\sum_{g=1}^{m}E(\gamma_{t+1}^g(1)^2|\mathcal{F}_l))\times$$
$$(m^{-1}\max_{1\le h\le m}\sum_{i:A_t^i=h}E(\gamma_{t+1}^i(1))^2|\mathcal{F}_l)) \xrightarrow{p} 0.$$

The last step is a direct result of Lemma 1 in the appendix.

Use the same trick we can prove that (3) converge to 0 in probability.

Case2. l is odd. Hence $t = i/2$ is an integer. We can decompose (2) the same as before:

$$(2) = m^{-1}\sum_{i=1}^{m}(\sum_{j:A_t^i=i}\eta_t^j)(\sum_{j:A_t^i=i}\eta_t^j)^T = \underbrace{m^{-1}\sum_{j=1}^{m}\eta_t^j(\eta_t^j)^T}_{(a')} + \underbrace{m^{-1}\sum_{j=1}^{m}\sum_{g:A_t^g=j,g\neq h}\sum_{h:A_t^h=j}\eta_t^g(\eta_t^h)^T}_{(b')}.$$

We use the conditional distribution of (a') to get the limitating distribution of (a'). This is the same trick as in the proof of (b) converging to 0 matrix. Note that $E(\eta_t^i|\mathcal{F}_{2t-1}) = \sum_{j:A_{t-1}^j=i}(\boldsymbol{u}_t(\tilde{\boldsymbol{x}}_t^j)\tilde{H}_t^j - \boldsymbol{u}_0)E(\#_t^j - mw_t^j|\mathcal{F}_{2t-1}) = 0$.

$$E((a')|\mathcal{F}_l) = m^{-1}\sum_{i=1}^{m}E(\eta_t^i(\eta_t^i)^T|\mathcal{F}_l) = m^{-1}\sum_{i=1}^{m}\mathrm{Var}(\eta_t^i|\mathcal{F}_l)$$
$$= m^{-1}\sum_{i=1}^{m}\mathrm{Var}(\sum_{j:A_{t-1}^j=i}(\#_t^j - mw_t^j)[\boldsymbol{u}_t(\tilde{\boldsymbol{x}}_{1:t}^j)\tilde{H}_t^j - \boldsymbol{u}_0]|\mathcal{F}_l)$$
$$= m^{-1}\sum_{j=1}^{m}\mathrm{Var}(((\#_t^j - mw_t^j)[\boldsymbol{u}_t(\tilde{\boldsymbol{x}}_{1:t}^j)\tilde{H}_t^j - \boldsymbol{u}_0])|\mathcal{F}_l)$$
$$= m^{-1}\sum_{j=1}^{m}mw_t^j(1-w_t^j)[\boldsymbol{u}_t(\tilde{\boldsymbol{x}}_{1:t}^j)\tilde{H}_t^j - \boldsymbol{u}_0][\boldsymbol{u}_t(\tilde{\boldsymbol{x}}_{1:t}^j)\tilde{H}_t^j - \boldsymbol{u}_0]^T$$
$$\xrightarrow{\text{p}} \Sigma_{l+1}.$$

We show in the appendix that (a') converges to Σ_{l+1} in probability, (b') and and that (3) coverge to 0 matrix.

Hence, we have: $m^{-1}\sum_{i=1}^{m}(\sum_{k=1}^{l+1}\tilde{M}_k^i)(\sum_{k=1}^{l+1}\tilde{M}_k^i)^T \xrightarrow{\text{p}} \sum_{k=1}^{l+1}\Sigma_k$. Combining with the above analysis, we conclude that if we assume:

$$m^{-1}\sum_{i=1}^{m}(\sum_{k=1}^{l}\tilde{M}_k^i)(\sum_{k=1}^{l}\tilde{M}_k^i)^T \xrightarrow{\text{p}} \sum_{k=1}^{l}\Sigma_k,$$

we have:

$$m^{-1}\sum_{i=1}^{m}(\sum_{k=1}^{l+1}\tilde{M}_k^i)(\sum_{k=1}^{l}\tilde{M}_k^i)^T \xrightarrow{\text{p}} \sum_{k=1}^{l+1}\Sigma_k.$$

By induction, we have $\hat{\Sigma} \xrightarrow{\text{p}} \Sigma$.
By now, we can analyze the first part of error $\hat{\boldsymbol{x}}_T - E(\boldsymbol{x}_T|\boldsymbol{z}_{1:T})$ when we generate the PF estimator. Specifically, we can calculate the limit distribution of this quantity. This result shows us how far the PF estimator is away from the "best" estimator which is incalculable. Our future work will focus on the second part of error $E(\boldsymbol{x}_T|\boldsymbol{z}_{1:T}) - \boldsymbol{x}_T$ which is uncorrelated with the first part. However, we can not say that they are independent. Our target is to find out a reasonable bound for the second part and study their relation.

