Foundations of Data Science doi:10.3934/fods.2021018

Volume **3**, Number **3**, September **2021** pp. **543–561**

FEEDBACK PARTICLE FILTER FOR COLLECTIVE INFERENCE

Jin-Won Kim

Coordinated Science Laboratory
University of Illinois, Urbana-Champaign
1308 W. Main St.
Urbana, IL 61801, USA

Amirhossein Taghvaei

Department of Mechanical and Aerospace Engineering
University of California, Irvine
4100 Calit2 Building
Irvine, CA 92697-2800, USA

Yongxin Chen

School of Aerospace Engineering
Georgia Institute of Technology
Guggenheim 448B
Atlanta, GA 30332, USA

Prashant G. Mehta*

Coordinated Science Laboratory
University of Illinois, Urbana-Champaign
1308 W. Main St.
Urbana, IL 61801, USA

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 inference. This viewpoint is used to derive the equation for the empirical distribution of the hidden agent states. A feedback particle filter (FPF) algorithm for this problem is presented and illustrated via numerical simulations. Results are presented for the Euclidean and the finite state-space cases, both in continuous-time settings. The classical FPF algorithm is shown to be the special case (with $M = 1$) of these more general results. The simulations help show that the algorithm well approximates the empirical distribution of the hidden states for large M.

2020 *Mathematics Subject Classification.* Primary: 60G35, 62M20; Secondary: 94A12.

Key words and phrases. Feedback particle filter, ensemble Kalman filter, collective inference, data association, collective filtering.

Kim and Mehta are supported in part by the C3.ai Digital Transformation Institute sponsored by C3.ai Inc. and the Microsoft Corporation, and in part by the National Science Foundation grant NSF 1761622. Chen is supported by the NSF 2008513.

* Corresponding author: Prashant G. Mehta.

1. **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 [3, 20]. Consider, e.g., the problem of multiple target tracking 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., [3, 8, 9] and references therein. The feedback particle filter (FPF) algorithm for this problem appears in [22].

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 typically 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 [15, 16, 7].

In a recent important work [17], 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 these algorithms 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 main purpose in 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 FPF algorithm to solve the collective filtering problem in continuous-time settings. For this purpose, the Bayes' formula for collective inference is introduced (following the formulation of [17]) and compared with the standard Bayes' formula. The collective Bayes' formula is specialized to derive the equations of collective filtering in continuous-time settings. The FPF algorithm for this problem is presented for the Euclidean and the finite state-space cases. The classical FPF and ensemble Kalman filter (see [5, 18, 12, 13]) algorithms are shown to be the special case (with $M = 1$) of these more general results. The algorithm is illustrated using numerical simulations.

The outline of the remainder of this paper is as follows. The problem formulation appears in Section 2. The collective Bayes formula and the filter equations are derived in Section 3. The FPF algorithm is described in Section 4. The simulation results appear in Section 5. All the proofs are contained in the Appendix.

2. Problem formulation.

2.1. 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 independent state–observation pair. The state process for the j^{th} agent is $X^j = \{X_t^j : t \in I\}$, a Markov process and I is the index set (time). The associated observation for the j^{th} agent is $Z^j = \{Z_t^j : t \in I\}$.
3. At time t, the observations from all agents is aggregated while their labels are censored through random permutations $\sigma_t \in \Pi(\mathcal{M})$. The association $\sigma_t = (\sigma_t^1, \sigma_t^2, \ldots, \sigma_t^M)$ signifies that the j^{th} observation originates from agent σ_t^j. The random permutations is modeled as a Markov process on $\Pi(\mathcal{M})$.

For each state-observation process, we consider the following types of models.

Discrete-time model. The index set $I = \{0, 1, \ldots\}$. The state and observation space are denoted as $\mathcal{X}$ and $\mathcal{Y}$, respectively. The state and observation processes are modeled as follows:

$$\mathsf{P}(X_{t+1} \in A \mid X_t = x') = \int_A p(x \mid x') \, \mathrm{d}\mu(x)$$
$$\mathsf{P}(Z_t \in B \mid X_t = x') = \int_B o(z \mid x') \, \mathrm{d}\nu(z)$$

where μ and ν are reference measures defined over measurable sets A and B in $\mathcal{X}$ and $\mathcal{Y}$, respectively. Typical choices for the reference measures are the counting measure in discrete state-space settings and the Lebesgue measure in Euclidean settings. The set of joint probability densities on $\mathcal{X} \times \mathcal{Y}$ with respect to $\mu \times \nu$ is denoted as $\mathcal{P}(\mathcal{X}, \mathcal{Y})$.

A topical 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 viremia.

Continuous-time model. The index set $I = [0, \infty)$ and the state process $\{X_t : t \geq 0\}$ is a continuous-time Markov process with the generator denoted as $\mathcal{A}$. The associated adjoint is denoted as $\mathcal{A}^\dagger$. Three special cases of interest are: (i) Ito diffusions in the Euclidean state-space, (ii) the linear Gaussian case, and (iii) the discrete state-space case.

In continuous-time settings, we assume the following stochastic differential equation (SDE) model for scalar-valued observations:

$$\mathrm{d}Z_t = h(X_t) \, \mathrm{d}t + \sigma_w \, \mathrm{d}W_t, \quad Z_0 = 0 \tag{1}$$

where $\{W_t : t \geq 0\}$ is the standard Wiener process and $\sigma_w > 0$. It is assumed that the observation noise is independent of the state process. The reason for restricting the observation model to (1) is because we are primarily interested in deriving the FPF algorithm. The scalar-valued case is considered for notational ease. The generalization to vector-valued observation is straightforward.

2.2. **Standard filtering problem with data association.** In its most general form, the filtering problem is to assimilate the measurements $\mathcal{Z}_t = \sigma(Z_t^i : 1 \leq i \leq M, 0 \leq s \leq t)$ 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 σ_t.

Remark 1. 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 [14]. 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 [8, 4]. 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 [22].

2.3. **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:

$$q_t(z) := \frac{1}{M} \sum_{j=1}^{M} \delta_{Z_t^j}(z)$$

and use it to estimate the empirical distribution of the hidden states – denoted as π_t at time t. The problem is referred to as the *collective filtering* problem.

3. **Collective Bayesian filtering.** As with the standard Bayes' formula, it is easiest to introduce the collective Bayes' formula for the discrete-time model. This is done before presenting the generalization to the continuous-time model.

3.1. **Discrete-time model. Optimization problem:** Given π_t and q_{t+1} the one-step collective inference problem is

$$\underset{\tilde{\mathsf{P}} \in \mathcal{P}(X,Z)}{\text{Minimize}} : \quad \mathsf{D}\big(\tilde{\mathsf{P}} \mid \mathsf{P}\big) \tag{2a}$$

$$\text{Subject to:} \quad \int_{\mathcal{X}} \tilde{\mathsf{P}}(x,z)\,\mathrm{d}\mu(x) = q_{t+1}(z) \tag{2b}$$

where $\mathsf{P}(x,z) = \int_{\mathcal{X}} p(x \mid x')\pi_t(x')\,\mathrm{d}\mu(x')\,o(z|x)$ and $\mathsf{D}(\,\cdot \mid \cdot\,)$ is the K-L divergence.

The justification for considering this type of optimization objective to model the collective inference problem is given in [17, Sec. III-A]. The K-L divergence is the rate function in the large deviation theory and characterizes the exponential decay in probability of observing an empirical distribution [6]. The solution to this problem is described in the following proposition whose proof appears in Appendix A.1.

Proposition 1. *Consider the optimization problem (2). The optimal $\tilde{\mathsf{P}}$ has a density given by the following equations:*

$$\pi_{t+1|t}(x) = \int_{\mathcal{X}} p(x \mid x')\pi_t(x')\,\mathrm{d}\mu(x') \tag{3a}$$

$$\xi_{t+1}(z) = \int_{\mathcal{X}} o(z \mid x)\pi_{t+1|t}(x)\,\mathrm{d}\mu(x) \tag{3b}$$

$$\tilde{\mathsf{P}}(x,z) = \frac{o(z \mid x)\pi_{t+1|t}(x)}{\xi_{t+1}(z)} q_{t+1}(z) \tag{3c}$$

The optimal one-step estimate is the marginal of the optimal $\tilde{\mathsf{P}}$ on $\mathcal{X}$ which is:

$$\pi_{t+1}(x) = \int_{\mathcal{Y}} \tilde{\mathsf{P}}(x,z)\,\mathrm{d}\nu(z) = \int_{\mathcal{Y}} \frac{o(z \mid x)\pi_{t+1|t}(x)}{\xi_{t+1}(z)} q_{t+1}(z)\,\mathrm{d}\nu(z) \tag{4}$$

Remark 2. With $M = 1$, the one-step estimate (4) reduces to the Bayes' formula.

