Feedback Particle Filter for Collective Inference

Jin Won Kim and Prashant G. Mehta

Abstract—The purpose of this paper is to describe the feedback particle filter algorithm for problems where there are a large number \( (M) \) of non-interacting agents (targets) with a large number \( (M) \) of non-agent specific observations (measurements) that originate from these agents. In its basic form, the problem is characterized by data association uncertainty whereby the association between the observations and agents must be deduced in addition to the agent state. In this paper, the large-\( M \) limit is interpreted as a problem of collective filtering. This viewpoint is used to derive the equation for the empirical distribution of the hidden agent states. A feedback particle filter algorithm for this problem is presented and illustrated via numerical simulations. The simulations help show that the algorithm well approximates the empirical distribution of the hidden states for large \( M \).

I. INTRODUCTION

Filtering with data association uncertainty is important to a number of classical applications, including target tracking, weather surveillance, remote sensing, autonomous navigation and robotics [1], [2]. Consider, e.g., the problem of multiple target tracking (MTT) with radar. The targets can be multiple aircrafts in air defense, or multiple weather cells in weather surveillance, or multiple landmarks in autonomous navigation and robotics. In each of these applications, there exists data association uncertainty in the sense that one can not assign, in an apriori manner, individual observations (measurements) to individual targets. Given the large number of applications, algorithms for filtering problems with data association uncertainty have been extensively studied in the past; cf., [1], [3], [4] and references therein. The feedback particle filter algorithm for this problem appears in [5].

The filtering problem with data association uncertainty is closely related to the filtering problem with aggregate and anonymized data. Some of these problems have gained in importance recently because of COVID-19. Indeed, the spread of COVID-19 involves dynamically evolving hidden processes (e.g., number of infected, number of asymptomatic etc.) that must be deduced from noisy and partially observed data (e.g., number of tested positive, number of deaths, number of hospitalized etc.). In carrying out data assimilation for such problems, one only has aggregate observations. For example, while the number of daily tested positives is available, the information on the disease status of any particular agent in the population is not known. Such problems are referred to as collective or aggregate inference problems.

In a recent important work [6], algorithms are described for solving the collective inference problem in graphical models, based on the large deviation theory. These results are also specialized to the smoothing problems for the hidden Markov models (HMMs). Two significant features of the proposed algorithm are: (i) the complexity of the data assimilation does not grow with the size of the population; and (ii) for a single agent, the algorithm reduces to the classical forward-backward smoothing algorithm for HMMs.

The purpose of this paper is to interpret the collective inference problem as a limit of the data association problem, as the number of agents \( (M) \) become large. Indeed, for a small number of agents, data association can help reduce the uncertainty and improve the performance of the filter. However, as number of agents gets larger, the data association-based solutions become less practical and may not offer much benefit: On the one hand, the complexity of the filter grows because the number of associations for \( M \) agents with \( M \) observations is \( M! \). On the other hand, the performance of any practical algorithm is expected to be limited.

In this paper, the filtering problem for a large number of agents is formulated and solved as a collective inference problem. Our main goal is to develop the feedback particle filter algorithm to solve the collective filtering problem in continuous-time settings. For this purpose, the Bayes’ formula for collective inference is introduced (following [6]) and compared with the standard Bayes’ formula. The collective Bayes’ formula is specialized to derive the equations of collective filtering in linear Gaussian settings. A feedback particle filter algorithm for this problem is presented and illustrated using numerical simulations.

The outline of the remainder of this paper is as follows. The problem formulation appears in Section [II] The collective Bayes formula and Kalman filter are introduced in Section [III] The FPF algorithm is described in Section [IV] The simulation results appear in Section [V] All the proofs are contained in the Appendix.

II. PROBLEM FORMULATION

A. Dynamic model

1) There are \( M \) agents. The set of agents is denoted as \( \mathcal{M} = \{1, 2, \ldots, M\} \). The set of permutations of \( \mathcal{M} \) is denoted by \( \Pi(\mathcal{M}) \) whose cardinality \( |\Pi(\mathcal{M})| = M! \).

2) Each agent has its own state–observation pair. The state process for \( j \)th agent is \( X^j = \{X^j_t : t \in I\} \), a Markov
process and \( I \) is the index set (time). The associated observation \( Z^t = \{ Z^t_i : i \in I \} \) is modeled for each time \( t \in I \) by the observation kernel for \( P(Z^t | X_t) \).