IV. Appendix

A. Sketch of previous work

To help in understanding our work here, we re-state the result of our previous work and give a proof sketch:
Theorem: Given the HMP as (1), for PF estimator $\hat{\boldsymbol{x}}_k$ obtained by resampling at each step, if $\det(\Sigma_k) < \infty$ for all k. Then as $m \to \infty$ $\sqrt{m}(\hat{\boldsymbol{x}}_T - \boldsymbol{x}_T) \xrightarrow{\text{dist}} N(0, \Sigma)$.

To prove the theorem, we will state two lemmas without proof:
Lemma 2: Let G be a measurable vector function from history of state-space $\mathbb{R}^{t\times n}$ (t is time and n is dimension of state) to $\mathbb{R}^s$, where $s < \infty$. For any $1 \le k \le T$. Define g_k^* as (9). Then,
(i) if $||E[G(\boldsymbol{x}_{1:k})/g_{k-1}^*(\boldsymbol{x}_{1:k-1})]||_\infty < \infty$,
$m^{-1}\sum_{i=1}^{m}G(\tilde{\boldsymbol{x}}_{1:k}^i) \xrightarrow{p} E[G(\boldsymbol{x}_{1:k})/g_{k-1}^*(\boldsymbol{x}_{1:k-1})]$ as $m \to \infty$;
(ii) if $||E[G(\boldsymbol{x}_{1:k})/g_k^*(\boldsymbol{x}_{1:k})]||_\infty < \infty$,
$m^{-1}\sum_{i=1}^{m}G(\boldsymbol{x}_{1:k}^i) \xrightarrow{p} E[G(\boldsymbol{x}_{1:k})/g_k^*(\boldsymbol{x}_{1:k})]$ as $m \to \infty$.

Lemma 3: If the conditions in Lemma 2 are satisfied, then

$$\frac{H_k^i}{g_k^*(\boldsymbol{x}_{1:k}^i)} \xrightarrow{p} 1 \quad \text{as } m \to \infty.$$

Furthermore, if $E[||G(\tilde{\boldsymbol{x}})||/g_{k-1}^*(\boldsymbol{x}_{1:k-1})] < \infty$, then

$$m^{-1}\sum_{i=1}^{m}||G(\tilde{\boldsymbol{x}}_{1:k}^i)||\mathbf{1}_{\{||G(\tilde{\boldsymbol{x}}_{1:k}^i)||>\varepsilon\sqrt{m}\}} \xrightarrow{p} 0.$$

The proof idea of lemma 2 is same as lemma 1 which we will state later. Lemma 3 is a direct result of lemm 2.

We decompose the "error" into:

$$m(\hat{\boldsymbol{x}}_T^* - E(\boldsymbol{x}_T|\boldsymbol{z}_{1:t})) = \sum_{k=1}^{2T-1}\sum_{i=1}^{m}M_k^i \tag{15}$$

where

$$\begin{cases} M_{2k-1}^i = [\boldsymbol{u}_k(\tilde{\boldsymbol{x}}_{1:k}^i) - \boldsymbol{u}_{t-1}(\boldsymbol{x}_{1:k-1}^i)]H_{k-1}^i \\ M_{2k}^i = \boldsymbol{u}_k(\boldsymbol{x}_{1:k}^i)H_k^i - \sum_{j=1}^{m}w_k^j\boldsymbol{u}_k(\tilde{\boldsymbol{x}}_{1:k}^j)\tilde{H}_k^j. \end{cases} \tag{16}$$

Then by algebra computation and lemma 1, 2, we can conclude that:

$$\frac{1}{m}\sum_{i=1}^{m}E[M_{2k-1}^i{M_{2k-1}^i}^T|\mathcal{F}_{2k-2}] \xrightarrow{\text{p}} \Sigma_{2k-1}, \tag{17}$$

$$\frac{1}{m}\sum_{i=1}^{m}E[M_{2k}^i{M_{2k}^i}^T|\mathcal{F}_{2k-1}] \xrightarrow{\text{p}} \Sigma_{2k}. \tag{18}$$

At last, by induction, we can conclude that:

$$\sqrt{m}(\hat{\boldsymbol{x}}_T - \boldsymbol{x}_T) \xrightarrow{\text{dist}} N(0, \Sigma)$$

as $m \to \infty$.