Remark 3. The optimization problem (2) is a special case of the problem introduced in [17] over a time horizon.

Remark 4. An interesting feature of the optimization problem is that q_{t+1} need not necessarily be generated using the transition kernel $o(z \mid x)$. In large M settings, this is important for two reasons:

1. The practical reason is that the transition kernel $o(z \mid x)$ is a nominal model. There is invariably a degree of heterogeneity in observation models for agents.
2. The theoretical reason is simply the nature of the optimization problem at hand: $\tilde{\mathsf{P}}(x,z)$ is the joint distribution that is closest to the joint distribution $\mathsf{P}(x,z)$ of the nominal model and satisfies the marginal constraint. The theoretical reason is similar to the justification used in constructing a Schrodinger bridge.

In the remainder of this paper, we focus on the continuous-time model with observation model (1). For this case, the formula for the collective filter, counterpart of (3), appears in the following subsection. Because of the Remark 4, we present the filter for the more general case where q is arbitrary. The filter formulae are then also specialized to the case where q is generated from the nominal model.

3.2. **Continuous-time model.** In the continuous-time settings, the empirical distribution $q = \{q_t : t \geq 0\}$ of the observations is defined for an increment $\Delta Z_t^j := Z_{t+\Delta t}^j - Z_t^j$ for $j = 1, 2, \ldots, M$. We denote the mean process by:

$$\Delta \hat{Z}_t := \int \Delta z \; \mathrm{d}q_t(\Delta z) = \frac{1}{M}\sum_{j=1}^{M} \Delta Z_t^j$$

Due to the linearity of the definition, we can write $\Delta \hat{Z}_t = \hat{Z}_{t+\Delta t} - \hat{Z}_t$ where $\hat{Z} = \{\hat{Z}_t : t \geq 0\}$ (which itself may be an infinite variation process) is defined by

$$\hat{Z}_t = \frac{1}{M}\sum_{j=1}^{M} Z_t^j$$

The following quantities related to the second-moment are also of interest:

$$\hat{V}_t := \lim_{\Delta t\to 0} \frac{1}{\Delta t}\big(\Delta \hat{Z}_t\big)^2$$

$$V_t := \lim_{\Delta t\to 0} \frac{1}{\Delta t}\int (\Delta z)^2 \,\mathrm{d}q_t(\Delta z) - \hat{V}_t = \lim_{\Delta t\to 0} \frac{1}{M\Delta t}\sum_{j=1}^{M}(\Delta Z_t^j)^2 - \hat{V}_t$$

It is assumed that these limits are well-defined. The first limit represents the quadratic variation of the stochastic process $\{\hat{Z}_t : t \geq 0\}$ and the second limit is an empirical variance of the increments. It is further assumed that the variance of observation $V_t \leq \sigma_w^2$ for all $t \geq 0$.

Optimal recursive update. Using this notation, the continuous time counterpart of the Prop. 1 is as follows. In the statement of the proposition, π_t is a density with respect to the Lebesgue measure. The proof appears in the Appendix A.2.

Proposition 2. *The collective filter* $\{\pi_t : t \geq 0\}$ *solves the following evolution equation:*

$$\mathrm{d}\pi_t(x) = \mathcal{A}^\dagger \pi_t(x)\,\mathrm{d}t + \frac{1}{\sigma_w^2}\pi_t(x)(h(x) - \hat{h}_t)(\,\mathrm{d}\hat{Z}_t - \hat{h}_t\,\mathrm{d}t) + C_t\,\mathrm{d}t \tag{5}$$

where the correction term

$$C_t = \frac{1}{\sigma_w^2}\pi_t(x)\big(g(x) - \hat{g}_t\big)\Big(\frac{V_t + \hat{V}_t}{\sigma_w^2} - 1\Big)$$

and $\hat{h}_t = \int h(x)\pi_t(x)\,\mathrm{d}x$, $g(x) = \frac{1}{2}(h(x) - \hat{h}_t)^2$, $\hat{g}_t = \int g(x)\pi_t(x)\,\mathrm{d}x$. *For the special case where* q *is generated from* (1), *the correction term* $C_t \equiv 0$ *a.s.*

Remark 5. Equation (5) is an example of a parabolic SPDE with a well-developed theory of existence and uniqueness [1]. The first two terms of the right-hand side of (5) are similar to the Kushner-Stratonovich equation with $\mathrm{d}Z_t$ replaced by the empirical mean $\mathrm{d}\hat{Z}_t$. The third term is an additional correction term that arises on account of the second moment of q. If q is generated from the model, the term vanishes because $V_t + \hat{V}_t = \sigma_w^2$. The following special cases are also of interest:

1. $(M = 1)$. The empirical distribution q_t is a Dirac delta measure concentrated on a single trajectory $\hat{Z}_t = Z_t$. In this case the variance $V_t = 0$ and the third term vanishes (using the Ito's rule $\mathrm{d}\hat{Z}_t^2 = \hat{V}_t\,\mathrm{d}t = \sigma_w^2\,\mathrm{d}t$). Therefore, for $M = 1$, the collective filter reduces to the Kushner-Stratonovich equation.
2. $(M = \infty)$. In this case, $\hat{Z}_t$ is a deterministic process $\int_0^t \mathsf{E}\big(h(X_s)\big)\,\mathrm{d}s$ and $V_t = \sigma_w^2$ almost surely. In this case, both the second and third terms of (5) are zero and the collective filter reduces to the Kolomogorov's forward equation.

3.3. **Linear-Gaussian case.** A special case of the continuous-time model is the linear-Gaussian case where the state and observation processes are defined by:

$$\mathrm{d}X_t = AX_t\,\mathrm{d}t + Q^{1/2}\,\mathrm{d}B_t, \quad X_0 \sim \mathcal{N}(m_0, \Sigma_0) \tag{6a}$$

$$\mathrm{d}Z_t = HX_t\,\mathrm{d}t + \sigma_w\,\mathrm{d}W_t \tag{6b}$$

Here the drift term and the observation function are linear and the noise processes and initial condition are Gaussian.

The following is the counterpart of (5) for the linear-Gaussian case. Its proof appears in the Appendix A.3.

Proposition 3. *Consider the collective filtering problem for the continuous-time linear Gaussian model (6). Then π_t is a Gaussian whose mean m_t and variance Σ_t evolve according to*

$$\mathrm{d}m_t = Am_t\,\mathrm{d}t + \frac{1}{\sigma_w^2}\Sigma_t H^\top\big(\,\mathrm{d}\hat{Z}_t - Hm_t\,\mathrm{d}t\big) \tag{7a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\Sigma_t = A\Sigma_t + \Sigma_t A^\top + Q - \frac{1}{\sigma_w^2}\Sigma_t H^\top\Big(1-\frac{V_t}{\sigma_w^2}\Big)H\Sigma_t \tag{7b}$$

with the initial conditions specified by m_0 and Σ_0.

4. Collective feedback particle filter.

The feedback particle filter is a controlled interacting particle system to approximate the solution of (5). In this section, we extend the FPF formulae in [21, 23] to the collective filtering settings.

4.1. Euclidean setting.

The state process is defined by the following SDE:

$$\mathrm{d}X_t = a(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t, \quad X_0\sim\pi_0$$

where $a\in C^1(\mathbb{R}^d;\mathbb{R}^d)$, $\sigma\in C^2(\mathbb{R}^d;\mathbb{R}^{d\times p})$ and $B=\{B_t : t\ge 0\}$ is a standard Wiener process. π_0 denotes the initial distribution of X_0. It is assumed that π_0 has a probability density with respect to the Lebesgue measure. For (1), the observation function $h\in C^2(\mathbb{R}^d;\mathbb{R})$.