3) At time \( t \), the observations from all agents are aggregated while their labels are censored through random permutations \( \sigma_i \in \Pi(M) \). The association \( \sigma_i = (\sigma^1_i, \sigma^2_i, \ldots, \sigma^M_i) \) signifies that the \( j^{th} \) observation originates from agent \( \sigma^j_i \). The random permutations are modeled as a Markov process on \( \Pi(M) \).

Three examples are presented next:

**Example 1: Finite state Markov chain**: The index set \( I = \{0, 1, \ldots \} \). The state space \( \mathcal{X} = \{1, 2, \ldots, d\} \) and the observation space \( \mathcal{Y} = \{1, 2, \ldots, m\} \) are both finite. The system is described by state transition \( P(X_{t+1} = x^+ | X_t = x) = p(x^+ | x) \) and observation model \( P(Z_t = z | X_t = x) = o(z | x) \). A typical example is the COVID-19 testing data assimilation problem. The state-space \( \mathcal{X} \) is the set of epidemiological states, e.g., susceptible (S), infected (I), recovered (R) for the so-called SIR Markov chain model. The sensor is the binary-valued output of the PCR test for virema.

**Example 2: Discrete-time linear Gaussian model**: The index set \( I = \{0, 1, \ldots \} \). The state and observation processes are modeled as

\[
\begin{align*}
X_{t+1} &= AX_t + B_t, \quad X_0 \sim \mathcal{N}(m_0, \Sigma_0) \quad (1a) \\
Z_t &= HX_t + W_t \quad (1b)
\end{align*}
\]

where \( \{B_t\}_{t \geq 0}, \{W_t\}_{t \geq 0} \) are mutually independent i.i.d. Gaussian random variables with zero mean and variance \( Q \) and \( R \), respectively, and also assumed to be independent of \( X_0 \).

**Example 3: Continuous-time linear Gaussian model**: The index set \( I = [0, \infty) \) if the positive real line. The state and observation processes are defined by the stochastic differential equation (SDE):

\[
\begin{align*}
dX_t &= AX_t \, dt + Q^{1/2} \, dB_t, \quad X_0 \sim \mathcal{N}(m_0, \Sigma_0) \quad (2a) \\
dZ_t &= HX_t \, dt + R^{1/2} \, dW_t \quad (2b)
\end{align*}
\]

where \( \{B_t\}_{t \geq 0}, \{W_t\}_{t \geq 0} \) are standard Wiener processes. \( Q \) and \( R \) are noise variances. It is assumed that \( X_0, \{B_t\}_{t \geq 0}, \{W_t\}_{t \geq 0} \) are mutually independent. For continuous-time problems involving data association uncertainty, a Markov model of \( \{\sigma_i\}_{i \geq 0} \) appears in [5].

**B. Basic filtering problem**

In its most general form, the filtering problem is to assimilate the measurements \( Z_t = \sigma(Z_t^i : 1 \leq i \leq M, 0 \leq s \leq t) \) to deduce the posterior distribution of the hidden states \( \{X_t^i : 1 \leq i \leq M\} \). Given the associations are also hidden, the problem is solved through building a filter also to estimate the permutation \( \sigma_i \).

A number of approaches have been considered to solve the problem in a tractable fashion: Early approaches included multiple hypothesis testing (MHT) algorithm, requiring exhaustive enumeration [7]. However, exhaustive enumeration leads to an NP-hard problem because number of associations increases exponentially with time. The complexity issue led to development of the probabilistic MHT or its simpler “single-scan” version, the joint probabilistic data association (JPDA) filter [3], [8]. These algorithms are based on computation (or approximation) of the observation-to-target association probability. The feedback particle filter extension of the JPDA filter appears in [5].

**C. Collective filtering problem**

In the limit of large number of non-agent specific observations, it is more tractable to consider directly the empirical distribution of the observations:

\[
q(z) := \frac{1}{M} \sum_{j=1}^{M} \delta_{z_j}(z)
\]

and use it to compute the empirical distribution of the hidden state. Following the work of [6], this is the approach followed in the remainder of this paper.