B. Lemma 1

We now introduce a lemma similar to lemma 3 of [1]. Although we deal with high dimension converge instead of scalar case, we can use this lemma because we are dealing with matrix elementrywise. We will give a proof sketch instead of discussing every details.

Let G be a nonnegative value function on the state space, for any $1 \le k \le T$:

1) if $E(G(\boldsymbol{x}_{1:k})/g^*_{k-1}(\boldsymbol{x}_{1:k-1})) < \infty$, then

$$m^{-1} \max_{1\le j\le m} \sum_{i:A^i_{t-1}=j} G(\tilde{\boldsymbol{x}}^i_k) \xrightarrow{\text{P}} 0 \tag{19}$$

as $m \to \infty$

2) if $E(G(\boldsymbol{x}_{1:k})/g^*_{k}(\boldsymbol{x}_{1:k})) < \infty$, then

$$m^{-1} \max_{1\le j\le m} \sum_{i:A^i_{t-1}=j} G(\boldsymbol{x}^i_k) \xrightarrow{\text{P}} 0 \tag{20}$$

as $m \to \infty$

Proof: The proof process is similar to lemma in [3]. we use induction to show that: if 2 holds for $k-1$, 1 holds for k. If 1 holds for k, 2 holds for k.

For $k = 1$, $A^i_0 = i, g^*_0 = 1$, Note that $E(\boldsymbol{x}_1) < \infty$.

$$\begin{aligned}(19) &= P(m^{-1} \max_{1\le j\le m} G(\tilde{\boldsymbol{x}}^j_1) \ge \epsilon) \\ &= P(\max_{1\le j\le m} G(\tilde{\boldsymbol{x}}^j_1) \ge \epsilon m) \le mP(G(\boldsymbol{x}_1) \ge \epsilon m) \to 0\end{aligned}$$

as $m \to 0$.

Assume that 2 holds for $k-1$, we want to show that 1 holds for k. If not, there exist $\epsilon > 0$, $\{m_i\}^\infty_{i=1}$, $m_1 < m_2 < \cdots$ such that:

$$P(m_i^{-1} \max_{1\le j\le m_i} \sum_{i:A^i_{k-1}=j} G(\tilde{\boldsymbol{x}}^i_k) > 2\epsilon) > \epsilon(2 + E(G(\boldsymbol{x}_k)/g^*_{k-1}(\boldsymbol{x}_{k-1}))) \quad \textit{for all } m_i.$$

Parallel to lemma 1 in [3], we could conduct a contradiction.

Next, we want to show that if 1 holds for k, 2 holds for k. Note that $\sum_{i:A^i_k=j} G(\boldsymbol{x}^i_k) = \sum_{i:A^i_{k-1}} \#^i_k G(\tilde{\boldsymbol{x}}^i_k)$. Right hand side of (20) $= m^{-1} \max_{1\le j\le m} \sum_{i:A^i_{k-1}=j} \#^i_k G(\tilde{\boldsymbol{x}}^i_k)$. Suppose 2 doesn't hold, we have $\epsilon > 0$ and $\{m_i\}^\infty_{i=1}$, $m_1 < m_2 < \cdots$, such that:

$$P(m_i^{-1} \max_{1\le j\le m_i} \sum_{i:A^i_{k-1}=j} \#^i_k G(\tilde{\boldsymbol{x}}^i_k) > 2\epsilon) > \epsilon(2 + E(G(\boldsymbol{x}_k)/g^*_k(\boldsymbol{x}_k))).$$

We decompose $\#^i_k G(\tilde{\boldsymbol{x}}^i_k)$ into three terms: $\#^i_k G(\tilde{\boldsymbol{x}}^i_k) = \underbrace{(\#^i_k - mw^i_k)[G(\tilde{\boldsymbol{x}}^i_k) \wedge \epsilon^3 m]}_{(I)} + \underbrace{mw^i_k[G(\tilde{\boldsymbol{x}}^i_k) \wedge \epsilon^3 m]}_{(II)} + \underbrace{\#^i_k[G(\tilde{\boldsymbol{x}}^i_k) - \epsilon^3 m]^+}_{(III)}$.