Feedback particle filter. The formula for collective FPF is as follows:

$$\begin{aligned}\mathrm{d}X_t^i =& a(X_t^i)\,\mathrm{d}t + \sigma(X_t^i)\,\mathrm{d}B_t^i + K_t(X_t^i)\big(\,\mathrm{d}\hat{Z}_t - \bar{h}_t\,\mathrm{d}t\big) - \alpha_t K_t(X_t^i)\big(h(X_t^i)-\bar{h}_t\big)\,\mathrm{d}t\\ &+\frac{1}{2}\big(K_t^\top\nabla K_t\big)(X_t^i)\hat{V}_t\,\mathrm{d}t + \Big(\frac{V_t+\hat{V}_t}{\sigma_w^2}-1\Big)u_t(X_t^i)\,\mathrm{d}t, \quad X_0^i\sim\bar{\pi}_0\end{aligned} \tag{8}$$

where $B^i=\{B_t^i : t\ge 0\}$ is an independent copy of the process noise; and $\alpha_t := \frac{1}{2}(1-\frac{V_t}{\sigma_w^2})$. The gain functions $K_t=\nabla\phi_t$ and $u_t = -\frac{1}{2}K_t(h-\bar{h}_t)+\nabla\psi_t$ where ϕ and ψ are chosen to solve Poisson equations:

$$-\nabla\cdot(\bar{\pi}_t\nabla\phi_t) = \frac{1}{\sigma_w^2}\bar{\pi}_t(h-\bar{h}_t) \tag{9a}$$

$$-\nabla\cdot(\bar{\pi}_t\nabla\psi_t) = \frac{1}{\sigma_w^2}\bar{\pi}_t(g-\bar{g}_t) \tag{9b}$$

where $\bar{\pi}_t$ denotes the distribution of X_t^i and $\bar{h}_t=\int\bar{\pi}_t(x)h(x)\,\mathrm{d}x$, $\bar{g}_t=\int\bar{\pi}_t(x)g(x)\,\mathrm{d}x$.

Remark 6. The final term in the FPF (8) is identically zero in the special case where q is generated from the observation model (1) (see Remark 5). In this case, the algorithm is similar to the classical FPF [21]. In particular, one need not solve (9b) to compute u_t.

Remark 7. The aim of the paper is to present a representation of the filter. The representation is not unique. For example, one may add any drift term "$v_t(X_t^i)\,\mathrm{d}t$" to the right-hand side of (8) such that $\nabla\cdot(\bar{\pi}_t v_t)=0$. There are two technical issues pertaining to the well-posedness of the collective FPF: (i) Existence and uniqueness of the process $\{\bar{\pi}_t : t\ge 0\}$; and (ii) Existence and uniqueness of the solution of the Poisson equation (9). These are simply assumed to hold in the paper. A unique weak solution of the Poisson equation exists if $\bar{\pi}_t$ satisfies Poincaré inequality, and $\int h^2\bar{\pi}_t\,\mathrm{d}x<\infty$, $\int g^2\bar{\pi}_t\,\mathrm{d}x<\infty$ [10, Theorem 2.2]. A probability distribution satisfies the Poincaré inequality when it has Gaussian decay [2]. In this paper, we assume $\bar{\pi}_t$ satisfies Poincaré inequality and $\int h^4\bar{\pi}_t\,\mathrm{d}x<\infty$ for all $t\ge 0$ to ensure both Poisson equations (9) are well-posed. It is possible to prove Poincaré inequality for the solution of the collective filter $\bar{\pi}_t$ by extending the existing results that establish the Poincaré inequality for the nonlinear filtering in certain settings [11, Lemma 5.1], [10, Prop 2.1].

The following proposition states the exactness property of the filter. (That is, the distribution of the particles exactly matches the distribution of the collective filter). The proof appears in Appendix A.5.

Proposition 4. *Consider the FPF* (8) *and suppose its probability density function* $\{\bar{\pi}_t : t \geq 0\}$ *exists and is unique. Then it evolves according to the equation* (5). *Consequently, if* $\bar{\pi}_0 = \pi_0$, *then*

$$\bar{\pi}_t = \pi_t \quad \forall t \geq 0$$

Remark 8. Noting that the term $\frac{1}{2}K^\top \nabla K$ is the Wong-Zakai correction term, it is useful to express the collective FPF in its Stratonovich form:

$$\begin{aligned} \mathrm{d}X_t^i &= a(X_t^i)\,\mathrm{d}t + \sigma(X_t^i)\,\mathrm{d}B_t^i + K_t(X_t^i) \circ \big(\mathrm{d}\hat{Z}_t - \alpha_t h(X_t^i)\,\mathrm{d}t - (1-\alpha_t)\bar{h}_t\,\mathrm{d}t\big) \\ &\quad + \Big(\frac{V_t + \hat{V}_t}{\sigma_w^2} - 1\Big) u_t(X_t^i)\,\mathrm{d}t + v_t(X_t^i)\hat{V}_t\,\mathrm{d}t, \quad X_0^i \sim \bar{\pi}_0 \end{aligned} \tag{10}$$

Depending upon the interpretation of $\circ$, there are two definitions of the drift term v_t:

1. In the classical derivation of the FPF, the gain function $K_t(x)$ is interpreted as a function of space x and time t, and the $\circ$ in the Stratonovich form is interpreted *only* with respect to the space x. Using this interpretation, $v_t = 0$.
2. In a recent paper [11, Sec. 3], the gain function is defined and interpreted as a function of space x and the density. This is natural because the dependence upon time t comes because of the changes in density ($\bar{\pi}_t$) as the time evolves. Because the density is a stochastic process, it is argued that the appropriate interpretation of $\circ$ in the Stratonovich form should involve *both* space x and the density. Using such an interpretation, $v_t(x) = \nabla\varphi_t(x)$ where φ_t satisfies the following Poisson equation whose derivation appears in Appendix A.4:

$$\nabla \cdot (\bar{\pi}_t \nabla \varphi_t) = \frac{1}{\sigma_w^2}\bar{\pi}_t\big(\frac{1}{2}K_t^\top \nabla h - \frac{1}{\sigma_w^2}\bar{g}_t\big)$$

Remark 9. In a special case of $M = 1$, the variance $V_t = 0$ and $\hat{Z} = Z^1$ is a single trajectory, so $\hat{V}_t = \sigma_w^2$. Therefore (10) with $v_t = 0$ becomes:

$$\mathrm{d}X_t^i = a(X_t^i)\,\mathrm{d}t + \sigma(X_t^i)\,\mathrm{d}B_t^i + K_t(X_t^i) \circ \big(\mathrm{d}\hat{Z}_t - \frac{h(X_t^i) - \bar{h}_t}{2}\,\mathrm{d}t\big), \quad X_0^i \sim \bar{\pi}_0$$

where $\circ$ is interpreted only with respect to the space. This is precisely the FPF on Euclidean case appear in [21, Remark 1].

Note that FPF (8) is not practical because it requires the distribution $\bar{\pi}_t$. In practice, these are approximated by finite N number of particles. In this case, for $i = 1, \ldots, N$, X_0^i is N iid samples from $\bar{\pi}_0$ and noise processes B^i is also N copies of independent Wiener processes. In this case,

$$\bar{h}_t^{(N)} := \frac{1}{N}\sum_{i=1}^N h(X_t^i), \quad \bar{g}_t^{(N)} := \frac{1}{N}\sum_{i=1}^N g(X_t^i)$$

and the Poisson equations are also accordingly constructed. In the limit as $N \to \infty$, the empirical distribution converges to $\bar{\pi}$ because of the propagation of chaos property of the interacting particle systems [19].

4.2. Linear-Gaussian case.

In linear-Gaussian case (6), the distribution is Gaussian and completely characterized by its mean $\bar{m}_t$ and the varriance $\bar{\Sigma}_t$. In this case, the explicit formulae for the solution of the Poisson equations can be obtained, and it is stated in the following lemma.

Lemma 4.1. *In linear-Gaussian case,* $K_t(\cdot) = \frac{1}{\sigma_w^2}\bar{\Sigma}_t H^\top$ *and* $u_t(\cdot) = 0$ *solves the Poisson equations* (9).

