Convergence of the Markov Chain Distributed Particle Filter (MCDPF)

Sun Hwan Lee and Matthew West

***Abstract*—The Markov Chain Distributed Particle Filter (MCDPF) is an algorithm for the nodes in a sensor network to cooperatively run a particle filter, based on each sensor making updates to a local particle set using only local measurements, and then having particles exchanged between neighboring sensors based on a Markov chain on the network. This paper extends previously-known almost sure convergence results for the MCDPF to prove that the MCDPF convergences to the optimal filter in mean square as the number of particles and the number of Markov chain steps both go to infinity. The convergence proof derives an explicit error bound, showing that the convergence is inverse square-root in both parameters. A numerical example is provided to support the theoretical result.**

***Index Terms*—Bayesian estimation, distributed estimation, Markov chain, optimal filtering, particle filtering.**

I. Introduction

THERE is a long history of methods for numerically approximating the distributions that arise in the Bayesian filtering problem. The popular methods developed in 1960s and 1970s include the extended Kalman filter [1] and sequential Monte Carlo method [2]–[5]. The very first introduction of the sequential Monte Carlo method, also known as particle filtering, goes back to the calculation of a polymer growing [6], [7]. Particle filtering was not able to be broadly adopted initially, mainly because of its very high computational complexity and the lack of adequate computing resources at that time [8]. Along with a huge development in computing power, particle filtering has become a very active research topic and has been applied to various areas. Among those areas, signal processing started to take advantage of particle filtering following a seminal paper [9]. Various modifications of a standard particle filter to improve the performance are introduced in the tutorial paper [10] in a clear manner. Stratified sampling, residual sampling [11] and systematic resampling [12] were proposed as efficient resampling schemes. Pitt and Shephard [13] introduced the Auxiliary Sampling Importance Resampling (ASIR) filter for better estimation and the regularized particle filter (RPF) was proposed in [14] to solve the problem induced by a resampling step.

Given a collection of distributed sensors, together with a communication network linking them, it is natural to consider how particle filters can be applied in a distributed setting, for instance in target tracking with a robotic navigation system [15], [16]. The general benefits of distributed estimation include the robustness of the estimation and the reduction in information flow while maintaining estimation accuracy comparable to the centralized approach. Much effort has been directed toward the realization of decentralized Kalman filters [17]–[19] but decentralized particle filters were thought to be challenging due to the difficulty of merging probability distributions represented by particles and weights [20]. Existing methods (e.g., [15], [16]) thus use approximations such as Gaussian mixture models, but then give up some of the robustness and simplicity of standard particle filters. Communicating state estimates directly (that is, exchanging particles between sensor nodes) regains the simplicity of particle filters.

Past work on distributed particle filters can be broadly categorized into two approaches, namely message passing approaches and consensus-based local information exchange methods. Message passing approaches transfer information along a predetermined route covering an entire network. For example, [21] passes parameters of the parametric model of the posterior distribution while [22] transmits the raw information, particles and weights, or the parameters of a Gaussian mixture model approximation of the posterior distribution. Consensus-based methods communicate the information only between the nearest nodes and achieve global consistency by consensus filtering. The type of exchanged information can be, for example, the parameters of a Gaussian mixture model approximation [16] or the local mean and covariance [23].

In this paper we consider the Markov Chain Distributed Particle Filter (MCDPF) introduced in [24]. This is a direct particle-exchange method, in which nearest-neighbor nodes in a sensor network exchange particles and weights using a Markov chain random walk and thereby collectively estimate the system state. The MCDPF can have better scaling in the number of system states than methods based on full covariance information, particularly for highly nonlinear and non-Gaussian problems. Such systems arise, for example, in fault detection [25] and in estimating the indoor environment of building systems [26].

The almost sure convergence of the MCDPF to the standard Centralized Particle Filter (CPF) and to the optimal filter was proved in [24]. A numerical comparison between the MCDPF

Manuscript received March 02, 2011; revised August 08, 2011, January 04, 2012, May 30, 2012, and September 20, 2012; accepted October 27, 2012. Date of publication November 30, 2012; date of current version January 23, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Yufei Huang. The work of S. H. Lee was supported by a Samsung scholarship. The work of M. West was supported by the NSF by Grants ATM-0739404 and ATM-0934491, and by the DOE by Grant DE-SC0003921. The work in this paper was presented in part at the IEEE Conference on Decision and Control, Shanghai, China, December 2009.

S. H. Lee was with the Department of Aeronautics and Astronautics, Stanford University, Stanford, CA 94305 USA. He is now with Lattice Engines Inc., San Mateo, CA 94402 USA (e-mail: sunhwan.lee@gmail.com).

M. West is with the Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: mwest@illinois.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2012.2231075

1053-587X/$31.00 © 2012 IEEE

and other distributed filters was undertaken in [27]. In the current work we prove the convergence of the MCDPF to the optimal filter in the sense of mean square. In proving mean square convergence we also obtain explicit bounds that give the rate of convergence of the MCDPF. The convergence of non-distributed particle filters has been extensively considered (e.g., [28]–[33]), and [34] provides an excellent survey of existing results.

The contents of this paper are as follows. In Section II we review basic properties and theorems for random walks on graphs. Section III defines the Centralized Particle Filter (CPF) and the Markov Chain Distributed Particle Filter (MCDPF). In addition, the almost sure convergence of the posterior distribution of the MCDPF to that of the CPF and the optimal filter, which is the main result of [24], is reviewed in this section. Section IV contains a precise definition of mean square convergence and the proof of mean square convergence of the MCDPF to the optimal filter. Section V-B provides the quantitative comparison of performance between MCDPF and the distributed Extended Kalman Filter (EKF) with respect to system complexity and measurement frequency. Conclusions are made in Section VI.

II. Random Walks on Graphs

A sensor network system can be modeled as a graph, $G = (V, E)$, with adjacency matrix $\mathcal{A}$ normalized so that row-sums are one (i.e., $\mathcal{A}\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ denotes a vector with all entries equal to one). The vertices $V = \{1, \ldots, m\}$ correspond to nodes or sensors in the network system and edges E represent the communication links between sensors. The set of neighbors of node i is defined as $N_i = \{i \in V : a_{ij} \neq 0\}$. Matrix $\mathcal{A}$ is a Markov transition probability matrix defined on the graph because it satisfies

$$\begin{aligned} &\mathcal{A}_{ij} \geq 0 \text{ for all } i, j \\ &\mathcal{A}\mathbf{1} = \mathbf{1}. \end{aligned}$$

Consequently, a random walk on the graph can be defined by regarding matrix $\mathcal{A}$ as a Markov transition matrix. We assume that there are no self-loops in the chain. Suppose X_t a random variable giving the state (graph vertex) after t random walk steps on the graph with a Markov transition matrix $\mathcal{A}$. Then X_t satisfies the Markov property

$$\Pr(X_{t+1} = i \mid X_t = j) = \mathcal{A}_{ij}.$$

Furthermore, a probability distribution of the random variable X_t, which we denote as Π_t here, evolves by

$$\Pi_{t+1} = \mathcal{A}^T \Pi_t.$$

We next review several properties of random walks on graphs which are useful for a development of distributed particle filters.

Theorem 1: If $\mathcal{A}$ is the normalized adjacency matrix of an undirected, connected graph G then the Markov chain with transition matrix $\mathcal{A}$ has a unique stationary distribution Π with all non-zero entries. For any initial distribution,

$$\lim_{k \to \infty} \frac{M(\cdot, k)}{k} = \Pi,$$

where $M(\cdot, k) \in \mathbb{R}^m$ is a vector whose i-th element is the number of visits to state i during k steps of the Markov chain. Furthermore, $M(\cdot, k) \in \mathbb{R}^m$ converges to Π in distribution as $k \to \infty$. More precisely,

$$\sqrt{k}\left(\frac{M(\cdot, k)}{k} - \Pi\right) \xrightarrow[d]{\mathcal{N}} (0, V), \tag{1}$$

where $\mathcal{N}(0, V)$ is the normal distribution with mean zero and a covariance matrix V that is independent of k.

Proof: See [35, Theorem 42.VII]. ■

Theorem 2: If $\mathcal{A}$ is the normalized adjacency matrix of an undirected, connected graph G then the stationary distribution of the Markov chain with transition matrix $\mathcal{A}$ is given by $\Pi = (\Pi_1, \Pi_2, \ldots, \Pi_m)$ with $\Pi_i = \frac{d(i)}{2|E(G)|}$, where $d(i)$ is the degree of node i and $|E(G)|$ is the number of edges in the graph.