(a) we can easily get that $E((I)|\mathcal{F}_{2k-1}) = 0$, therefore, We can bound
$P(|m^{-1} \sum_{i:A^i_{k-1}=j} (\#^i_k - mw^i_k)[G(\tilde{\boldsymbol{x}}^i_k) \wedge \epsilon^3 m]| > \epsilon|\mathcal{F}_{2k-1})$
by:
$(\epsilon m)^{-2}\text{Var}(\sum_{i:A^i_{k-1}=j} (\#^i_k - mw^i_k)[G(\tilde{\boldsymbol{x}}^i_k) \wedge \epsilon^3 m]|\mathcal{F}_{2k-1})$.
Using facts that $\text{Cov}(\#^i_k, \#^j_k) < 0$ for any pair of $1 \le i, j \le m$, $\text{Var}(\#^i_k|\mathcal{F}_{2k-1}) \le mw^i_k = \alpha(\tilde{\boldsymbol{x}}^i_k)/\bar{\alpha}$, $E(\alpha(\boldsymbol{x}_k)G(\boldsymbol{x}_k)/g^*_{k-1}(\boldsymbol{x}_{k-1}))/\bar{\alpha} \xrightarrow{\text{P}} E(G(\boldsymbol{x}_k)/g^*_k(\boldsymbol{x}_k))$ by lemma 2 in the appendix, we could conclude:
$\sum^m_{j=1} P(|m^{-1} \sum_{i:A^i_{k-1}=j} (\#^i_k - mw^i_k)[G(\tilde{\boldsymbol{x}}^i_k) \wedge \epsilon^3 m]| > \epsilon) \le \epsilon(1 + E(G(\boldsymbol{x}_k)/g^*_k(\boldsymbol{x}_k)))$ for large m. We can get rid of $\mathcal{F}_{2k-1}$ since we sum up over j and take expectation on both sides.

(b) By lemma in [3], we can conduct that: $m^{-1} \max_{1\le j\le m} \sum_{i:A^i_{k-1}=j} mw^i_k G(\tilde{\boldsymbol{x}}^i_k) \wedge \epsilon^3 m \xrightarrow{\text{P}} 0$.

(c) Noting $P(\#^i_k[G(\tilde{\boldsymbol{x}}^i_k) - \epsilon^3 m]^+ \neq 0|\mathcal{F}_{2k-1}) = P(\#^i_k > 0|\mathcal{F}_{2k-1})\mathbf{1}_{\{[G(\tilde{\boldsymbol{x}}^i_k)-\epsilon^3 m]^+\neq 0\}}$, by appling lemma 2 in the appendix, we have:
$\sum^m_{i=1} P(\#^i_k[G(\tilde{\boldsymbol{x}}^i_k) - \epsilon^3 m]^+ \neq 0|\mathcal{F}_{2k-1}) \xrightarrow{\text{P}} 0$.

Combine (a), (b), (c) we can conduct a contradiction.

C. Complement main proof

(a) We first show that variance of each components of $m^{-1} \sum^m_{i=1} \eta^i_t(\eta^i_t)^T$ on conditional of $\mathcal{F}_{l-1}$ converge to 0. It's sufficient to show that for any row r and column s, $m^{-1} \sum^m_{i=1} \eta^i_t(\eta^i_t)^T(r,s) = m^{-1} \sum^m_{i=1} \eta^i_t(r)\eta^i_t(s)$ satisfies:

$$\begin{aligned}&E((m^{-1} \sum^m_{i=1} \eta^i_t(r)\eta^i_t(s))^2|\mathcal{F}_l) \\ &= m^{-2} \sum^m_{i=1} E((\eta^i_t(r)\eta^i_t(s))^2) + m^{-2} \sum_{i\neq j} E((\eta^i_t(r)\eta^j_t(s))^2) \\ &\xrightarrow{\text{P}} (\Sigma_l(r,s))^2\end{aligned}$$

Above result is from Lemma 2 in the appendix and following two equations:

$$\begin{aligned}&E[(\#^i_t - mw^i_t)^2(\#^j_t - mw^j_t)^2|\mathcal{F}_{2t-1}] \\ &= (3m^{-2} - 6m^{-3})(mw^i_t)^2(mw^j_t)^2 + (-m^{-1} + \\ &2m^{-2})((mw^i_t)^2 mw^j_t + mw^i_t(mw^j_t)^2) + (1 - m^{-1})mw^i_t mw^j_t\end{aligned}$$

and

$$E((\#_t^i - mw_t^i)^4 | \mathcal{F}_{2t-1})$$
$$= m(w_t^i(1-w_t^i)^4 + (1-w_t^i)(w_t^i)^4)$$
$$+ 3m(m-1)((w_t^i)^2(1-w_t^i)^4$$
$$+ 2(w_t^i)^3(1-w_t^i)^3 + (w_t^i)^4(1-w_t^i)^2)$$