**III. COLLECTIVE BAYESIAN FILTERING**

**A. One-step estimator for HMM**

We begin by recalling the Bayes’ formula for the discrete-time HMM model introduced in Example 1. The standard filtering problem is to compute the conditional distribution. Let \( \bar{\pi}_t(x) := P(X_t = x | Z_\tau) \) for \( x \in \mathcal{X} \). The solution is given by the nonlinear filter. Given \( \bar{\pi}_t[i] \) and \( Z_{t+1} = z_{t+1} \),

\[
\bar{\pi}_{t+1}[i](x_{t+1}) = \sum_{x_i \in \mathcal{X}} p(x_{t+1} | x_i) \bar{\pi}_t[i](x_i)
\]

The denominator for the estimation step is given by

\[
P(z_{t+1} | Z_t) = \sum_{x_i \in \mathcal{X}} o(z_{t+1} | x_{t+1}) \bar{\pi}_{t+1}[i](x_{t+1})
\]

For collective problem, denote \( \pi_t[i] \) to be the (estimated) empirical distribution of \( X_t \) given the empirical distribution of the observation \( q(z) \) up to time \( \tau \). Given \( \pi_t[i] \), the one-step estimate is given by:

\[
\begin{align*}
\pi_{t+1}[i](x_{t+1}) &= \sum_{x_i \in \mathcal{X}} p(x_{t+1} | x_i) \pi_t[i](x_i) \quad (3a) \\
\pi_{t+1}[i](x_{t+1}) &\propto \sum_{z_{t+1} \in \mathcal{Z}} o(z_{t+1} | x_{t+1}) \pi_t[i](x_{t+1}) q_t(z_{t+1}) \\
\hat{\pi}_{t+1}(z_{t+1}) &= \sum_{x_{t+1} \in \mathcal{X}} o(z_{t+1} | x_{t+1}) \pi_{t+1}[i](x_{t+1}) \quad (3c)
\end{align*}
\]

The justification of this formula appears in Appendix A.

**B. Linear Gaussian case**

The collective Bayes’ formula [3] naturally extends to the continuous state-space settings. It is now specialized to the discrete-time linear Gaussian model [1] introduced in Example 2. It is assumed that the empirical distribution \( q_t \) is approximated by a Gaussian \( \mathcal{N}(\hat{\xi}_t, V_t) \). The collective filtering equations appear in the following proposition whose proof appears in Appendix B.
**Proposition 1:** Consider the collective filtering problem for the discrete-time linear Gaussian model (1). Suppose $q_t = \mathcal{N}(\hat{z}_t, V_t)$ and $\pi_{t|t} = \mathcal{N}(m_t, \Sigma_t)$ are both Gaussian. Then $\pi_{t+1|t}$ and $\pi_{t+1|t+1}$ are also Gaussian whose mean and variance evolve according to the following recursion:

\[ m_{t+1|t} = A m_t \]
\[ \Sigma_{t+1|t} = A \Sigma_t A^T + Q \quad (4a) \]
\[ K_{t+1} = \Sigma_{t+1|t} (H \Sigma_{t+1|t} H^T + R)^{-1} \quad (4c) \]
\[ m_{t+1|t+1} = m_{t+1|t} + K_{t+1} (\hat{z}_{t+1} - H m_{t+1|t}) \quad (4d) \]
\[ \Sigma_{t+1|t+1} = \Sigma_{t+1|t} - K_{t+1} (H \Sigma_{t+1|t} H^T + R - V_{t+1}) K_{t+1}^{-1} \quad (4e) \]

**Remark 1:** The recursion for the mean is exactly the same form as the Kalman filter where $\hat{V}$ now has the meaning of the empirical mean of the observations $\{Z_{t+1}^i : 1 \leq j \leq M\}$. Note however that the equation for the variance also now depends upon $V_t$. This in turn affects the Kalman gain. For a single agent, $\hat{z}_{t+1} = Z_{t+1}^i$ and $V_{t+1} = 0$. Therefore, one obtains the Kalman filter for this case. This was noted also in [6] for the smoothing algorithm presented in their paper.

**Remark 2:** The Gaussian approximation of the empirical distribution $q_t$ is justified only in the large $M$ limit. For large $M$, one again obtains the Kalman filter if $V_{t+1} \to 0$.