Proof: See [24, Theorem 2.2]. ■

III. Particle Filters

Suppose we have the general state space model

$$\begin{aligned} x_{t+1} &= f(x_t, \zeta_t) \\ y_t &= g(x_t, \xi_t), \end{aligned} \tag{2}$$

where $x_t \in \mathbb{R}^n$ is the state, $y_t \in \mathbb{R}^p$ is the measurement, ζ_t is the process noise, and ξ_t is the measurement noise. We define two stochastic processes, $X = \{X_t \mid t \in \mathbb{N}\}$ and $Y = \{Y_t \mid t \in \mathbb{N}\}$, where X and Y are a signal process and an observation process, respectively. The signal process X is a Markov process with an initial distribution $\mu(x_0)$ and transition kernel $K(dx_t \mid x_{t-1})$ and the observation process Y is conditionally independent given X_t, which means that $\Pr(Y_t \in B \mid X, Y \setminus Y_t) = \Pr(Y_t \in B \mid X_t)$. For simplicity, we assume that the kernel and conditional probability distribution of Y can be written as densities with respect to the Lebesgue measure. Thus

$$\begin{aligned} \Pr(X_t \in A \mid X_{t-1} = x_{t-1}) &= \int_A K(x_t \mid x_{t-1}) dx_t \\ \Pr(Y_t \in B \mid X_t = x_t) &= \int_B \rho(y_t \mid x_t) dy_t, \end{aligned}$$

where $\rho(y_t \mid x_t)$ is the transition probability density of a measurement y_t given the state x_t. To facilitate the convergence proof in Section IV, we will further assume that $\rho(y_t \mid x_t)$ is bounded above with the definition of $\|\rho\|_\infty = \sup_{y_t \in \mathbb{R}^p} |\rho(y_t \mid x_t)|$, so that $\rho(y_t \mid x_t) \leq \|\rho\|_\infty < \infty$, and that ρ is bounded away from zero, so that $\rho \geq \rho_0 > 0$ pointwise for a fixed ρ_0.

The filtering problem is to estimate the true state x_t at time t given the time series of observations $y_{1:t} = \{y_1, \ldots, y_t\}$. The prediction and updating of the optimal filtering based on Bayes' recursion are given as follows.

$$\begin{aligned} p(x_t \mid y_{1:t-1}) &= \int_{\mathbb{R}^n} p(x_{t-1} \mid y_{1:t-1}) K(x_t \mid x_{t-1})\, dx_{t-1} \\ p(x_t \mid y_{1:t}) &= \frac{\rho(y_t \mid x_t) p(x_t \mid y_{1:t-1})}{\int_{\mathbb{R}^n} \rho(y_t \mid x_t) p(x_t \mid y_{1:t-1})\, dx_t}. \end{aligned} \tag{3}$$

Analytic solutions for the posterior distribution in (3) do not generally exist except in special cases, such as linear dynamical systems with Gaussian noise. In the particle filtering setting, the posterior distribution is represented by a group of particles and associated weights so that the integral in (3) is approximated by the sum of discrete values.

A. The Centralized Particle Filter (CPF)

The Centralized Particle Filter (CPF) is a recursive method to estimate the true state, given the time series of measurements [28], [34]. The algorithm consists of the initialization and iteration steps. At $t = 0$ we have the empirical distribution with N particles

$$\pi_{0|0}(dx_0) \approx \pi^N_{0|0}(dx_0). \tag{4}$$

Now the iteration step is repeated as follows. Suppose the posterior distribution at time $t-1$, $\pi_{t-1|t-1}(dx_{t-1})$, is approximated by N particles $\{x^i_{t-1}\}^N_{i=1}$. Then we have

$$\begin{aligned} p(x_{t-1} \mid y_{1:t-1}) &\triangleq \pi_{t-1|t-1}(dx_{t-1}) \approx \pi^N_{t-1|t-1}(dx_{t-1}) \\ &\triangleq \frac{1}{N}\sum_{i=1}^{N} \delta_{x^i_{t-1}}(dx_{t-1}). \end{aligned} \tag{5}$$

where particle i is at position x^i_{t-1} in state space. The particles then go through the prediction and measurement update steps to approximate the posterior distribution at time t. The N particles are propagated through the transition kernel density,

$$\tilde{x}^i_t \sim \pi^N_{t-1|t-1}K(dx_t), \tag{6}$$

where the integration of transition kernel with respect to measures is defined [34, Section III.A] as

$$\pi^N_{t-1|t-1}K(dx_t) \triangleq \frac{1}{N}\sum_{i=1}^{N} K(x_t \mid x^i_{t-1}). \tag{7}$$

Now this new set of particles is the approximation of $\pi_{t|t-1}$. That is,

$$\begin{aligned} p(x_t \mid y_{1:t-1}) &\triangleq \pi_{t|t-1}(dx_t) \\ &\approx \tilde{\pi}^N_{t|t-1}(dx_t) \triangleq \frac{1}{N}\sum_{i=1}^{N} \delta_{\tilde{x}^i_t}(d\tilde{x}_t). \end{aligned} \tag{8}$$

If the empirical distribution in (8) is substituted in (3), we have the following distribution approximating the posterior distribution $p(x_t \mid y_{1:t})$.

$$\begin{aligned} \tilde{\pi}^N_{t|t}(dx_t) &\triangleq \frac{\rho(y_t \mid x_t)\tilde{\pi}^N_{t|t-1}(dx_t)}{\int_{\mathbb{R}^n} \rho(y_t \mid x_t)\tilde{\pi}^N_{t|t-1}(dx_t)\,dx_t} \\ &= \frac{\sum_{i=1}^{N} \rho(y_t \mid \tilde{x}^i_t)\delta_{\tilde{x}^i_t}(d\tilde{x}_t)}{\sum_{i=1}^{N} \rho(y_t \mid \tilde{x}^i_t)} = \sum_{i=1}^{N} w^i_t \delta_{\tilde{x}^i_t}(d\tilde{x}_t) \end{aligned} \tag{9}$$

where $\sum_{i=1}^N w^i_t = 1$ and w^i_t are called the importance weights. To avoid the degeneracy problem, new particles are selected according to a resampling step that samples N particles from the empirical distribution, $\tilde{\pi}^N_{t|t}(dx_t)$, each with weight $\frac{1}{N}$. We then have the empirical distribution approximating the posterior at time t given by

$$\pi^N_{t|t}(dx_t) \triangleq \frac{1}{N}\sum_{i=1}^{N} \delta_{x^i_t}(dx_t). \tag{10}$$

While the above particle filter is perhaps the simplest and most direct particle discretization of the Bayes recursion (3), there are many possible extensions which can result in superior performance. For example, the prediction step (6) can be replaced with the classic bootstrap filter [9] sampler $\tilde{x}^i_t \sim K(dx_t \mid x^i_{t-1})$, or a new kernel $\tilde{K}$ can be used with $\tilde{x}^i_t \sim \pi^N_{t-1|t-1}\tilde{K}(dx_t)$, where $\tilde{K}$ is chosen to better approximate the true distribution (see [34], [36], [37]). The Markov chain distribution method described in Section III-B can be applied to any such variant, and the convergence proofs in Sections III-C, III-D, and IV extend with some adjustments. The precise version of the algorithm presented in this paper was chosen to simplify the analysis and presentation.

B. The Markov Chain Distributed Particle Filter (MCDPF)

The main difference between centralized and distributed particle filters is that centralized filters have a central unit to collect the measurements from all nodes and update particles using all measurements simultaneously. During the process of data collection, the centralized filter might suffer from bottlenecks in information flow. On the other hand, a distributed filter can overcome this problem by passing information only locally between connected nodes. If we have m nodes measuring the partial observations independently, then we can decompose the general state space model (2) as follows.

$$\begin{aligned} x_{t+1} &= f(x_t, \zeta_t) \\ \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{m,t} \end{bmatrix} &= \begin{bmatrix} g_1(x_t, \xi_{1,t}) \\ g_2(x_t, \xi_{2,t}) \\ \vdots \\ g_m(x_t, \xi_{m,t}) \end{bmatrix}. \end{aligned} \tag{11}$$

Here $x_t \in \mathbb{R}^n$, $y_{v,t} \in \mathbb{R}^{p_v}$ with $\sum_{v=1}^m p_v = p$ and subscript v represents node v. In addition, the measurement noise at each nodes is assumed to be uncorrelated, $\mathbf{E}[\xi_t \xi_t^T] = \operatorname{diag}(R_1, R_2, \ldots, R_m)$. Uncorrelated noise structure enables us to have conditionally independent measurements at each node, $y_{i,t}$, given the true state x_t. As a consequence of this assumption, the function $\rho(y_t \mid x_t)$ in (3) can be factorized by a product of $\rho_v(y_{v,t} \mid x_t)$ at each node,

$$\rho(y_t \mid x_t) = \prod_{v=1}^{m} \rho_v(y_{v,t} \mid x_t). \tag{12}$$

We propose a distributed particle filtering method using a random walk on the graph defined by the network topology. In the sensor network, node v measures the partial observation $y_{v,t}$ at time t and data at every node has to be fused to reach the global estimation of the true state. While achieving a global estimate by exchanging data, it is desirable to maintain a robustness with respect to the unexpected changes of global properties such as losing a node. The distributed particle filter proposed here is robust since the information, consisting of particles and weights, is transferred only to the connected neighborhood of each node. In other words, every node only needs local information. Transferring particle data is inefficient for low-dimensional systems, but scales well (only linearly) with dimension of the state space, as opposed to existing methods using Gaussian mixture model approximations of posterior distribution [16], [23] [22]. Further performance comparison with the numerical example is provided in [27] and reviewed in Section V-B.