Since we know that $E(m^{-1}\sum_{i=1}^{m}\eta_t^i(r)\eta_t^i(s)|\mathcal{F}_l) \xrightarrow{\text{p}} \Sigma_l(r,s)$, we know that $\text{Var}(m^{-1}\sum_{i=1}^{m}\eta_t^i(r)\eta_t^i(s)|\mathcal{F}_l) \xrightarrow{\text{p}} 0$, we done.

(b) $m^{-1}\sum_{j=1}^{m}\sum_{g,h,A_t^g=A_t^h=j,g\neq h}\eta_t^g(\eta_t^h)^T \xrightarrow{\text{p}} 0 \iff m^{-1}\sum_{j=1}^{m}\sum_{g,h,A_t^g=A_t^h=j,g\neq h}\eta_t^g(r)\eta_t^h(s) \xrightarrow{\text{p}} 0$ for any pair of (r,s). It's sufficient to show that:

$$m^{-1}\sum_{j=1}^{m}E((\sum_{g:A_t^g=j,g\neq h}\sum_{h:A_t^h=j}\eta_t^g(r)\eta_t^h(s))^2|\mathcal{F}_l) \to 0.$$

The left hand side is equivalent to:

$$m^{-2}(3m^{-2}-6m^{-3})\{\sum_{j=1}^{m}\sum_{g:A_t^g=j,g\neq h}\sum_{h:A_t^h=j}$$
$$mw_t^g(\boldsymbol{u}_t(\tilde{\boldsymbol{x}}_t^g(s))\tilde{H}_t^g - \boldsymbol{u}_0(s))*$$
$$mw_t^h(\boldsymbol{u}_t(\tilde{\boldsymbol{x}}_t^h(r))\tilde{H}_t^h - \boldsymbol{u}_0(r))\}^2$$
$$+ m^{-2}(1-m^{-1})\sum_{j=1}^{2}\sum_{g,h,A_t^g=A_t^h=j,g\neq h}$$
$$w_t^g(\boldsymbol{u}_t(\tilde{\boldsymbol{x}}_t^g(s))\tilde{H}_t^g - \boldsymbol{u}_0(s))^2*$$
$$mw_t^h(\boldsymbol{u}_t(\tilde{\boldsymbol{x}}_t^h(r))\tilde{H}_t^h - \boldsymbol{u}_0(r))^2$$
$$-4m^{-2}(m^{-1}-2m^{-2})\sum_{j=1}^{m}(\sum_{g:A_{t-1}^g=j}$$
$$mw_t^g(\boldsymbol{u}_t(\tilde{\boldsymbol{x}}_t^g(r))\tilde{H}_t^g - \boldsymbol{u}_0(r))^2*$$
$$(\sum_{h:A_{t-1}^h=j}\boldsymbol{u}_t(\tilde{\boldsymbol{x}}_t^h(s))\tilde{H}_t^h - \boldsymbol{u}_0(s))^2) \leq$$
$$3(m^{-1}(\sum_{j=1}^{m}mw_t^g|\boldsymbol{u}_t(\tilde{\boldsymbol{x}}_t^g(s))\tilde{H}_t^g - \boldsymbol{u}_0(s)|)*$$
$$m^{-1}\max_{1\leq j\leq m}\sum_{i:A_{t-1}^i=j}mw_t^i|\boldsymbol{u}_t(\tilde{\boldsymbol{x}}_t^h(r))\tilde{H}_t^h - \boldsymbol{u}_0(r)|)^2+$$
$$m^{-1}(\sum_{j=1}^{m}mw_t^g(\boldsymbol{u}_t(\tilde{\boldsymbol{x}}_t^g(s))\tilde{H}_t^g - \boldsymbol{u}_0(s))^2)*$$
$$m^{-1}\max_{1\leq j\leq m}\sum_{i:A_{t-1}^i=j}mw_t^h(\boldsymbol{u}_t(\tilde{\boldsymbol{x}}_t^h(r))\tilde{H}_t^h - \boldsymbol{u}_0(r))^2 \xrightarrow{\text{p}} 0.$$