**IV. Feedback Particle Filter for Collective Filtering**

In its basic form, the feedback particle filter (FPF) is a controlled interacting particle system algorithm to approximate the solution of the nonlinear filtering problem in continuous-time settings [9]. Although the algorithm was developed for the general nonlinear case, here we only consider the linear Gaussian model (2) introduced in Example 3. As in the discrete-time settings, a Gaussian model is assumed for the aggregated observations $q(z)$. This is formalized by introducing the following model for the observation:

\[ Z_t^i = \hat{z}_t + \int_0^t V_t^{1/2} d\tilde{W}_t^i, \quad 1 \leq j \leq M \]

where $\hat{z}_t = \frac{1}{M} \sum_{j=1}^M Z_t^j$ is the empirical mean and $\tilde{W}_t^i$ is a standard Wiener process introduced to ‘fit’ the empirical distribution to the observations. The coefficient $V_t$ has the meaning of the quadratic variation of the random component.

The following is the counterpart of Prop. 1 in continuous-time settings. Its proof appears in the Appendix C.

**Proposition 2:** Consider the collective filtering problem for the continuous-time linear Gaussian model (1). Suppose $q_t$ is as defined according to (5), and the prior is a Gaussian $\mathcal{N}(m_0, \Sigma_0)$. Then the posterior $\pi_t$ is a Gaussian whose mean $m_t$ and variance $\Sigma_t$ evolve according to:

\[ dm_t = Am_t dt + \Sigma_t H^T R^{-1} (d\hat{z}_t - H m_t dt) \quad (6a) \]
\[ \frac{d}{dt} \Sigma_t = A \Sigma_t + \Sigma_t A^T + Q - \Sigma_t H^T (R^{-1} - R^{-1} V_t R^{-1}) H \Sigma_t \quad (6b) \]

with the initial conditions specified by $m_0$ and $\Sigma_0$. 

**A. Feedback Particle Filter**

The feedback particle filter (FPF) algorithm is designed to sample from the posterior $\pi_t$. For the collective filtering problem, FPF is comprised of $N$ particles $\{X_t^i : 1 \leq i \leq N, t \geq 0\}$ which are defined according to the mean-field model:

\[ dX_t^i = AX_t^i dt + dB^i_t + \tilde{K}_t (d\hat{Z}_t - (\alpha_t H X_t^i + (I - \alpha_t) H \tilde{m}_t) dt) \quad (7) \]

where $\{B_t^i : 1 \leq i \leq N, t \geq 0\}$ are $N$ independent copies of the process noise $\{B_t\}_{t \geq 0}$, $\tilde{K} = \bar{\Sigma} H^T R^{-1}$, $\alpha_t = \frac{1}{2}(I - V_t R^{-1})$, and $\tilde{m}_t$ and $\bar{\Sigma}$ are the mean and the variance, respectively, of $X_t^i$.

The following proposition states the exactness property of the filter. (That is, the mean and variance of the particles exactly match the mean and the variance of the collective filter). Its proof appears in the Appendix D.

**Proposition 3:** Consider the FPF (7). Then its mean $\tilde{m}_t$ and variance $\bar{\Sigma}_t$ evolve according to equations for $m_t$ and $\Sigma_t$, respectively. Consequently, if $\tilde{m}_0 = m_0$ and $\bar{\Sigma}_0 = \Sigma_0$ then

\[ m_t = \tilde{m}_t, \quad \Sigma_t = \bar{\Sigma}_t \quad \forall t > 0 \]

Notice that the particle system (7) is not practical because it requires the mean and variance of the collective $\{X_t^i : 1 \leq i \leq N\}$.

**V. Simulation**

In this section, we simulate the collective filtering algorithm for the linear Gaussian system. There are two objectives: (i) To evaluate the collective filter described in Prop. 2 as the number of agents $M$ increases; (ii) To show the convergence of the estimates using the FPF algorithm (6) as the number of particles $N \to \infty$.

Comparisons of the collective filtering algorithm are made against the gold standard of running independent Kalman filters with known data association. It will also be interesting to compare the results using joint probabilistic data association (JPDA) filter and are planned as part of the continuing research.