The MCDPF moves particles around the network according to the Markov chain on the network defined by the normalized

adjacency matrix $\mathcal{A}$ to compute the importance weights. Precisely speaking, once the measurement is received in each node or sensor, each node updates the weights of its current particles by using its local measurement to perform a partial update (defined precisely below). Then particles (and associated weights) are then moved to the neighboring nodes, with each particle passed to a neighboring node chosen at random uniformly. These update/move steps are then repeated k times before another measurement is taken.

The above description is from the point of view of a fixed node. This process can also be viewed from the point of view of a moving particle. The particle takes k steps of a Markov chain random walk on the sensor network, with transition probabilities given by the normalized adjacency matrix $\mathcal{A}$. For each node that the particle passes through, it updates its associated weight using the node's current measurement (defined precisely below).

To define the partial measurement update, suppose the sensor network has graph $G = (V, E)$ with normalized adjacency matrix $\mathcal{A}$. Then $|E(G)|$ is the number of edges and $d(v)$ is the degree of node v. The partial measurement update applied at node v to a particle with state $\tilde{x}_t$ changes its weight $\tilde{w}_t$ by

$$\tilde{w}_t \leftarrow \tilde{w}_t \times \rho_v(y_{v,t} \mid \tilde{x}_t)^{\frac{2|E(G)|}{kd(v)}}.$$

This is repeated k times as the particle moves between nodes (the particle view) or as the nodes exchange particle sets (the nodal view).

Observe that as the number of steps k is increased, each partial measurement update has less effect. However, performing k such updates still gives the correct total update. This paper proves that as $k \to \infty$, the total effect of these partial updates converges to the update of the CPF.

To express the total effect of the k steps of this movement/update, recall that the number of visits by the particle to the v-th node is denoted $M(v, k)$. Then we can multiply the partial updates for each of the N particles to obtain the posterior distribution of the MCDPF as follows.

$$\begin{aligned}\tilde{\pi}^N_{t|t,k}(dx_t) &\triangleq \frac{\sum_{i=1}^{N} \prod_{v=1}^{m} \rho_v(y_{v,t} \mid \tilde{x}^i_t)^{\frac{2|E(G)|}{kd(v)} \times M(v,k)} \delta_{\tilde{x}^i_t}(d\tilde{x}_t)}{\sum_{i=1}^{N} \prod_{v=1}^{m} \rho_v(y_{v,t} \mid \tilde{x}^i_t)^{\frac{2|E(G)|}{kd(v)} \times M(v,k)}} \\ &= \sum_{i=1}^{N} w^i_{t,k} \delta_{\tilde{x}^i_t}(d\tilde{x}_t). \end{aligned} \tag{13}$$

The MCDPF is defined in Algorithm 1. We use the notation $x^i_{v,t}$ for the i-th particle of node v at time t and $N(v)$ for the number of particle at node v. Thus

$$\pi^{N(v)}_{v,t-1|t-1}(dx_{t-1}) \triangleq \frac{1}{N(v)} \sum_{i=1}^{N(v)} \delta_{x^i_{v,t-1}}(dx_{t-1}) \tag{14}$$

$$\pi^{N(v)}_{v,t-1|t-1} K(dx_t) \triangleq \frac{1}{N(v)} \sum_{i=1}^{N(v)} K(x_t \mid x^i_{v,t-1}) \tag{15}$$

are the approximate posterior distribution and integrated transition kernel at node v. Also $\mathcal{I}_{\ell \to v}$ is the indices of particles moving from node ℓ to v in the current Markov chain step and we recall that $\mathcal{A}$ is the adjacency matrix of the network.

Algorithm 1 Markov Chain Distributed Particle Filter (MCDPF)

Initialization for time $t = 0$, node v, particle i:

$x^i_{v,0} \sim p(x_0)$

$w^i_{v,0} = \frac{1}{N(v)}$

Prediction step for time t, node v, particle i:

$\tilde{x}^i_{v,t} \sim \pi^{N(v)}_{v,t-1|t-1} K(dx_t)$

$\tilde{w}^i_{v,t} = 1$

Measurement step for time t:

for k iterations **do**

Sample particle movement sets $\mathcal{I}_{\ell \to v}$ using $\mathcal{A}$

for node $v = 1$ **to** m **do**

$\{\tilde{x}^i_{v,t}\}_{i=1}^{N(v)} \leftarrow \bigcup_{l=1}^{m} \{\tilde{x}^i_{l,t}\}_{i \in \mathcal{I}_{l \to v}}$

$\{\tilde{w}^i_{v,t}\}_{i=1}^{N(v)} \leftarrow \bigcup_{l=1}^{m} \{\tilde{w}^i_{l,t}\}_{i \in \mathcal{I}_{l \to v}}$

for particle $i = 1$ **to** $N(v)$ **do**

$\tilde{w}^i_{v,t} \leftarrow \tilde{w}^i_{v,t} \times \rho_v(y_{v,t} \mid \tilde{x}^i_{v,t})^{\frac{2|E(G)|}{kd(v)}}$

end for

end for

end for

Resample step for time t, node v, particle i:

$x^i_{v,t} \sim \tilde{\pi}^N_{v,t|t}(d\tilde{x}_t) = \sum_{l=1}^{N(v)} \tilde{w}^l_{v,t} \delta_{\tilde{x}^l_{v,t}}(d\tilde{x}_t)$

$w^i_{v,t} = \frac{1}{N(v)}$

C. Convergence to CPF and Algorithm

We showed [24] that the empirical posterior distribution of the MCDPF converges almost surely to that of the CPF as the number of Markov chain steps k per measurement goes to infinity. We briefly summarize this result here in preparation for the proof of mean square convergence in Section IV.

We largely follow the notation of Crisan and Doucet [34]. In the stochastic filtering problem, we consider the functions $a_t : \pi_{t|t-1} \mapsto \pi_{t|t}$ and $b_t : \pi_{t-1|t-1} \mapsto \pi_{t|t-1}$, which are continuous functions mapping a metric space (E, d) to itself. Additionally, $a^k_t : \pi_{t|t-1} \mapsto \pi_{t|t,k}$ is also a continuous function, given by

$$\begin{aligned} &a^k_t(p(x_t \mid y_{1:t-1})) \\ &\quad = \frac{\prod_{v=1}^{m} \rho_v(y_{v,t} \mid x_t)^{\frac{2|E(G)|}{kd(v)} \times M(v,k)} p(x_t \mid y_{1:t-1})}{\int_{\mathbb{R}^n} \prod_{v=1}^{m} \rho_v(y_{v,t} \mid x_t)^{\frac{2|E(G)|}{kd(v)} M(v,k)} p(x_t \mid y_{1:t-1})\, dx_t}. \end{aligned}$$

The perturbation c^N is defined as a function that maps from a measure μ to a random sample size of size N of the measure, so that

$$c^{N,w}(\mu) \triangleq \frac{1}{N} \sum_{j=1}^{N} \delta_{\{V_j(w)\}},$$

where $V_j : \Omega \to \mathbb{R}^n$ is an i.i.d. random variable with distribution μ, and $w \in \Omega$. For notational simplicity, let $h^N_t, h^N_{1:t}$ be defined as the composition of functions a_t, b_t, c^N, giving

$$h^N_t \triangleq c^N \circ a_t \circ c^N \circ b_t, \quad h^N_{1:t} \triangleq h^N_t \circ \cdots \circ h^N_1$$
$$h^N_{t,k} \triangleq c^N \circ a^k_t \circ c^N \circ b_t, \quad h^N_{1:t,k} \triangleq h^N_{t,k} \circ \cdots \circ h^N_{1,k}.$$

Thus the posterior distribution of the CPF and MCDPF at time t can then be expressed as

$$\pi^N_{t|t} = h^N_t(\pi^N_{t-1|t-1}) = h^N_{1:t}(\pi_0)$$
$$\pi^N_{t|t,k} = h^N_{t,k}(\pi^N_{t-1|t-1,k}) = h^N_{1:t,k}(\pi_0).$$

Almost sure convergence of the MCDPF to the CPF as the number of Markov chain steps k goes to infinity can be stated as follows.