(c) To show that $m^{-1}(\sum_{i=1}^{m}\tilde{M}_{l+1}^i(\sum_{k=1}^{l}\tilde{M}_k^i)^T + \sum_{i=1}^{m}(\sum_{k=1}^{l}\tilde{M}_k^i)(\tilde{M}_{l+1}^i)^T) \to 0$, we only need to show that for any row r and column s, we have:

$$m^{-1}\sum_{i=1}^{m}E((\tilde{M}_{l+1}^i(r)(\sum_{k=1}^{l}\tilde{M}_k^i(s)))^2|\mathcal{F}_l) \to 0 \iff$$
$$m^{-1}\sum_{i=1}^{m}(\sum_{k=1}^{l}\tilde{M}_k^i(s))^2E((\tilde{M}_{l+1}^i(r))^2|\mathcal{F}_l) \to 0.$$

By computation we have:

$$E((\tilde{M}_{l+1}^i(r))^2|\mathcal{F}_l)$$
$$= m^{-1}\sum_{i:A_{t-1}^j=j}mw_t^i(\boldsymbol{u}_t(\tilde{\boldsymbol{x}})(r)\tilde{H}_t^i - \boldsymbol{u}_0(r))^2$$
$$-(m^{-1}\sum_{i:A_{t-1}^j=j}mw_t^i(\boldsymbol{u}_t(\tilde{\boldsymbol{x}})(r)\tilde{H}_t^i - \boldsymbol{u}_0(r)))^2 \xrightarrow{\text{p}} 0.$$

Note that $m^{-1}\sum_{i=1}^{m}(\sum_{k=1}^{l}\tilde{M}_k^i(s))^2$ will converge to the (s,s) entry of $\sum_{k=1}^{l}\Sigma_k$, which is finite, we have $m^{-1}\sum_{i=1}^{m}(\sum_{k=1}^{l}\tilde{M}_k^i(s))^2E((\tilde{M}_{l+1}^i(r))^2|\mathcal{F}_l) \to 0.$

References

[1] Chan, H. P. and Lai, T. L., ”A general theory of particle filters in hidden Markov models and some applications”, Annals of Statistics, vol.41, no.6, pp.2877–2904, 2013.

[2] Chan, H. P. and Lai, T. L., ”A sequential Monte Carlo approach to computing tail probabilities in stochastic models”, Annals of Applied Probability, vol.21, no.6, pp.2315–2342, 2011.

[3] Liu, Z., Wei, S., and Spall, J. C. (2019), “Error Analysis for the Particle Filter,” Proceedings of the American Control Conference, Philadelphia, PA, 10–12 July 2019, to appear.

[4] Gordon, N., Salmond, D., and Smith, A., ”Novel approach to nonlinear/non-Gaussian Bayesian state estimation”, IEE Proceedings F Radar and Signal Processing, vol.140, no.2, pp.107, 1993.

[5] Chopin, N, “Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference”, Ann. Statist. 32 2385–2411, 2004.

[6] Liu, J. and Chen, R, “Sequential Monte Carlo methods for dynamic systems”, J. Amer. Statist. Assoc. 93 1032–1044, 1998.

[7] Del Moral, P. and Guionnet, “Central limit theorem for nonlinear filtering and interacting particle systems”, Ann. Appl. Probab. 9 275–297, 1999.

[8] Del Moral, P. and Guionnet, “Central limit theorem for nonlinear filtering and interacting particle systems”, Ann. Appl. Probab. 9 275–297, 1999.

[9] Kunsch, H.-R, “State space and hidden Markov models. In Complex Stochastic Systems” (O. E. Barndorff-Nielsen, D. R. Cox and C. Kluppelberg, eds.) 109–173, Chapman and Hall, London, 2001.

[10] Douc, R. and Moulines, E, “Limit theorems for weighted samples with applications to sequential Monte Carlo methods”, Ann. Statist. 36 2344–2376. MR2458190, 2008.

[11] Maryak, J. L., Spall, J. C., and Heydon, B. D. , “Use of the Kalman Filter for Inference in State-Space Models with Unknown Noise Distributions,” IEEE Transactions on Automatic Control, vol. 49, pp. 87–90, 2004.

[12] Spall, J. C. and Wall, K. D., “Asymptotic Distribution Theory for the Kalman Filter State Estimator,” Communications in Statistics: Theory and Methods, vol. 13, pp. 1981–2003, 1984.