The continuous-time system (2) is simulated using the parameters

\[ A = \begin{pmatrix} 0 & 1 \\ -1 & -0.5 \end{pmatrix}, \quad H = \begin{pmatrix} 0 & 1 \end{pmatrix} \]
The process noise covariance $Q = 0.1I$ and the measurement noise covariance $R = 0.7$. The initial condition is sampled from a Gaussian prior with parameters

$$m_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \Sigma_0 = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}$$

The sample path for each agent is generated by using an Euler method of numerical integration with a fixed step size $\Delta t = 0.01$ over the total simulation time interval $[0, 5]$. At each discrete time-step, $q$ is approximated as a Gaussian whose mean and variance are defined as follows:

$$z_t = \frac{1}{M} \sum_{i=1}^{M} Z_t^i$$

$$V_t = \frac{1}{\Delta t (M-1)} \sum_{j=1}^{M} \left( (Z_{t+\Delta t}^j - \bar{Z}_{t+\Delta t}) - (Z_t^j - \bar{Z}_t) \right)^2$$

The comparison is carried out for the following three filtering algorithms:

(i) (KF) This involves simulating $M$ independent Kalman filters for the $M$ sample paths $\{Z_t^j : 0 \leq t \leq 5\}$ for $j = 1, 2, \ldots, M$. The data association is fully known.

(ii) (CKF) This involves simulating the mean and the variance of a single collective Kalman-Bucy filter using the filtering equations 6 in Prop. 2 with $N$ particles.

(iii) (FPF) This involves simulating a single FPF 8 with $N$ particles.

At the terminal time $T$, KF simulation yields $M$ Gaussians (posterior distributions for each of the $M$ independent Kalman filters). We use $m_{T}^{\text{KF}}$ and $\Sigma_{T}^{\text{KF}}$ to denote the mean and the variance of the sum (mixture) of these $M$ Gaussians. The mean and the variance for the CKF is denoted as $m_{T}^{\text{CKF}}$ and $\Sigma_{T}^{\text{CKF}}$, respectively. Similarly, $m_{T}^{\text{FPF}}$ and $\Sigma_{T}^{\text{FPF}}$ are the empirical mean and variance computed using the FPF algorithm with $N$ particles.

Figure 1 depicts the normalized difference for the mean and the variance between (KF) and (CKF), as the number of agents increase from $M = 2$ to $M = 200$. It is observed that both the differences converge to zero as the number of agents increases. Though omitted in the plot, for the $M = 1$ case, these do match exactly (see Remark 1).

Figure 2 depicts the distance between normalized difference for the mean and the variance between (CKF) and (FPF). In this simulation, $M = 30$ is fixed and $N$ varies from 30 to 1000. The plots show that the difference converges to zero as $N$ increases. Therefore, the FPF is able to provide a solution to the collective inference problem.

REFERENCES

[1] Y. Bar-Shalom, F. Daum, and I. Huang, “The probabilistic data association filter,” IEEE Control Systems Magazine, vol. 29, no. 6, pp. 82–100, 2009.
[2] S. Thrun, “Probabilistic robotics,” Communications of the ACM, vol. 45, no. 3, pp. 52–57, 2002.
[3] T. Kirubarajan and Y. Bar-Shalom, “Probabilistic data association techniques for target tracking in clutter,” Proceedings of the IEEE, vol. 92, no. 3, pp. 536–557, 2004.
[4] I. Kyriakides, D. Morrell, and A. Papandreou-Suppappola, “Sequential monte carlo methods for tracking multiple targets with deterministic and stochastic constraints,” IEEE Transactions on Signal Processing, vol. 56, no. 3, pp. 937–948, 2008.
[5] T. Yang and P. G. Mehta, “Probabilistic data association-feedback particle filter for multiple target tracking applications,” Journal of Dynamic Systems, Measurement, and Control, vol. 140, no. 3, 2018.
[6] R. Singh, I. Haasler, Q. Zhang, J. Karlsson, and Y. Chen, “Inference with aggregate data: An optimal transport approach,” arXiv preprint arXiv:2003.13933, 2020.
[7] D. Reid, “An algorithm for tracking multiple targets,” IEEE transactions on Automatic Control, vol. 24, no. 6, pp. 843–854, 1979.
[8] Y. Bar-Shalom and T. E. Fortmann, Tracking and data association / Yaakov Bar-Shalom, Thomas E. Fortmann., ser. Mathematics in science and engineering; v. 179. Boston: Academic Press, 1988.
[9] T. Yang, R. S. Laugesen, P. G. Mehta, and S. P. Meyn, “Multivariable feedback particle filter,” Automatica, vol. 71, pp. 10–23, 2016.
[10] A. Taghvaei and P. G. Mehta, “An optimal transport formulation of the ensemble kalman filter,” IEEE Transactions on Automatic Control, 2020.
APPENDIX