Theorem 3: Consider a connected sensor network with measurements at different nodes conditionally independent given the true state. Then the estimated distribution of the MCDPF converges almost surely to the estimated distribution of the CPF as the number of Markov chain steps k per measurement goes to infinity. That is,

$$\lim_{k\to\infty} \pi^N_{t|t,k} = \pi^N_{t|t}$$

pointwise.

Proof: See [24, Section C]. ■

D. Convergence to Optimal Filtering

Having established that the MCDPF converges almost surely to the CPF, it is relatively straightforward to show that the MCDPF converges to the optimal filter as both the number of particles N and the number of Markov chain steps k go to infinity. The convergence of the CPF to the optimal filter is shown in [34], [38]. The only difference between the MCDPF and the standard particle filter used in that proof is the function a^k_t, which needs to satisfy the following condition to ensure the convergence of the MCDPF to the optimal filter. For all sequences $e_N \to e \in E$ we have

$$\lim_{N\to\infty} \lim_{k\to\infty} a^k_t(e_N) = a_t(e). \tag{16}$$

This property is not obvious because the function a^k_t converges only pointwise to a function a_t. Fortunately, however, a_t is continuous, which is sufficient to give (16). For reference, we restate the main convergence result here.

Theorem 4: Assume that the kernel K is Feller and the function ρ is bounded, continuous and strictly positive. Then the estimated distribution of the MCDPF algorithm converges to the optimal filtering distribution as the number of particles N and the number of Markov chain steps k per measurement go to infinity:

$$\lim_{N\to\infty} \lim_{k\to\infty} \pi^N_{t|t,k} = \pi_{t|t}. \tag{17}$$

Proof: See [24, Theorem 3.7]. ■

IV. Mean Square Convergence

In the previous section, almost sure convergence of the MCDPF to the optimal filter was discussed. We now prove that this convergence is in the sense of mean square.

We say that the sequence of random probability measures $(\mu_N)_{N=1}^{\infty}$ converges in mean square to μ if for any $\varphi \in B(\mathbb{R}^n)$, where $B(\mathbb{R}^n)$ is the set of bounded Borel measurable functions,

$$\lim_{N\to\infty} \mathbf{E}\left[((\mu_N,\varphi) - (\mu,\varphi))^2\right] = 0.$$

Here we define

$$(\mu,\varphi) \triangleq \int \varphi\mu, \quad K\varphi(x) \triangleq \int K(dz \mid x)\varphi(z). \tag{18}$$

A. Preliminaries

We will need several lemmas to prove the main mean square convergence result. We first recall the main results on the Markov chain central limit theorem. If $(X_i)_{i=0}^{\infty}$ is an irreducible, positive recurrent Markov chain on a countable state space $\mathcal{X}$ with stationary distribution Π, then for $g : \mathcal{X} \to \mathbb{R}$ the random variable $W_k(g)$ defined by

$$W_k(g) = \sqrt{k}\sum_{i=0}^{k-1} (g(X_i) - \Pi(g)) \tag{19}$$

can be shown [39] to converge to the normal random variable with mean zero and variance

$$\gamma^2_g = \Pi(x_0)\mathbf{E}_{x_0}\left[\left(\sum_{i=0}^{\tau_{x_0}-1} [g(X_i) - \Pi(g)]\right)^2\right], \tag{20}$$

where τ_{x_0} is the first return time to the initial state and $\Pi(g)$ is an expected value of function g with respect to the stationary distribution Π, $\Pi(g) = \mathbf{E}[g(X)]$ for $X \in \mathcal{X}$ distributed by $X \sim \Pi$. If we let P^i_x be the distribution of X_i with $X_0 = x$, then it can also be shown [39] that $\left|P^i_x(g) - \Pi(g)\right| \le R\nu^i$, where R is a constant and ν is the second largest eigenvalue modulus (SLEM) of the Markov transition matrix $\mathcal{A}$ that defines the Markov chain.

We also have the following theorem [40, Corollary 1] on the convergence of the moment generating function.

Theorem 5: If there is a positive c such that $|g(x)-\Pi(g)| \le c$ for all $x \in \mathcal{X}$, then for any $\lambda \le \frac{1}{(\sqrt[3]{3}L\vee L')}$, all $k \ge 1$, and all $x \in \mathcal{X}$,

$$\begin{aligned}\left|\mathbf{E}_x \exp\{\lambda W_k(g)\} - \mathbf{E}\exp\{\lambda\gamma_g U\}\right| &\le k^{-\frac{1}{2}}V(x)\Big[C'L'\lambda e^{(\lambda L')^2} + k^{-\frac{1}{2}}Ce^{(\sqrt[3]{3}\lambda L)^2} \\ &\quad + k^{-1}\frac{C'L'\lambda}{1-(\lambda L')^2} + k^{-\frac{3}{2}}\frac{C}{1-(\sqrt[3]{3}\lambda L)^2}\Big],\end{aligned} \tag{21}$$

where U is a standard normal random variable and $V(x)$ is a function to ensure the V-uniform ergodicity of a Markov chain. Furthermore, the following positive constants C, L, C' and L' depend on the SLEM, ν.

$$C = \frac{(4e)^2 + 3 + 2(1-\nu)}{1-\nu}, \tag{22}$$

$$L = e^{\frac{\log 3}{2\cdot 3}} R^{\frac{25}{32}}(1-\nu)^{-\frac{1}{2}} 2\left(\sqrt{e} \vee \sqrt{\frac{2}{1-\nu}}\right), \tag{23}$$

$$C' = \sqrt{\frac{2R}{1-\nu}}, \tag{24}$$

$$L' = \sqrt{\frac{2R}{1-\nu}} e^{\frac{\log 3}{2(2\cdot 3+1)}}. \tag{25}$$

where $a \vee b$ denotes the minimum of a and b and $\sqrt[3]{3}L \vee L' = (\sqrt[3]{3}L) \vee L'$.

Proof: The proof of this theorem is given by Steinsaltz [40, Sections 4 and 5.1]. Here we will show how the constants L, L', C and C' are related to the SLEM. The upper bound of the error of the $2n$-th moment given in [40, equation (42)] reduces to

$$k^{-1}\frac{n(2n)!}{(n-1)!}\frac{(4e)^2+3+2(1-\nu)}{1-\nu} \times \left(e^{\frac{\log 3}{2\cdot 3}}R^{\frac{25}{32}}(1-\nu)^{-\frac{1}{2}}2\left(\sqrt{e}\vee\sqrt{\frac{2}{1-\nu}}\right)\right)^{2n} \times \left(1+\frac{n!}{k}\right)V(x).$$

With C and L as defined above, the error of the even moments has the bound

$$\left|\mathbf{E}_x W_k(g)^{2n} - (2n-1)(\gamma_g^2)^n\right| \leq k^{-1}CL^{2n}\frac{n(2n)!}{(n-1)!}\left(1+\frac{n!}{k}\right)V(x).$$

Similarly for the error of odd moments, the upper bound given in [40, equation (43)] can be written as

$$\left|\mathbf{E}_x W_k(g)^{2n+1} - 2n(\gamma_g^2)^{n+1}\right| \leq k^{-\frac{1}{2}}\frac{(2n+1)!}{n!}\sqrt{\frac{2R}{1-\nu}}\left(\sqrt{\frac{2R}{1-\nu}}e^{\frac{\log 3}{2(2\cdot 3+1)}}\right)^{2n+1} \times \left(1+\frac{n!}{k}\right)V(x).$$

Using C' and L' from above, the error of the odd moments has the bound

$$\left|\mathbf{E}_x W_k(g)^{2n+1}\right| \leq k^{-\frac{1}{2}}C'L'^{2n+1}\frac{(2n+1)!}{n!}\left(1+\frac{n!}{k}\right)V(x).$$

Hence the error bound (21) for the moment generating function is given by

$$\begin{aligned}&\left|\mathbf{E}_x \exp\{\lambda W_k(g)\} - \mathbf{E}\exp\{\lambda\gamma_g U\}\right| \\ &\leq \sum_{n=0}^{\infty} V(x)\left[\frac{\lambda^{2n}n^2}{n!}k^{-1}CL^{2n}\left(1+\frac{n!}{k}\right) + \frac{\lambda^{2n+1}}{n!}k^{-\frac{1}{2}}C'L'^{2n+1}\left(1+\frac{n!}{k}\right)\right] \\ &\leq k^{-\frac{1}{2}}V(x)\left[C'L'\lambda e^{(\lambda L')^2} + k^{-\frac{1}{2}}Ce^{(\sqrt[3]{3}\lambda L)^2} + k^{-1}\frac{C'L'\lambda}{1-(\lambda L')^2} + k^{-\frac{3}{2}}\frac{C}{1-(\sqrt[3]{3}\lambda L)^2}\right].\end{aligned}$$

■

The following lemma is a trivial inequality from [38, Lemma 7.2].