The proof is straightforward by directly using (9). Consequently, the FPF algorithm for the linear-Gaussian case is as follows:

$$\begin{aligned} \mathrm{d}X_t^i &= AX_t^i \,\mathrm{d}t + \mathrm{d}\bar{B}_t^i + \bar{K}_t\big(\mathrm{d}\hat{Z}_t - (\alpha_t H X_t^i + (I-\alpha_t)H\bar{m}_t)\,\mathrm{d}t\big) \\ X_0^i &\overset{\text{iid}}{\sim} \mathcal{N}(m_0, \Sigma_0) \end{aligned} \tag{11}$$

where $\{\bar{B}_t^i : 1 \le i \le N, t \ge 0\}$ are N independent copies of the process noise $\{B_t\}_{t\ge 0}$, $\bar{K}_t = \frac{1}{\sigma_w^2}\bar{\Sigma}_t H^\top$, $\alpha_t = \frac{1}{2}(I - \frac{V_t}{\sigma_w^2})$. It is readily seen that the mean and the variance of the particles evolves exactly according to (7).

4.3. Finite-state case.

Consider the continuous-time Markov chain $\{X_t : t \ge 0\}$ defined on the finite state space case $\mathcal{X} := \{e_1, e_2, \dots, e_d\}$ where e_k are standard bases on $\mathbb{R}^d$ for $k = 1, \dots, d$. The dynamics of the Markov chain is given by

$$\mathrm{d}X_t = \sum_{x\in\mathcal{X}, x\neq X_t} (x - X_t)\,\mathrm{d}\zeta_t^{x,X_t}$$

where $\zeta_t^{x,y}$ is a Poisson processes with rate $r_{x,y}$ for $x, y \in \mathcal{X}$. A count from Poisson process ζ_t^{x,X_t} causes the Markov-chain to jump to state x. The observation model is (1). The FPF update law for the finite-state case is as follows:

$$\begin{aligned} \mathrm{d}X_t^i = &\sum_{x\in\mathcal{X}, x\neq X_t^i} (x - X_t^i)\,\mathrm{d}\zeta_t^{x,X_t^i} + \sum_{x\in\mathcal{X}, x\neq X_t^i} (x - X_t^i)\,\mathrm{d}\ell_t^{x,X_t^i} \\ &+ \sum_{x\in\mathcal{X}, x\neq X_t^i} (x - X_t^i)\,\mathrm{d}\tilde{\ell}_t^{x,X_t^i}, \quad X_0^i \sim \pi_0 \end{aligned} \tag{12}$$

where $\ell_t^{x,y}, \tilde{\ell}_t^{x,y}$ are time-modulated Poisson processes of the following form

$$\ell_t^{x,y} = N^{x,y}(U_t^x), \quad \tilde{\ell}_t^{x,y} = \tilde{N}^{x,y}(\tilde{U}_t^x)$$

Here, $N^{x,y}(\cdot)$ and $\tilde{N}^{x,y}(\cdot)$ are standard Poisson processes with rate equal to one. The inputs U_t^x and $\tilde{U}_t^x$ are defined according to

$$\mathrm{d}U_t^x = K_t(x)(\,\mathrm{d}\hat{Z}_t - \bar{h}_t\,\mathrm{d}t), \quad \mathrm{d}\tilde{U}_t^x = \tilde{K}_t(x)\Big(\frac{V_t + \hat{V}_t}{\sigma_w^2} - 1\Big)\,\mathrm{d}t$$

where the gain vectors $K(\cdot)$ and $\tilde{K}(\cdot)$ solve the finite-state space counterpart of the Poisson equations (9)

$$K_t(x) - \bar{\pi}_t(x) \sum_{y\in\mathcal{X}} K_t(y) = \frac{1}{\sigma_w^2} \bar{\pi}_t(x)(h(x) - \bar{h}_t)$$
$$\tilde{K}_t(x) - \bar{\pi}_t(x) \sum_{y\in\mathcal{X}} \tilde{K}_t(y) = \frac{1}{\sigma_w^2} \bar{\pi}_t(x)(g(x) - \bar{g}_t)$$

The general form of the solution is explicitly known:

$$K_t(x) = \frac{1}{\sigma_w^2} \bar{\pi}_t(x)(h(x) - c), \quad \tilde{K}_t(x) = \frac{1}{\sigma_w^2} \bar{\pi}_t(x)(g(x) - \tilde{c})$$

where c and $\tilde{c}$ are constants. The constants are chosen so that U_t^x and $\tilde{U}_t^x$ are non-decreasing leading to a well-posed Poisson processes $N^{x,y}(U_t^x)$ and $\tilde{N}^{x,y}(\tilde{U}_t^x)$. In particular,

$$c = \begin{cases} \min_x h(x), & \text{if} \quad \mathrm{d}\hat{Z}_t - \bar{h}_t \,\mathrm{d}t \geq 0 \\ \max_x h(x), & \text{else} \end{cases}, \quad \tilde{c} = \begin{cases} \min_x g(x), & \text{if} \quad (\frac{V_t + \hat{V}_t}{\sigma_w^2} - 1) \geq 0 \\ \max_x g(x), & \text{else} \end{cases}$$

Remark 10. The FPF for finite state-space Markov chain is proposed in [23]. It simulates the Wonham filter. Notice that the first line of (12) has the same structure with the algorithm proposed in [23] and it is indeed identical when $M = 1$.

5. **Simulations.** In this section, we simulate the collective filtering algorithm for a simple linear-Gaussian system. There are two objectives: (i) To evaluate the collective filter described in Prop. 3 as the number of agents M increases; (ii) To show the convergence of the estimates using the FPF algorithm as the number of particles $N \to \infty$. In order to avoid using any other approximation, the mean and the variance from the each algorithms are directly compared.

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 this is planned as part of the continuing research.

The continuous-time system (6) 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 $\sigma_w^2 = 0.7$. 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

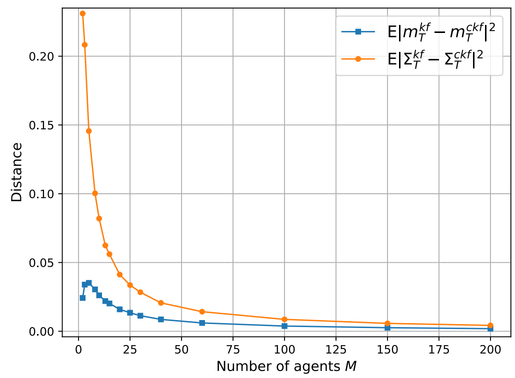

FIGURE 1. Normalized error for mean (blue circle) and variance (orange circle) with the KF and CKF algorithms. The KF algorithms were run as M independent Kalman filters with fully known data associations.

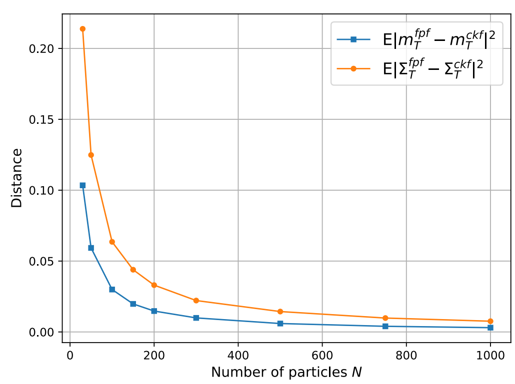

FIGURE 2. Normalized error for mean (blue circle) and variance (orange circle) with the CKF and FPF algorithms. The number of agents is fixed to $M = 30$ for this simulation.

variance are defined as follows:

$$\hat{Z}_t = \frac{1}{M}\sum_{j=1}^{M} Z_t^j$$

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

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

- (KF) This involves simulating M independent Kalman-Bucy filters for the M sample paths $\{Z_t^j : 0 \le t \le 5\}$ for $j = 1, 2, \ldots, M$. The data association is fully known.
- (CKF) This involves simulating the mean and the variance of a single collective Kalman-Bucy filter using the filtering equations (7) in Prop. 3.

- (FPF) This involves simulating a single FPF (11) with N particles.