A. Justification of the one-step estimator \([3]\)

\(\pi_{t+1|j}\) is the usual prediction step. For \(\Pi_{t+1|j}\), consider a maximum entropy problem: The argument is joint distribution over \(X_{t+1}\) and \(Z_{t+1}\) and the objective function is the KL divergence:

\[
\min_{\tilde{P}} \sum_{x, z} \tilde{P}(x, z) \log \frac{\tilde{P}(x, z)}{\pi_{t+1|j}(x) o(z | x)}
\]

and the constraint is that the marginal of \(\tilde{P}\) on \(Z_{t+1}\) must be \(q_{t+1}\).

Hence a Lagrange multiplier \(\lambda(z)\) is introduced and the objective function becomes

\[
\sum_{x, z} \tilde{P}(x, z) \log \frac{\tilde{P}(x, z)}{\pi_{t+1|j}(x) o(z | x)} + \sum_{z} \lambda(z) \left( \sum_{x} \tilde{P}(x, z) - q_{t+1}(z) \right)
\]

Differentiate with respect to \(P(x, z)\) yields

\[
\log \frac{\tilde{P}(x, z)}{\pi_{t+1|j}(x) o(z | x)} + 1 + \lambda(z) = 0
\]

The solution is

\[
\tilde{P}(x, z) = \pi_{t+1|j}(x) o(z | x) \exp \left( -1 - \lambda(z) \right)
\]

It is substituted to the constraint;

\[
\sum_{x} \pi_{t+1|j}(x) o(z | x) \exp \left( -1 - \lambda(z) \right) - q_{t+1}(z) = 0
\]

Therefore,

\[
\exp \left( -1 - \lambda(z) \right) = \frac{q_{t+1}(z)}{\sum_{z} \pi_{t+1|j}(x) o(z | x)}
\]

Denote the denominator by \(\xi_{t+1}(z) = \sum_{z} \pi_{t+1|j}(x) o(z | x)\) and collect the result to conclude:

\[
\pi_{t+1|j+1}(z) = \sum_{x} \pi_{t+1|j}(x) o(z | x) q_{t+1}(z) \xi_{t+1}(z)
\]

Therefore, the integrand \(\frac{O(z | x) q_{t+1}(z)}{\xi_{t+1}(z)} \propto \exp \left( -\frac{1}{2} E_{1} \right)\) where

\[
E_{1} = (z-Hx)^{T} R^{-1} (z-Hx) + (z-\hat{z}_{t+1})^{T} V_{t+1}^{-1} (z-\hat{z}_{t+1})
\]

\[
- (z-Hm_{t+1|j})^{T} (\Sigma_{t+1|j} H^{T} + R)^{-1} (z-Hm_{t+1|j})
\]

\[
= (z-c_{0})^{T} C_{0} (z-c_{0}) + (Hx-\hat{z}_{t+1})^{T} C_{1} (Hx-\hat{z}_{t+1})
\]

\[
+ (Hm_{t+1|j} - \hat{z}_{t+1})^{T} C_{2} (Hm_{t+1|j} - \hat{z}_{t+1})
\]

\[
+ (Hx-Hm_{t+1|j})^{T} C_{3} (Hx-Hm_{t+1|j})
\]

where

\[
C_{i} = R^{-1} \left( R^{-1} + V_{t+1}^{-1} - (\Sigma_{t+1|j} H^{T} + R)^{-1} \right)^{-1} V_{t+1}^{-1}
\]

\[
C_{2} = -V_{t+1}^{-1} \left( R^{-1} + V_{t+1}^{-1} - (\Sigma_{t+1|j} H^{T} + R)^{-1} \right)^{-1} (\Sigma_{t+1|j} H^{T} + R)^{-1}
\]

\[
C_{3} = -R^{-1} \left( R^{-1} + V_{t+1}^{-1} - (\Sigma_{t+1|j} H^{T} + R)^{-1} \right)^{-1} (\Sigma_{t+1|j} H^{T} + R)^{-1}
\]