Lemma 6: Let Y be a random variable. If the p-th moment of Y is finite, $\mathbf{E}|Y|^p < \infty$, then for any $p \geq 1$,

$$\mathbf{E}|Y - \mathbf{E}(Y)|^p \leq 2^p\mathbf{E}|Y|^p.$$

Proof: From Jensen's inequality with $p \geq 1$, we have

$$(\mathbf{E}|Y|)^p \leq \mathbf{E}|Y|^p. \tag{26}$$

Minkowski's inequality gives

$$\begin{aligned}(\mathbf{E}|Y-\mathbf{E}(Y)|^p)^{\frac{1}{p}} &\leq (\mathbf{E}|Y|^p)^{\frac{1}{p}} + (|\mathbf{E}Y|^p)^{\frac{1}{p}} \\ &= (\mathbf{E}|Y|^p)^{\frac{1}{p}} + |\mathbf{E}Y| \\ &\leq 2\,(\mathbf{E}|Y|^p)^{\frac{1}{p}},\end{aligned}$$

where the last inequality is from (26). ■

Proof of Mean Square Convergence

We start this section by considering a single particle and obtaining a bound on the difference of likelihood for this particle in the centralized and decentralized cases, for a given set of measurements. The weighting of a particle in the CPF is computed from the entire measurement, whereas the weighting of a particle in the MCDPF is computed sequentially as the particle jumps around the nodes, updating its likelihood at each node. Hence, understanding how the weights determined by likelihood are different in the two cases is a fundamental step for the main proof.

Let x_t be a particle state and $M(v,k)$ be the number of visits by this particle to the v-th node in k steps. Define the random variable $\rho_{t,k}$ and the function ρ_t by

$$\rho_{t,k}(x_t) = \prod_{v=1}^{m}\rho_v(y_{v,t} \mid x_t)^{\frac{2|E(G)|}{kd(v)}\times M(v,k)} \tag{27}$$

$$\rho_t(x_t) = \prod_{v=1}^{m}\rho_v(y_{v,t} \mid x_t) \tag{28}$$

for a given set of measurements $\{y_{v,t}\}_{v=1}^m$. Lemma 7 below shows that the expectation of the difference of likelihood function between the MCDPF and CPF given a set of particles is bounded above.

Lemma 7: Suppose we have the MCDPF on a finite, connected graph with a measurement density ρ that factorizes as in (12) and is bounded above and away from zero, so that $0 < \rho_0 \leq \rho(y_t \mid x_t) \leq \|\rho\|_\infty < \infty$. Then there exists $k^\star$ such that $\frac{2}{\sqrt{k^\star}} \leq \frac{1}{\sqrt[3]{3}L\vee L'}$ and we have the upper bound on the expected error between $\rho_{t,k}$ and ρ_t for $k \geq k^\star$ given by

$$\mathbf{E}\left[|\rho_{t,k}(x_t) - \rho_t(x_t)|^2\right] \leq \|\rho\|_\infty^2\Phi(k,\nu),$$

where the expectation is taken with respect to the dynamics of the particle on the Markov chain, and

$$\begin{aligned}&\Phi(k,\nu) \\ &= \frac{1}{k}\left[2C'L'\left(e^{\frac{(2L')^2}{k}}+e^{\frac{L'^2}{k}}\right)+C\left(e^{\frac{(2\sqrt[3]{3}L)^2}{k}}+2e^{\frac{(\sqrt[3]{3}L)^2}{k}}\right) + 2C'L'\left(\frac{1}{k-(2L')^2}+\frac{1}{k-(L')^2}\right) + C\left(\frac{1}{k-(2\sqrt[3]{3}L)^2}+\frac{2}{k-(\sqrt[3]{3}L)^2}\right)\right] \\ &\quad + \left(e^{\frac{2\gamma_g^2}{k}}-2e^{\frac{\gamma_g^2}{2k}}+1\right).\end{aligned} \tag{29}$$

The positive constants L, L', C, and C' depend on the SLEM of the Markov chain transition matrix as defined in (22)–(25).

Proof: For simplicity of notation, we omit the arguments of the function $\rho_v(y_{v,t} \mid x_t)$. We compute

$$\begin{aligned}
&\mathbf{E}\left[|\rho_{t,k} - \rho_t|^2\right] \\
&= \mathbf{E}\left[\left(\prod_{v=1}^{m} \rho_v^{\frac{2|E(G)|}{kd(v)} \times M(v,k)} - \prod_{v=1}^{m} \rho_v\right)^2\right] \\
&= \left(\prod_{v=1}^{m} \rho_v^2\right) \mathbf{E}\left[\left(\prod_{v=1}^{m} \rho_v^{\frac{2|E(G)|}{kd(v)} \times M(v,k)-1} - 1\right)^2\right] \\
&= \rho_t^2 \mathbf{E}\left[\left(\prod_{v=1}^{m} \rho_v^{\frac{2|E(G)|}{d(v)}\left(\frac{M(v,k)}{k} - \frac{d(v)}{2|E(G)|}\right)} - 1\right)^2\right] \\
&= \rho_t^2 \mathbf{E}\left[\left(\prod_{v=1}^{m} \rho_v^{\frac{2|E(G)|}{d(v)} Z_{k,v}} - 1\right)^2\right] \\
&= \rho_t^2 \mathbf{E}\left[\left(e^{\sum_{v=1}^{m} C_v Z_{k,v}} - 1\right)^2\right],
\end{aligned}$$

where $C_v = \frac{2|E(G)|}{d(v)} \ln \rho_v$ and $Z_{k,\cdot} = \frac{M(\cdot,k)}{k} - \frac{d(\cdot)}{2|E(G)|}$.

Let $(X_i)_{i=0}^{\infty}$ be a Markov chain on the state space $\mathcal{X}$ starting at the node for the particle x_t, so that $M(v,k)$ is the visitation count of X_i to node v within k steps. Let the function $g : \mathcal{X} \to \mathbb{R}$ be defined as $g(X_i) = \sum_{v=1}^{m} I(X_i = v) C_v$. Now we have

$$\begin{aligned}
\sum_{v=1}^{m} C_v Z_{k,v} &= \sum_{v=1}^{m} C_v \frac{M(v,k)}{k} - \sum_{v=1}^{m} C_v \frac{d(v)}{2|E(G)|} \\
&= \sum_{v=1}^{m} C_v \frac{M(v,k)}{k} - \sum_{v=1}^{m} \ln \rho_v \\
&= \frac{1}{k} \sum_{i=0}^{k-1} g(X_i) - \ln \rho_t \\
&= \frac{1}{k} \sum_{i=0}^{k-1} g(X_i) - \Pi(g).
\end{aligned}$$

Defining $W_k(g) = \sqrt{k} \sum_{v=1}^{m} C_v Z_{k,v}$, the Markov chain central limit theorem gives the convergence of $W_k(g)$ to the normal random variable with mean zero and variance γ_g^2 defined in (20). With the definition of $W_k(g)$ and a standard normal random variable U, the following expectation over $W_k(g)$ is

$$\begin{aligned}
&\mathbf{E}\left[\left(e^{\frac{1}{\sqrt{k}} W_k(g)} - 1\right)^2\right] \\
&= \mathbf{E}\left[e^{\frac{2}{\sqrt{k}} W_k(g)} - 2e^{\frac{1}{\sqrt{k}} W_k(g)} + 1\right] \\
&\le \left|\mathbf{E}\left[e^{\frac{2}{\sqrt{k}} W_k(g)}\right] - \mathbf{E}\left[e^{\frac{2}{\sqrt{k}} \gamma_g U}\right]\right| \\
&\quad + 2\left|\mathbf{E}\left[e^{\frac{1}{\sqrt{k}} W_k(g)}\right] - \mathbf{E}\left[e^{\frac{1}{\sqrt{k}} \gamma_g U}\right]\right| \\
&\quad + \mathbf{E}\left[e^{\frac{2}{\sqrt{k}} \gamma_g U}\right] - 2\mathbf{E}\left[e^{\frac{1}{\sqrt{k}} \gamma_g U}\right] + 1 \\
&= \left|\mathbf{E}\left[e^{\frac{2}{\sqrt{k}} W_k(g)}\right] - \mathbf{E}\left[e^{\frac{2}{\sqrt{k}} \gamma_g U}\right]\right| \\
&\quad + 2\left|\mathbf{E}\left[e^{\frac{1}{\sqrt{k}} W_k(g)}\right] - \mathbf{E}\left[e^{\frac{1}{\sqrt{k}} \gamma_g U}\right]\right| \\
&\quad + e^{\frac{2\gamma_g^2}{k}} - 2e^{\frac{\gamma_g^2}{2k}} + 1.
\end{aligned}$$