At the terminal time T, KF simulation yields M Gaussians (posterior distributions for each of the M independent Kalman filters) whose mean and variance are $m_T^{(j)}$ and $\Sigma_T^{(j)}$, respectively for $j = 1, 2, \ldots, M$. We use m_T^{kf} and Σ_T^{kf} to denote the mean and the variance of the sum (mixture) of these M Gaussians. Note the mean and the variance is computed by:

$$m_T^{\text{kf}} = \frac{1}{M}\sum_{j=1}^{M} m_T^{(j)}$$

$$\Sigma_T^{\text{kf}} = \frac{1}{M}\sum_{j=1}^{M} \Sigma_T^{(j)} + \big(m_T^{(j)} - m_T^{\text{kf}}\big)\big(m_T^{(j)} - m_T^{\text{kf}}\big)^\top$$

The mean and the variance for the CKF is denoted as m_T^{ckf} and Σ_T^{ckf}, respectively. Similarly, m_T^{fpf} and Σ_T^{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.

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.

Appendix A. Proofs of the propositions.

A.1. Proof of Proposition 1.

Eq. (3a) for $\pi_{t+1|t}$ is the usual prediction step, and the nominal distribution $\mathsf{P}(x,z) = \pi_{t+1|t}(x)o(z \mid x)$. Therefore the optimization problem becomes

$$\min_{\tilde{\mathsf{P}}} \iint_{\mathcal{X}\times\mathcal{Y}} \tilde{\mathsf{P}}(x,z)\log\frac{\tilde{\mathsf{P}}(x,z)}{\pi_{t+1|t}(x)o(z\mid x)}\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(z)$$

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

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

$$\text{Minimize:}\quad \iint_{\mathcal{X}\times\mathcal{Y}} \tilde{\mathsf{P}}(x,z)\log\frac{\tilde{\mathsf{P}}(x,z)}{\pi_{t+1|t}(x)o(z\mid x)}\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(z) + \int_{\mathcal{Y}} \lambda(z)\Big(\int_{\mathcal{X}} \tilde{\mathsf{P}}(x,z)\,\mathrm{d}\mu(x) - q_{t+1}(z)\Big)\,\mathrm{d}\nu(z)$$

Differentiate with respect to $\tilde{\mathsf{P}}(x,z)$ yields

$$\log\frac{\tilde{\mathsf{P}}(x,z)}{\pi_{t+1|t}(x)o(z\mid x)} + 1 + \lambda(z) = 0$$

The solution is

$$\tilde{\mathsf{P}}(x,z) = \pi_{t+1|t}(x)o(z\mid x)\exp\big(-1-\lambda(z)\big)$$

It is substituted to the constraint:

$$\int_{\mathcal{X}} \pi_{t+1|t}(x) o(z \mid x) \, \mathrm{d}\mu(x) \exp\big(-1-\lambda(z)\big) - q_{t+1}(z) \, \mathrm{d}\mu(x) = 0$$

Therefore,

$$\exp\big(-1-\lambda(z)\big) = \frac{q_{t+1}(z)}{\int_{\mathcal{X}} \pi_{t+1|t}(x) o(z \mid x) \, \mathrm{d}\mu(x)}$$

Denote the denominator by $\xi_{t+1}(z) = \int_{\mathcal{X}} \pi_{t+1|t}(x) o(z \mid x) \, \mathrm{d}\mu(x)$ and collect the result to conclude:

$$\pi_{t+1}(x) = \int_{\mathcal{Y}} \tilde{\mathsf{P}}(x,z) \, \mathrm{d}\nu(z) = \int_{\mathcal{Y}} \frac{\pi_{t+1|t}(x) o(z \mid x) q_{t+1}(z)}{\xi_{t+1}(z)} \, \mathrm{d}\nu(z)$$

□

A.2. **Proof of Proposition 2.** The proof is given for the Euclidean case where the reference measures μ and ν are both Lebesgue measures. Let Δt be the discrete time-step. From time $t \to t + \Delta t$, the likelihood function is approximated as a Gaussian density

$$o(Z_{t+\Delta t} - Z_t | X_t = x) \propto \exp\Big(-\frac{|Z_{t+\Delta t} - Z_t - h(x)\Delta t|^2}{2\sigma_w^2 \Delta t}\Big)$$

Denoting $\Delta Z := Z_{t+\Delta t} - Z_t$, upon suitably adapting the formulae from Proposition 1,

$$\pi_{t+\Delta t}(x) - \pi_t(x) = \mathcal{A}^\dagger \pi_t(x) \Delta t + \int_{\mathcal{Y}} \frac{\exp(-\frac{|\Delta Z - h(x)\Delta t|^2}{2\sigma_w^2 \Delta t}) \pi_t(x)}{\xi_t(\Delta Z)} \, \mathrm{d}q_t(\Delta Z) + O(\Delta t^2)$$

where

$$\xi_t(\Delta Z) = \int_{\mathcal{X}} \exp\big(-\frac{|\Delta Z - h(x)\Delta t|^2}{2\sigma_w^2 \Delta t}\big) \pi_t(x) \, \mathrm{d}x$$

Note that the term $\exp(-\frac{\Delta Z^2}{2\sigma_w^2 \Delta t})$ appears both in numerator and denominator and therefore cancel. For the other terms, we use the expansion keeping the terms up to second-order for ΔZ and first-order for Δt. Since $\exp(x) = 1 + x + \frac{1}{2}x^2 + O(x^3)$, the numerator inside the integral becomes:

$$\begin{aligned} &\exp\Big(-\frac{h(x)^2 \Delta t - 2h(x)\Delta Z}{2\sigma_w^2}\Big) \pi_t(x) \\ &= (1 - \frac{1}{2\sigma_w^2} h(x)^2 \Delta t + \frac{1}{\sigma_w^2} h(x) \Delta Z + \frac{1}{2\sigma_w^4} h(x)^2 \Delta Z^2 + O(\Delta t^2, \Delta Z^3)) \pi_t(x) \end{aligned}$$

Similarly the denominator is also expressed by:

$$\int_{\mathcal{X}} \big(1 - \frac{1}{2\sigma_w^2} h(x)^2 \Delta t + \frac{1}{\sigma_w^2} h(x) \Delta Z + \frac{1}{2\sigma_w^4} h(x)^2 \Delta Z^2 + O(\Delta t^2, \Delta Z^3)\big) \pi_t(x) \, \mathrm{d}x$$

We use $\frac{1+x}{1+y} = 1 + x - y + y^2 - xy + O(xy^2, y^3)$ to express the ratio, and then it simplifies to

$$\begin{aligned} \pi_{t+\Delta t}(x) - \pi_t(x) = \mathcal{A}^\dagger \pi_t(x) \Delta t &+ \int_{\mathcal{Y}} \frac{1}{\sigma_w^2} \pi_t(x) (h(x) - \hat{h}_t)(\Delta Z - \hat{h}_t \Delta t) \, \mathrm{d}q_t(\Delta Z) \\ &+ \int_{\Delta Z} \frac{1}{\sigma_w^2} \pi_t(x) (g(x) - \hat{g}_t)(\frac{1}{\sigma_w^2} \Delta Z^2 - \Delta t) \, \mathrm{d}q_t(\Delta Z) \\ &+ O(\Delta t^2, \Delta Z^3) \end{aligned}$$

Take the limit as $\Delta t \to 0$ to conclude

$$\mathrm{d}\pi_t(x) = \mathcal{A}^\dagger \pi_t(x)\,\mathrm{d}t + \frac{1}{\sigma_w^2}\pi_t(x)(h(x)-\hat{h}_t)(\,\mathrm{d}\hat{Z}_t - \hat{h}_t\,\mathrm{d}t) + \underbrace{\frac{1}{\sigma_w^2}\pi_t(x)\big(g(x)-\hat{g}_t\big)\Big(\frac{V_t+\hat{V}_t}{\sigma_w^2}-1\Big)}_{=:C_t}\,\mathrm{d}t$$

If q is generated from M i.i.d. copies of the model (1), then

$$V_t + \hat{V}_t = \lim_{\Delta t \to 0} \frac{1}{M\Delta t}\sum_{j=1}^{M}(\Delta Z_t^j)^2 = \sigma_w^2$$

almost surely, and $C_t = 0$ in this case.