Also, \(\pi_{t+1|j+1}(x) \propto \exp \left( -\frac{1}{2} E_{2} \right)\)

\[
E_{2} = \left( x-m_{t+1|j} \right)^{T} \Sigma_{t+1|j}^{-1} \left( x-m_{t+1|j} \right) + (Hx-\hat{z}_{t+1})^{T} C_{1} (Hx-\hat{z}_{t+1})
\]

\[
+ (Hx-Hm_{t+1|j})^{T} C_{3} (Hx-Hm_{t+1|j})
\]

\[
= x^{T} \left( \Sigma_{t+1|j}^{-1} + H^{T} C_{1} H \right) x
\]

\[
- 2x^{T} \left( \Sigma_{t+1|j}^{-1} m_{t+1|j} + H^{T} C_{1} \hat{z}_{t+1} + H^{T} C_{3} H m_{t+1|j} \right) + \text{(const.)}
\]

Therefore, \(\pi_{t+1|j+1}(x)\) is a Gaussian pdf with mean and variance is given by:

\[
m_{t+1|j+1} = \Sigma_{t+1|j+1}^{-1} \sum_{z} \pi_{t+1|j}(x) m_{t+1|j} \hat{z}_{t+1} + H^{T} C_{3} H m_{t+1|j}
\]

\[
\Sigma_{t+1|j+1} = \Sigma_{t+1|j}^{-1} + H^{T} C_{1} H + \Sigma_{t+1|j}^{-1} H^{T} C_{3} H m_{t+1|j}
\]

B. Proof of Prop. \([7]\)

The collective one-step estimator \([3]\) in continuous state-space settings is:

\[
\pi_{t+1|j+1}(x) \propto \int o(z | x) \pi_{t+1|j}(x) q_{t+1}(z) dz
\]

\[
= \pi_{t+1|j}(x) \int \frac{o(z | x)}{\xi_{t+1}(z)} q_{t+1}(z) dz
\]

\[
= \pi_{t+1|j}(x) o(z | x) \exp \left( -\frac{1}{2} (z-Hx)^{T} R^{-1} (z-Hx) \right)
\]

\[
\xi_{t+1}(z) = \int o(z | x) \pi_{t+1|j}(x) dx
\]

where the probability density is involved instead of probability mass function. Note \(\xi_{t+1}(z)\) is the pdf of a Gaussian with mean \(Hm_{t+1|j}\) and variance \(H \Sigma_{t+1|j} H^{T} + R\), and therefore

\[
o(z | x) \propto \exp \left( -\frac{1}{2} (z-Hx)^{T} R^{-1} (z-Hx) \right)
\]

\[
\xi_{t+1}(z) \propto \exp \left( -\frac{1}{2} (z-Hm_{t+1|j})^{T} (H \Sigma_{t+1|j} H^{T} + R)^{-1} (z-Hm_{t+1|j}) \right)
\]

\[
q_{t+1}(z) \propto \exp \left( -\frac{1}{2} (z-\hat{z}_{t+1})^{T} V_{t+1}^{-1} (z-\hat{z}_{t+1}) \right)
\]

Therefore, the variance \(\Sigma_{t+1|j+1}\) is obtained. It is substituted to \([30]\) to obtain \([34]\).
The variance of the linear system is given by the Lyapunov equation for the mean is obtained as
\[
\frac{d}{dt} \tilde{m}_t = (A - \bar{K}_t \alpha_t H) \tilde{m}_t + \bar{K}_t \alpha_t H (X_t - \tilde{m}_t)
\]

The variance of the linear system is given by the Lyapunov equation
\[
\frac{d}{dt} \tilde{\Sigma}_t = (A - \bar{K}_t \alpha_t H) \tilde{\Sigma}_t + \tilde{\Sigma}_t (A - \bar{K}_t \alpha_t H)^\top + Q
\]

Set \( \alpha_t = \frac{1}{2} (I - V_t R_t^{-1}) \) to conclude the exactness property.