Since $\frac{1}{\sqrt{k}}$ is decreasing in k, there exists $k^\star$ such that $\frac{2}{\sqrt{k^\star}} \le \frac{1}{L \vee L'}$, where L and L' are defined in Theorem 5. Furthermore if we define c by

$$c(x_t) = \max_v \left| \ln \frac{\rho_v(y_{v,t} \mid x_t)^{\frac{2|E(G)|}{d(v)}}}{\rho_0} \right|$$

then it satisfies the condition $|g(x) - \Pi(g)| = |C_j - \ln \rho_t| \le c(x_t)$ as $\rho_0 \le \rho_t$. Therefore, Theorem 5 implies that for $k \ge k^\star$,

$$\begin{aligned}
&\mathbf{E}\left[\left(e^{\frac{1}{\sqrt{k}} W_k(g)} - 1\right)^2\right] \\
&\le \frac{1}{k}\left[2C'L'\left(e^{\frac{(2L')^2}{k}} + e^{\frac{L'^2}{k}}\right)\right. \\
&\quad + C\left(e^{\frac{(2\sqrt[3]{3}L)^2}{k}} + 2e^{\frac{(\sqrt[3]{3}L)^2}{k}}\right) \\
&\quad + 2C'L'\left(\frac{1}{k - (2L')^2} + \frac{1}{k - (L')^2}\right) \\
&\quad \left. + C\left(\frac{1}{k - (2\sqrt[3]{3}L)^2} + \frac{2}{k - (\sqrt[3]{3}L)^2}\right)\right] \\
&\quad + e^{\frac{2\gamma_g^2}{k}} - 2e^{\frac{\gamma_g^2}{2k}} + 1.
\end{aligned}$$

Multiplying by ρ_t^2 and using $\rho_t^2 \le \|\rho\|_\infty^2$ gives the desired result. ■

Now we prove mean square convergence of the MCDPF by considering in turn each of the steps of prediction, measurement, and resampling. This proof is a modification of [34, Lemma 3–5]. The main difference from that proof is that the effect of Markov chain step k must be considered and we obtain an explicit bound for the mean square error in terms of the SLEM of the underlying Markov chain.

We begin with the following lemma for the prediction update step.

Lemma 8: Assume that time dependent constants ϕ_{t-1} and $c_{t-1|t-1}$ exists and for any $\varphi \in B(\mathbb{R}^n)$,

$$\begin{aligned}
&\mathbf{E}\{[(\pi^N_{t-1|t-1}, \varphi) - (\pi_{t-1|t-1}, \varphi)]^2\} \\
&\qquad \le \|\varphi\|_\infty^2 \left(\phi_{t-1}\sqrt{\Phi(k,\nu)} + \frac{\sqrt{c_{t-1|t-1}}}{\sqrt{N}}\right)^2. \quad (30)
\end{aligned}$$

Then

$$\begin{aligned}
&\mathbf{E}\{[(\pi^N_{t|t-1}, \varphi) - (\pi_{t|t-1}, \varphi)]^2\} \\
&\qquad \le \|\varphi\|_\infty^2 \left(\phi_{t-1}\sqrt{\Phi(k,\nu)} + \frac{\sqrt{c_{t|t-1}}}{\sqrt{N}}\right)^2,
\end{aligned}$$

where $c_{t|t-1} = (2 + \sqrt{c_{t-1|t-1}})^2$.

Proof: From Minkowski's inequality,

$$\begin{aligned}
&\{\mathbf{E}[(\pi^N_{t|t-1}, \varphi) - (\pi_{t|t-1}, \varphi)]^2\}^{\frac{1}{2}} \\
&\le \{\mathbf{E}[(\pi^N_{t|t-1}, \varphi) - (\pi^N_{t-1|t-1}, K\varphi)]^2\}^{\frac{1}{2}} \\
&\quad + \{\mathbf{E}[(\pi^N_{t-1|t-1}, K\varphi) - (\pi_{t-1|t-1}, K\varphi)]^2\}^{\frac{1}{2}}. \quad (31)
\end{aligned}$$

The first term on the right hand side of (31) is bounded above as follows. From Lemmas 6 and the equality, $\mathbf{E}\left[(\pi^N_{t|t-1}, \varphi) \mid \mathcal{G}_{t-1}\right] = (\pi^N_{t-1|t-1}, K\varphi)$ where $\mathcal{G}_{t-1}$ is

the σ-algebra generated by $\{x_{t-1}^i\}_{i=1}^N$, for the first inequality and the first equality,

$$\begin{aligned}
&\mathbf{E}\left[|(\pi_{t|t-1}^N,\varphi)-(\pi_{t-1|t-1}^N,K\varphi)|^2 \mid \mathcal{G}_{t-1}\right]\\
&=\mathbf{E}\left[\left|(\pi_{t|t-1}^N,\varphi)-\mathbf{E}[(\pi_{t|t-1}^N,\varphi)\mid \mathcal{G}_{t-1}]\right|^2 \mid \mathcal{G}_{t-1}\right]\\
&=\frac{1}{N^2}\mathbf{E}\left[\left|\sum_{i=1}^N \varphi(x_t^i)-\mathbf{E}[\varphi(x_t^i)\mid \mathcal{G}_{t-1}]\right|^2 \mid \mathcal{G}_{t-1}\right]\\
&=\frac{1}{N^2}\sum_{i=1}^N \mathbf{E}\left[\left|\varphi(x_t^i)-\mathbf{E}[\varphi(x_t^i)\mid \mathcal{G}_{t-1}]\right|^2 \mid \mathcal{G}_{t-1}\right]\\
&\le \frac{4}{N^2}\sum_{i=1}^N \mathbf{E}\left[\varphi^2(x_t^i)\mid \mathcal{G}_{t-1}\right]\\
&=\frac{4}{N}\mathbf{E}\left[(\pi_{t|t-1}^N,\varphi^2)\mid \mathcal{G}_{t-1}\right]\\
&=\frac{4}{N}(\pi_{t-1|t-1}^N,K\varphi^2). \qquad (32)
\end{aligned}$$

Since Markov operators are contractions [41], $\|K\varphi\|_\infty \le \|\varphi\|_\infty$, and so

$$\mathbf{E}\left[|(\pi_{t|t-1}^N,\varphi)-(\pi_{t-1|t-1}^N,K\varphi)|^2 \mid \mathcal{G}_{t-1}\right]\le \frac{4}{N}\|\varphi\|_\infty^2 .$$

The upper bound of the second term (31) is given by assumption (30) in the lemma together with the contracting behavior of K, so that

$$\mathbf{E}\left[|(\pi_{t-1|t-1}^N,K\varphi)-\left(\pi_{t-1|t-1},K\varphi\right)|^2\right] \le \|\varphi\|_\infty^2\left(\phi_{t-1}\sqrt{\Phi(k,\nu)}+\frac{\sqrt{c_{t-1|t-1}}}{\sqrt{N}}\right)^2 .$$

Thus

$$\{\mathbf{E}[(\pi_{t|t-1}^N,\varphi)-(\pi_{t|t-1},\varphi)]^2\}^{\frac{1}{2}} \le \|\varphi\|_\infty\left(\phi_{t-1}\sqrt{\Phi(k,\nu)}+\frac{\sqrt{c_{t|t-1}}}{\sqrt{N}}\right),$$

where $\sqrt{c_{t|t-1}}=2+\sqrt{c_{t-1|t-1}}$. ■

Given the result of the prediction step, together with Lemma 7, the following lemma gives the error bound after the Markov chain iteration step.

Lemma 9: Assume that for any $\varphi\in B(\mathbb{R}^n)$ and given measurements,

$$\mathbf{E}\{[(\pi_{t|t-1}^N,\varphi)-(\pi_{t|t-1},\varphi)]^2\} \le \|\varphi\|_\infty^2\left(\phi_{t-1}\sqrt{\Phi(k,\nu)}+\frac{\sqrt{c_{t|t-1}}}{\sqrt{N}}\right)^2$$

and ρ is a measurement density satisfying the requirements of Lemma 7. Then

$$\mathbf{E}\{[(\pi_{t|t-1}^N,\rho_{t,k}\varphi)-(\pi_{t|t-1},\rho_t\varphi)]^2\} \le \|\varphi\|_\infty^2\left(\tilde{\phi}_t\sqrt{\Phi(k,\nu)}+\frac{\sqrt{c_{t|t-1}}}{\sqrt{N}}\right)^2,$$

where $\tilde{\phi}_t=\phi_{t-1}+\|\rho\|_\infty$ and $\rho_{t,k}$ and ρ_t are defined by (27) and (28).