The formulae for discrete state-space models are entirely identical. The integral over $\mathcal{X}$ is replaced by a summation. □

A.3. **Proof of Proposition 3.** For linear-Gaussian example, the mean and the variance fully characterize the distribution. Thus, we repeat the procedure in 3.2 for the linear-Gaussian case. Although the linear Gaussian is a special case of the general Euclidean result, we provide here a proof as a continuous limit of a discrete-time model (which is of independent interest). The results are stated and proved in somewhat more general settings with vector-valued observations.

A.3.1. *Discrete-time linear-Gaussian problem.* The model is

$$X_{t+1} = AX_t + B_t, \quad X_0 \sim \mathcal{N}(m_0, \Sigma_0) \tag{13a}$$
$$Z_t = HX_t + W_t \tag{13b}$$

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. The observation q_t is assumed to be a Gaussian with mean $\hat{Z}_t$ and variance V_t. The discrete-time update (3) for the mean and the variance of linear-Gaussian problem is illustrated in the following proposition.

Proposition 5. *Consider the collective filtering problem for the discrete-time linear Gaussian model (13). Suppose $q_t = \mathcal{N}(\hat{Z}_t, V_t)$ and $\pi_t = \mathcal{N}(m_t, \Sigma_t)$ are both Gaussian. It is assumed that $R - V_t$ is positive semi-definite. Then $\pi_{t+1|t}$ and π_{t+1} are also Gaussian whose mean and variance evolve according to the following recursion:*

$$m_{t+1|t} = Am_t \tag{14a}$$
$$\Sigma_{t+1|t} = A\Sigma_t A^\top + Q \tag{14b}$$
$$K_{t+1} = \Sigma_{t+1|t}H^\top (H\Sigma_{t+1|t}H^\top + R)^{-1} \tag{14c}$$
$$m_{t+1} = m_{t+1|t} + K_{t+1}(\hat{Z}_{t+1} - Hm_{t+1|t}) \tag{14d}$$
$$\Sigma_{t+1} = \Sigma_{t+1|t} - K_{t+1}(H\Sigma_{t+1|t}H^\top + R - V_{t+1})K_{t+1}^\top \tag{14e}$$

Proof. Recall the one-step estimator (3)-(4) in Euclidean state-space settings:

$$\begin{aligned}\pi_{t+1}(x) &= \int \frac{o(z \mid x)\pi_{t+1|t}(x)}{\xi(z)} q_{t+1}(z)\,\mathrm{d}z \\ &= \pi_{t+1|t}(x) \int \frac{o(z \mid x)}{\xi(z)} q_{t+1}(z)\,\mathrm{d}z \\ \xi(z) &= \int o(z \mid x)\pi_{t+1|t}(x)\,\mathrm{d}x\end{aligned}$$

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

$$\begin{aligned}o(z \mid x) &\propto \exp\Big(-\frac{1}{2}(z - Hx)^\top R^{-1}(z - Hx)\Big) \\ \xi(z) &\propto \exp\Big(-\frac{1}{2}(z - Hm_{t+1|t})^\top (H\Sigma_{t+1|t}H^\top + R)^{-1}(z - Hm_{t+1|t})\Big) \\ q_{t+1}(z) &\propto \exp\Big(-\frac{1}{2}(z - \hat{Z}_{t+1})^\top V_{t+1}^{-1}(z - \hat{Z}_{t+1})\Big)\end{aligned}$$

Therefore, the integrand $\frac{o(z|x)q_{t+1}(z)}{\xi(z)} \propto \exp\big(-\frac{1}{2}E_1\big)$ where

$$\begin{aligned}E_1 = &(z - Hx)^\top R^{-1}(z - Hx) + (z - \hat{Z}_{t+1})^\top V_{t+1}^{-1}(z - \hat{Z}_{t+1}) \\ &- (z - Hm_{t+1|t})^\top (H\Sigma_{t+1|t}H^\top + R)^{-1}(z - Hm_{t+1|t})\end{aligned}$$

Since we will integrate over z, any perfect square of z becomes a constant. Therefore, we complete the square terms on z,

$$\begin{aligned}E_1 &= z^\top \big(R^{-1} + V_{t+1}^{-1} + (H\Sigma_{t+1|t}H^\top + R)^{-1}\big)z \\ &\quad - 2z^\top \big(R^{-1}Hx + V_{t+1}^{-1}\hat{Z}_{t+1} + (H\Sigma_{t+1|t}H^\top + R)^{-1}Hm_{t+1|t}\big) + (\cdots) \\ &= (z - C_0^{-1}c_0(x))^\top C_0(z - C_0^{-1}c_0(x)) - c_0(x)^\top C_0^{-1}c_0(x) + (\cdots)\end{aligned}$$

where

$$\begin{aligned}C_0 &= R^{-1} + V_{t+1}^{-1} + (H\Sigma_{t+1|t}H^\top + R)^{-1} \\ c_0(x) &= R^{-1}Hx + V_{t+1}^{-1}\hat{Z}_{t+1} + (H\Sigma_{t+1|t}H^\top + R)^{-1}Hm_{t+1|t}\end{aligned}$$

Collect non-constant term of $c_0(x)^\top C_0^{-1}c_0(x)$ and remaining term of E_1, one can obtain

$$\begin{aligned}E_1 = &(Hx - \hat{Z}_{t+1})^\top C_1(Hx - \hat{Z}_{t+1}) \\ &+ (Hm_{t+1|t} - \hat{Z}_{t+1})^\top C_2(Hm_{t+1|t} - \hat{Z}_{t+1}) \\ &+ (Hx - Hm_{t+1|t})^\top C_3(Hx - Hm_{t+1|t}) \\ &+ (z - C_0^{-1}c_0(x))^\top C_0(z - C_0^{-1}c_0(x)) + (\text{const.})\end{aligned}$$

where

$$\begin{aligned}C_1 &= R^{-1}\big(R^{-1} + V_{t+1}^{-1} - (H\Sigma_{t+1|t}H^\top + R)^{-1}\big)^{-1}V_{t+1}^{-1} \\ C_2 &= -V_{t+1}^{-1}\big(R^{-1} + V_{t+1}^{-1} - (H\Sigma_{t+1|t}H^\top + R)^{-1}\big)^{-1}(H\Sigma_{t+1|t}H^\top + R)^{-1} \\ C_3 &= -R^{-1}\big(R^{-1} + V_{t+1}^{-1} - (H\Sigma_{t+1|t}H^\top + R)^{-1}\big)^{-1}(H\Sigma_{t+1|t}H^\top + R)^{-1}\end{aligned}$$

Also, $\pi_{t+1}(x) \propto \exp\big(-\frac{1}{2}E_2\big)$ where

$$\begin{aligned}E_2 &= (x - m_{t+1|t})^\top \Sigma_{t+1|t}^{-1}(x - m_{t+1|t}) + (Hx - \hat{Z}_{t+1})^\top C_1 (Hx - \hat{Z}_{t+1}) \\ &\quad + (Hx - Hm_{t+1|t})^\top C_3 (Hx - Hm_{t+1|t}) \\ &= x^\top \big(\Sigma_{t+1|t}^{-1} + H^\top (C_1 + C_3) H\big) x \\ &\quad - 2x^\top \big(\Sigma_{t+1|t}^{-1} m_{t+1|t} + H^\top C_1 \hat{Z}_{t+1} + H^\top C_3 H m_{t+1|t}\big) + (\text{const.})\end{aligned}$$

Therefore, π_{t+1} is a Gaussian pdf with mean and variance is given by:

$$m_{t+1} = \Sigma_{t+1}\big(\Sigma_{t+1|t}^{-1} m_{t+1|t} + H^\top C_1 \hat{Z}_{t+1} + H^\top C_3 H m_{t+1|t}\big) \tag{15}$$

$$\Sigma_{t+1} = \big(\Sigma_{t+1|t}^{-1} + H^\top (C_1 + C_3) H\big)^{-1} \tag{16}$$

By the matrix inversion lemma, the variance formula becomes

$$\Sigma_{t+1} = \Sigma_{t+1|t} - \Sigma_{t+1|t} H^\top \big((C_1 + C_3)^{-1} + H\Sigma_{t+1|t} H^\top\big)^{-1} H \Sigma_{t+1|t}$$