Proof: From Minkowski's inequality,

$$\begin{aligned}
&\{\mathbf{E}[(\pi_{t|t-1}^N,\rho_{t,k}\varphi)-(\pi_{t|t-1},\rho_t\varphi)]^2\}^{\frac{1}{2}}\\
&\le\{\mathbf{E}[(\pi_{t|t-1}^N,\rho_{t,k}\varphi)-(\pi_{t|t-1}^N,\rho_t\varphi)]^2\}^{\frac{1}{2}}\\
&\quad+\{\mathbf{E}[(\pi_{t|t-1}^N,\rho_t\varphi)-(\pi_{t|t-1},\rho_t\varphi)]^2\}^{\frac{1}{2}}. \qquad (33)
\end{aligned}$$

The first term (33) on the right hand side is bounded as follows, using Lemma 8.

$$\begin{aligned}
&\mathbf{E}\{[(\pi_{t|t-1}^N,\rho_{t,k}\varphi)-(\pi_{t|t-1}^N,\rho_t\varphi)]^2\}\\
&=\mathbf{E}\left[\frac{1}{N^2}\left(\sum_{i=1}^N(\rho_{t,k}(x_t^i)\varphi(x_t^i)-\rho_t(x_t^i)\varphi(x_t^i))\right)^2\right]\\
&\le q\mathbf{E}\left[\frac{1}{N}\sum_{i=1}^N(\rho_{t,k}(x_t^i)\varphi(x_t^i)-\rho_t(x_t^i)\varphi(x_t^i))^2\right]\\
&\le\mathbf{E}\left[\frac{\|\varphi\|_\infty^2}{N}\sum_{i=1}^N(\rho_{t,k}(x_t^i)-\rho_t(x_t^i))^2\right]\\
&\le\|\varphi\|_\infty^2\|\rho\|_\infty^2\Phi(k,\nu).
\end{aligned}$$

The first inequality is a trivial application of Hölder's inequality. Thus

$$\{\mathbf{E}[(\pi_{t|t-1}^N,\rho_{t,k}\varphi)-(\pi_{t|t-1},\rho_t\varphi)]^2\}^{\frac{1}{2}} \le\|\varphi\|_\infty\left((\phi_{t-1}+\|\rho\|_\infty)\sqrt{\Phi(k,\nu)}+\frac{\sqrt{c_{t|t-1}}}{\sqrt{N}}\right).$$

■

The next MCDPF step is a measurement update and the following lemma provides an error bound after this step.

Lemma 10: From the result of lemma 9, assume that for any $\varphi\in B(\mathbb{R}^n)$,

$$\mathbf{E}\{[(\pi_{t|t-1}^N,\rho_{t,k}\varphi)-(\pi_{t|t-1},\rho_t\varphi)]^2\} \le\|\varphi\|_\infty^2\left(\tilde{\phi}_t\sqrt{\Phi(k,\nu)}+\frac{\sqrt{c_{t|t-1}}}{\sqrt{N}}\right)^2.$$

Then

$$\mathbf{E}\{[(\tilde{\pi}_{t|t,k}^N,\varphi)-(\pi_{t|t},\varphi)]^2\} \le\|\varphi\|_\infty^2\left(\phi_t\sqrt{\Phi(k,\nu)}+\frac{\sqrt{\tilde{c}_{t|t}}}{\sqrt{N}}\right)^2.$$

Proof: We have

$$\begin{aligned}
&(\tilde{\pi}_{t|t,k}^N,\varphi)-(\pi_{t|t},\varphi)\\
&\quad=\frac{(\pi_{t|t-1}^N,\rho_{t,k}\varphi)}{(\pi_{t|t-1}^N,\rho_{t,k})}-\frac{(\pi_{t|t-1},\rho\varphi)}{(\pi_{t|t-1},\rho)}\\
&\quad=\frac{(\pi_{t|t-1}^N,\rho_{t,k}\varphi)}{(\pi_{t|t-1}^N,\rho_{t,k})}-\frac{(\pi_{t|t-1}^N,\rho_{t,k}\varphi)}{(\pi_{t|t-1},\rho)}\\
&\qquad+\frac{(\pi_{t|t-1}^N,\rho_{t,k}\varphi)}{(\pi_{t|t-1},\rho)}-\frac{(\pi_{t|t-1},\rho\varphi)}{(\pi_{t|t-1},\rho)}.
\end{aligned}$$

These terms can be bounded by

$$\left|\frac{(\pi^N_{t|t-1},\rho_{t,k}\varphi)}{(\pi^N_{t|t-1},\rho_{t,k})}-\frac{(\pi^N_{t|t-1},\rho_{t,k}\varphi)}{(\pi_{t|t-1},\rho)}\right| = \frac{(\pi^N_{t|t-1},\rho_{t,k}\varphi)\left|(\pi^N_{t|t-1},\rho_{t,k})-(\pi_{t|t-1},\rho)\right|}{(\pi^N_{t|t-1},\rho_{t,k})(\pi_{t|t-1},\rho)} \le \frac{\|\varphi\|_\infty\left|(\pi^N_{t|t-1},\rho_{t,k})-(\pi_{t|t-1},\rho)\right|}{(\pi_{t|t-1},\rho)}$$

and

$$\left|\frac{(\pi^N_{t|t-1},\rho_{t,k}\varphi)}{(\pi_{t|t-1},\rho)}-\frac{(\pi_{t|t-1},\rho\varphi)}{(\pi_{t|t-1},\rho)}\right| = \frac{\left|(\pi^N_{t|t-1},\rho_{t,k}\varphi)-(\pi_{t|t-1},\rho\varphi)\right|}{(\pi_{t|t-1},\rho)}.$$

Using Minkowski's inequality again gives

$$\begin{aligned}&\{\mathbf{E}[(\tilde\pi^N_{t|t,k},\varphi)-(\pi_{t|t},\varphi)]^2\}^{\frac12}\\ &\le \frac{\|\varphi\|_\infty}{(\pi_{t|t-1},\rho)}\{\mathbf{E}[(\pi^N_{t|t-1},\rho_{t,k})-(\pi_{t|t-1},\rho)]^2\}^{\frac12}\\ &\quad+\frac{1}{(\pi_{t|t-1},\rho)}\{\mathbf{E}[(\pi^N_{t|t-1},\rho_{t,k}\varphi)-(\pi_{t|t-1},\rho\varphi)]^2\}^{\frac12}\\ &\le \frac{2\|\varphi\|_\infty}{(\pi_{t|t-1},\rho)}\left(\tilde\phi_t\sqrt{\Phi(k,\nu)}+\frac{\sqrt{c_{t|t-1}}}{\sqrt{N}}\right),\end{aligned}$$

where $\phi_t=\frac{2\tilde\phi_t}{(\pi_{t|t-1},\rho)}$ and $\tilde c_{t|t}=\frac{4c_{t|t-1}}{(\pi_{t|t-1},\rho)^2}$. ■

The next lemma provides the error bound after the resampling step, which completes the steps in the MCDPF.