Observe that

$$\begin{aligned}C_1 + C_3 &= R^{-1}\big(R^{-1} + V_{t+1}^{-1} - (H\Sigma_{t+1|t}H^\top + R)^{-1}\big)^{-1} \\ &\qquad \cdot \big(V_{t+1}^{-1} - (H\Sigma_{t+1|t}H^\top + R)^{-1}\big) \\ &= \big(R + (V_{t+1}^{-1} - (H\Sigma_{t+1|t}H^\top + R)^{-1})^{-1}\big)^{-1}\end{aligned}$$

Therefore

$$\begin{aligned}&(C_1 + C_3)^{-1} + H\Sigma_{t+1|t}H^\top \\ &\quad = R - ((H\Sigma_{t+1|t}H^\top + R)^{-1} - V_{t+1}^{-1})^{-1} + H\Sigma_{t+1|t}H^\top \\ &\quad = -V_{t+1}(H\Sigma_{t+1|t}H^\top + R - V_{t+1})^{-1}(H\Sigma_{t+1|t}H^\top + R) \\ &\qquad + (H\Sigma_{t+1|t}H^\top + R) \\ &\quad = (H\Sigma_{t+1|t}H^\top + R)(H\Sigma_{t+1|t}H^\top + R - V_{t+1})^{-1} \\ &\qquad \cdot (H\Sigma_{t+1|t}H^\top + R)\end{aligned}$$

Substituting back to the variance equation (16), the Riccati equation (14e) is obtained. It is substituted to (15) to obtain (14d). □

A.3.2. *Proof of Proposition 3.* The previous proposition is extended to continuous-time problem by considering suitable limits.

Consider the continuous-time system (6) with a discrete time-step Δt,

$$\begin{aligned}X_{t+\Delta t} &= (I + A\Delta t)X_t + \Delta B_t \\ Z_{t+\Delta t} &= Z_t + HX_t\Delta t + \Delta W_t\end{aligned}$$

where ΔB_t and ΔW_t are normal random variables with variance $Q\Delta t$, $\sigma_w^2 I \Delta t$, respectively. $Z_{t+\Delta t} - Z_t$ is assumed to be a normal random variable and its mean and variance are $\Delta\hat{Z}_t := \hat{Z}_{t+\Delta t} = \hat{Z}_t$, and $V_t\Delta t$ respectively.

By the Proposition 5, the prediction step is:

$$\begin{aligned}m_{t+\Delta t|t} &= (I + A\Delta t)m_t \\ \Sigma_{t+\Delta t|t} &= (I + A\Delta t)\Sigma_t(I + A\Delta t)^\top + Q\Delta t\end{aligned}$$

For the estimation step, omit the higher order terms such as $H\Sigma_{t+1|t}H^\top \Delta t^2$ to simplify the equation, and then we have

$$
\begin{aligned}
m_{t+\Delta t|t+\Delta t} &= m_{t+\Delta t|t} - \Sigma_{t+\Delta t|t} H^\top R^{-1}(\Delta \hat{Z}_t - H m_{t+\Delta t|t}\Delta t)\\
\Sigma_{t+\Delta t|t+\Delta t} &= \Sigma_{t+\Delta t|t} - \Sigma_{t+\Delta t|t} H^\top R^{-1}(R - V_t)R^{-1} H \Sigma_{t+\Delta t|t}\Delta t
\end{aligned}
$$

up to $o(\Delta t)$ error. Substitute $m_{t+\Delta t|t}$ and $\Sigma_{t+\Delta t|t}$ to the equation, and ignoring higher order terms,

$$
\begin{aligned}
m_{t+\Delta t} - m_t &= A m_t \Delta t + \Sigma_t H^\top R^{-1}(\Delta \hat{Z}_t - H m_t \Delta t)\\
\Sigma_{t+\Delta t} - \Sigma_t &= \Big(A\Sigma_t + \Sigma_t A^\top + Q - \Sigma_t H^\top \big(R^{-1} - R^{-1} V_t R^{-1}\big) H \Sigma_t\Big)\Delta t
\end{aligned}
$$

The differential formula is obtained by letting $\Delta t \to 0$, and $R = \sigma_w^2$. □

A.4. **Derivation of the Stratonovich expression (10).** By the relation between Itô stochastic integral and Stratonovich integral,

$$
K_t(X_t^i) \circ \mathrm{d}\hat{Z}_t = K_t(X_t^i)\,\mathrm{d}\hat{Z}_t + \frac{1}{2}\,\mathrm{d}\big[K_t(X_t^i), \hat{Z}_t\big]
$$

where $\big[K_t(X_t^i), \hat{Z}_t\big]$ denotes the quadratic covariation term of the two processes. This term is expressed by:

$$
\begin{aligned}
\mathrm{d}\big[K_t(X_t^i), \hat{Z}_t\big] &= \nabla K_t(X_t^i)\,\mathrm{d}[X_t^i, \hat{Z}_t] + \mathrm{d}[K_t(\cdot), \hat{Z}](X_t^i)\\
&= \underbrace{\big(K_t^\top \nabla K_t\big)(X_t^i)\hat{V}_t\,\mathrm{d}t}_{\text{spatial correction}} + \underbrace{\mathrm{d}[K_t(\cdot), \hat{Z}](X_t^i)}_{\text{density correction}}
\end{aligned}
$$

The spatial correction term appears in (8), and it motivates the Stratonovich expression of the FPF with $v_t = 0$. In order to compute the density correction, assume that K_t satisfies a stochastic partial differential equation of the form:

$$
\mathrm{d}K_t(x) = (\cdots)\,\mathrm{d}t - 2v_t(x)\,\mathrm{d}\hat{Z}_t
$$

then the density correction term is exactly $v_t(X_t^i)\hat{V}_t\,\mathrm{d}t$. To derive equation for v_t, differentiate the Poisson equation (9a) with respect to time:

$$
-\nabla \cdot \big(\bar{\pi}_t\,\mathrm{d}K_t + K_t\,\mathrm{d}\bar{\pi}_t + \mathrm{d}[\bar{\pi}_t, K_t]\big) = \frac{1}{\sigma_w^2}\big(\mathrm{d}\bar{\pi}_t(h - \bar{h}_t) - \bar{\pi}_t(\mathrm{d}\bar{h}_t) - \mathrm{d}[\bar{\pi}_t, \bar{h}_t]\big)
$$

Since we are interested in only stochastic term, remove all finite variation terms. Recall from (5) that $\mathrm{d}\bar{\pi}_t = (\cdots)\,\mathrm{d}t + \frac{1}{\sigma_w^2}\bar{\pi}_t(h - \bar{h}_t)\,\mathrm{d}\hat{Z}_t$, we obtain the equation for $\mathrm{d}\hat{Z}_t$ terms:

$$
2\nabla \cdot \big(\bar{\pi}_t v_t\big) - \frac{1}{\sigma_w^2}\nabla \cdot \big(K_t \bar{\pi}_t (h - \bar{h}_t)\big) = \frac{1}{\sigma_w^4}\bar{\pi}_t(h - \bar{h}_t)^2 - \frac{1}{\sigma_w^4}\bar{\pi}_t(2\bar{g}_t)
$$

Upon using (9a),

$$
\nabla \cdot \big(K_t \bar{\pi}_t (h - \bar{h}_t)\big) = (h - \bar{h}_t)\nabla \cdot (\bar{\pi}_t K_t) + \bar{\pi}_t K_t^\top \nabla h = -\frac{1}{\sigma_w^2}\bar{\pi}_t(h - \bar{h}_t)^2 + \bar{\pi}_t K_t^\top \nabla h
$$

This is substituted back to the previous calculation to conclude that v_t satisfies

$$
\nabla \cdot \big(\bar{\pi}_t v_t\big) = \frac{1}{\sigma_w^2}\bar{\pi}_t\big(\frac{1}{2}K_t^\top \nabla h - \frac{1}{\sigma_w^2}\bar{g}_t\big)
$$

A.5. **Proof of Proposition 4.** In order to check if the FPF update law gives the required distribution, we express the Fokker-Planck equation for X_t^i:

$$\begin{aligned} \mathrm{d}\pi_t =& \mathcal{A}^\dagger \pi_t \,\mathrm{d}t - \nabla\cdot(\pi_t K_t)(\,\mathrm{d}\hat{Z}_t - \bar{h}_t \,\mathrm{d}t) + \frac{1}{2}\sum_{n,m=1}^{d} \frac{\partial^2}{\partial x_n \partial x_m}(K_t^n K_t^m \pi_t)\hat{V}_t \,\mathrm{d}t \\ &+ \alpha_t \nabla\cdot(\pi_t K_t(h-\bar{h}_t))\,\mathrm{d}t - \frac{1}{2}\sum_{n,m=1}^{d} \frac{\partial}{\partial x_n}(\pi_t K_t^m \frac{\partial}{\partial x_m} K_t^n)\hat{V}_t \,\mathrm{d}t \\ &- \Big(\frac{V_t+\hat{V}_t}{\sigma_w^2} - 1\Big)\nabla\cdot(\pi_t u_t)\,\mathrm{d}t \end{aligned}$$

where K_t^n denotes the n-th element of the gain $K_t(x) = (K_t^1(x),\dots,K_t^d(x)) \in \mathbb{R}^d$. Upon using the Poisson equations (9) and collecting the terms

$$\begin{aligned} \mathrm{d}\pi_t =& \mathcal{A}^\dagger \pi_t \,\mathrm{d}t + \frac{1}{\sigma_w^2}\pi_t(h-\bar{h}_t)(\,\mathrm{d}\hat{Z}_t - \bar{h}_t \,\mathrm{d}t) + \frac{1}{2}\nabla\cdot(K_t \nabla\cdot(\pi_t K_t))\hat{V}_t \,\mathrm{d}t \\ &+ \alpha_t \nabla\cdot(\pi_t K_t(h-\bar{h}_t))\,\mathrm{d}t + \Big(\frac{V_t+\hat{V}_t}{\sigma_w^2} - 1\Big)\Big(\frac{1}{2}\nabla\cdot(\pi_t K_t(h-\bar{h}_t)) \\ &+ \frac{1}{\sigma_w^2}\pi_t(g-\bar{g}_t)\Big)\,\mathrm{d}t \\ =& \mathcal{A}^\dagger \pi_t \,\mathrm{d}t + \frac{1}{\sigma_w^2}\pi_t(h-\bar{h}_t)(\,\mathrm{d}\hat{Z}_t - \bar{h}_t \,\mathrm{d}t) - \frac{1}{2\sigma_w^2}\nabla\cdot(\pi_t K_t(h-\bar{h}_t))\hat{V}_t \,\mathrm{d}t \\ &+ \alpha_t \nabla\cdot(\pi_t K_t(h-\bar{h}_t))\,\mathrm{d}t + \frac{1}{2}\Big(\frac{V_t+\hat{V}_t}{\sigma_w^2} - 1\Big)\nabla\cdot(\pi_t K_t(h-\bar{h}_t))\,\mathrm{d}t \\ &+ \frac{1}{\sigma_w^2}\Big(\frac{V_t+\hat{V}_t}{\sigma_w^2} - 1\Big)\pi_t(g-\bar{g}_t)\,\mathrm{d}t \\ =& \mathcal{A}^\dagger \pi_t \,\mathrm{d}t + \frac{1}{\sigma_w^2}\pi_t(h-\bar{h}_t)(\,\mathrm{d}\hat{Z}_t - \bar{h}_t \,\mathrm{d}t) + \frac{1}{\sigma_w^2}\Big(\frac{V_t+\hat{V}_t}{\sigma_w^2} - 1\Big)\pi_t(g-\bar{g}_t)\,\mathrm{d}t \end{aligned}$$

concluding the update law (5) for the collective filter. □

REFERENCES

[1] A. Bain and D. Crisan, *Fundamentals of Stochastic Filtering*, Stochastic Modelling and Applied Probability, 60, Springer, New York, 2009.

[2] D. Bakry, F. Barthe, P. Cattiaux and A. Guillin,A simple proof of the Poincaré inequality for a large class of probability measures including the log-concave case, *Electron. Commun. Probab.*, **13** (2008), 60–66.

[3] Y. Bar-Shalom, F. Daum and J. Huang, The probabilistic data association filter: Estimation in the presence of measurement origin uncertainty, *IEEE Control Syst. Mag.*, **29** (2009), 82–100.

[4] Y. Bar-Shalom and T. E. Fortmann, *Tracking and Data Association*, Mathematics in Science and Engineering, 179, Academic Press, Inc., San Diego, CA, 1988.

[5] K. Bergemann and S. Reich, An ensemble Kalman-Bucy filter for continuous data assimilation, *Meteorol. Z.*, **21** (2012), 213–219.

[6] Y. Chen, T. T. Georgiou and M. Pavon, On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint, *J. Optim. Theory Appl.*, **169** (2016), 671–691.

[7] Y. Chen and J. Karlsson, State tracking of linear ensembles via optimal mass transport, *IEEE Control Syst. Lett.*, **2** (2018), 260–265.

[8] T. Kirubarajan and Y. Bar-Shalom, Probabilistic data association techniques for target tracking in clutter, *Proc. IEEE*, **92** (2004), 536–557.

[9] I. Kyriakides, D. Morrell and A. Papandreou-Suppappola, Sequential Monte Carlo methods for tracking multiple targets with deterministic and stochastic constraints, *IEEE Trans. Signal Process.*, **56** (2008), 937–948.

[10] R. S. Laugesen, P. G. Mehta, S. P. Meyn and M. Raginsky, Poisson's equation in nonlinear filtering, *SIAM J. Control Optim.*, **53** (2015), 501–525.

[11] S. Pathiraja, S. Reich and W. Stannat, McKean-Vlasov sdes in nonlinear filtering, preprint, arXiv:2007.12658.

[12] S. Reich, A dynamical systems framework for intermittent data assimilation, *BIT*, **51** (2011), 235–249.

[13] S. Reich and C. Cotter, *Probabilistic Forecasting and Bayesian Data Assimilation*, Cambridge University Press, New York, 2015.

[14] D. Reid, An algorithm for tracking multiple targets, *IEEE Trans. Automat. Control*, **24** (1979), 843–854.

[15] D. Sheldon, T. Sun, A. Kumar and T. Dietterich, Approximate inference in collective graphical models, *Proceedings of the 30th International Conference on Machine Learning*, **28** (2013), 1004–1012. Available from: http://proceedings.mlr.press/v28/sheldon13.pdf.

[16] D. R. Sheldon and T. G. Dietterich, Collective graphical models, in *Advances in Neural Information Processing Systems*, NIPS, 2011, 1161–1169. Available from: https://papers.nips.cc/paper/2011/file/fccb3cdc9acc14a6e70a12f74560c026-Paper.pdf.

[17] R. Singh, I. Haasler, Q. Zhang, J. Karlsson and Y. Chen, Inference with aggregate data: An optimal transport approach, preprint, arXiv:2003.13933.

[18] S. C. Surace, A. Kutschireiter and J.-P. Pfister, How to avoid the curse of dimensionality: Scalability of particle filters with and without importance weights, *SIAM Rev.*, **61** (2019), 79–91.

[19] A. Taghvaei and P. G. Mehta, An optimal transport formulation of the ensemble Kalman filter, *IEEE Trans. Automat. Control*, **66** (2021), 3052–3067.

[20] S. Thrun, Probabilistic robotics, *Communications of the ACM*, **45** (2002), 52–57.

[21] T. Yang, R. S. Laugesen, P. G. Mehta and S. P. Meyn, Multivariable feedback particle filter, *Automatica J. IFAC*, **71** (2016), 10–23.

[22] T. Yang and P. G. Mehta, Probabilistic data association - Feedback particle filter for multiple target tracking applications, *J. Dyn. Sys. Meas. Control*, **140** (2018), 14pp.

[23] T. Yang, P. G. Mehta and S. P. Meyn, Feedback particle filter for a continuous-time Markov chain, *IEEE Trans. Automat. Control*, **61** (2016), 556–561.

Received February 2021; 1st revision April 2021; 2nd revision June 2021; early access August 2021.

E-mail address: jkim684@illinois.edu
E-mail address: amirhoseintghv@gmail.com
E-mail address: yongchen@gatech.edu
E-mail address: mehtapg@illinois.edu