Lemma 11: Assume that for any $\varphi\in B(\mathbb{R}^n)$,

$$\mathbf{E}\{[(\tilde\pi^N_{t|t,k},\varphi)-(\pi_{t|t},\varphi)]^2\} \le \|\varphi\|^2_\infty\left(\phi_t\sqrt{\Phi(k,\nu)}+\frac{\sqrt{\tilde c_{t|t}}}{\sqrt{N}}\right)^2.$$

Then

$$\mathbf{E}\{[(\pi^N_{t|t,k},\varphi)-(\pi_{t|t},\varphi)]^2\} \le \|\varphi\|^2_\infty\left(\phi_t\sqrt{\Phi(k,\nu)}+\frac{\sqrt{c_{t|t}}}{\sqrt{N}}\right)^2.$$

Proof: We have

$$(\pi^N_{t|t,k},\varphi)-(\pi_{t|t},\varphi) = (\pi^N_{t|t,k},\varphi)-(\tilde\pi^N_{t|t,k},\varphi)+(\tilde\pi^N_{t|t,k},\varphi)-(\pi_{t|t},\varphi).$$

If $\mathcal{H}_t$ is a σ-algebra generated by particles $\{\tilde x^i_t\}^N_{i=1}$, then

$$\mathbf{E}\left[(\pi^N_{t|t,k},\varphi)\,\middle|\,\mathcal{H}_t\right]=\frac1N\sum_{i=1}^N\mathbf{E}\varphi(x^i_t) =\frac1N\sum_{i=1}^N w^i_t\varphi(x^i_t)=(\tilde\pi^N_{t|t,k},\varphi).$$

Thus by the same procedure from (32) in Lemma 8, for some constant $\bar C$,

$$\mathbf{E}\left[((\pi^N_{t|t,k},\varphi)-(\tilde\pi^N_{t|t,k},\varphi))^2\,\middle|\,\mathcal{H}_t\right]\le\bar C\frac{\|\varphi\|^2_\infty}{N}.$$

Finally we have

$$\{\mathbf{E}[(\pi^N_{t|t,k},\varphi)-(\pi_{t|t},\varphi)]^2\}^{\frac12} \le\sqrt{\bar C}\frac{\|\varphi\|_\infty}{\sqrt{N}}+\|\varphi\|_\infty\left(\phi_t\sqrt{\Phi(k,\nu)}+\frac{\sqrt{\tilde c_{t|t}}}{\sqrt{N}}\right),$$

where $\sqrt{c_{t|t}}=\sqrt{\bar C}+\sqrt{\tilde c_{t|t}}$. ■

Combing the above sequence of lemmas now gives the following main result.

Theorem 12: Consider a connected sensor network with given measurements at different nodes conditionally independent given the true state. Furthermore assume that the kernel K is Feller and the function ρ is bounded above, continuous, positive, and bounded away from zero. Then there exist time dependent constants $c_{t|t}$ and ϕ_t, and a constant $\Phi(k,\nu)$ given in (29) depending on the number of Markov chain steps k and the SLEM ν of the Markov chain, such that

$$\{\mathbf{E}[(\pi^N_{t|t,k},\varphi)-(\pi_{t|t},\varphi)]^2\}^{\frac12} \le\|\varphi\|_\infty\left(\phi_t\sqrt{\Phi(k,\nu)}+\frac{\sqrt{c_{t|t}}}{\sqrt{N}}\right) \quad (34)$$

for all $t\ge0$ and bounded Borel measurable $\varphi\in B(\mathbb{R}^n)$. Thus, the root mean square error of the MCDPF converges proportionally to $O(\frac{1}{\sqrt{k}}+\frac{1}{\sqrt{N}})$ and increases proportionally to $O(\frac{1}{\sqrt{\delta}}e^{\frac{1}{\delta}})$ as $\delta\to0$, where δ is the spectral gap $\delta=1-\nu$.

Proof: Combining Lemmas 8, 9, 10, and 11 gives the result. The constants $c_{t|t}$ and ϕ_t are defined recursively as specified in Lemmas 8, 9, and 10. The convergence rate of the root mean square error of the MCDPF in terms of the number of particles N is easily obtained from the upper bound in (34). We also have $\Phi(k,\nu)=O(k^{-1})$ for $k\ge k^\star$ and a fixed ν as given in (29). Similarly, if k is fixed then $\Phi(k,\nu)=O(\frac{e^{\frac{1}{\delta^2}}}{\delta})$ because the $Ce^{\frac{(2\sqrt[3]{3}L)^2}{k}}$ term dominates and $C=O(\delta^{-1})$ and $L=O(\delta^{-1})$.

■

V. Numerical Results

To quantify the behavior of the distributed filter described in this paper, we consider a tracking problem for K vehicles moving in the plane, with interaction dynamics specified by a nonlinear flocking model due to [42] (see also [27]). This vehicle movement control law includes global cooperation and local collision avoidance. There are 4 sensors in the plane, each of which measures the distance between its position and each of the K vehicles, giving a total of $4K$ distances.

There is Gaussian process and measurement noise, both with zero mean. The covariance matrix of the process noise is proportional to Δt^2, the square of the timestep, while the covariance of the measurement noise is proportional to the distance from the sensor to the vehicle. In the sensor network, each node is connected to two neighbors, forming a square as shown in

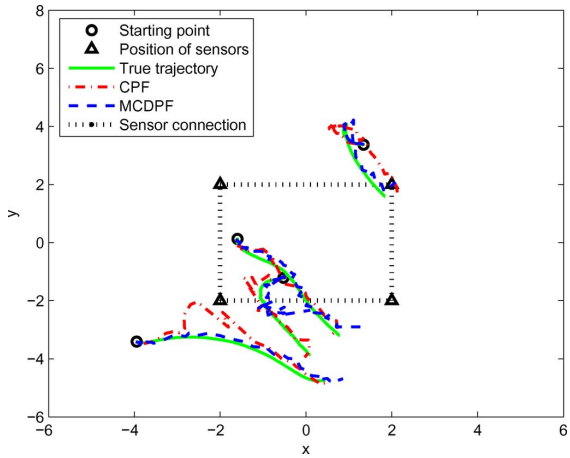

Fig. 1. Sample estimated trajectories for the centralized (CPF) and decentralized (MCDPF) particle filters, for the range-measurement-only vehicle tracking problem described in Section V.

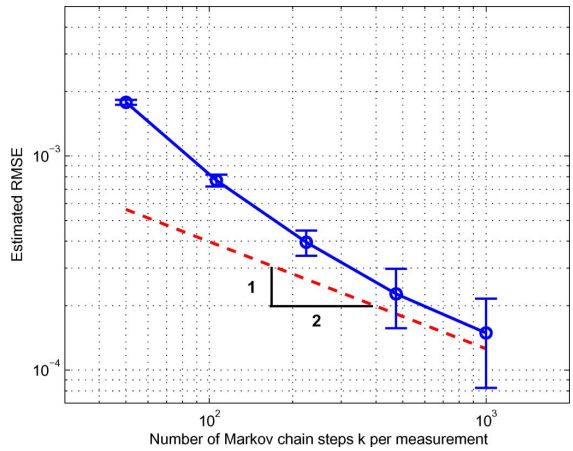

Fig. 2. RMSE between the mean state estimated by the MCDPF and CPF, plotted against the number of Markov chain steps k per measurement in the MCDPF. The error bars for the estimated RMSE are the 95% confidence intervals from Student's t-test. The dashed line in the lower figure has slope $-\frac{1}{2}$ and is provided for reference, showing that these results agree with Theorem 12.

Fig. 1. Therefore the normalized adjacency matrix in this case is

$$\mathcal{A} = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix}.$$

where the nodes are numbered clockwise around the square. See [27] for full details of this system.

We compare six different estimators for this problem: the Extended Kalman Filter (EKF), the Distributed EKF (DEKF), the Regularized EKF (REKF), the Regularized DEKF (RDEKF), the Centralized Particle Filter (CPF), and the Markov Chain Distributed Particle Filter (MCDPF) described in this paper. The EKF is the standard filter (see, e.g., [1]), while the DEKF is distributed using the consensus method of [19], [43]. Due to the nonlinearities in the system dynamics, both the EKF and the DEKF become unstable at moderate timestep values. To mitigate this, we define regularized versions (REKF and RDEKF) which have mollified system dynamics and for which, if the condition number of the information matrix exceeds 10^6, the singular values of the information matrix are constrained to lie within $[1, 10^6]$ by taking an SVD and capping values outside this range (see [27] for details). These values were chosen heuristically to maximize estimation accuracy. The CPF and MCDPF filters are as described in Sections III-A and III-B, implemented with a bootstrap prediction step and varying number of particles N and number of Markov chain steps k.

A. Numerical Evidence of Mean Square Convergence

We first consider the convergence of the distributed MCDPF to the centralized CPF particle filter as a function of the number of Markov chain steps k, showing the $O(\frac{1}{\sqrt{k}})$ behavior as $k \to \infty$ as predicted by Theorem 12. The RMSE (root mean square error) between the MCDPF estimate and the CPF estimate is defined by

$$\mathrm{RMSE}^N_{t|t,k} \triangleq \left(\mathbf{E}\left[\left\| \hat{x}^N_{t|t,k} - \hat{x}^N_{t|t} \right\|^2 \right] \right)^{\frac{1}{2}}, \qquad (35)$$

where $\hat{x}^N_{t|t,k}$ is the mean of the state distribution estimated by the MCDPF at time t with N particles and k Markov chain steps per measurement. The mean state estimate $\hat{x}_{\mathrm{CPF}}$ for the CPF is defined similarly.

Fig. 2 shows the RMSE versus the number of Markov chain steps k per measurement. Both the MCDPF and CPF used $N = 400$ particles, one vehicle, four sensors connected in a square network, timestep $\Delta t = 0.01$, and are at time $t = 0.1$. At $t = 0$ the number of particles at each node is 100, so $N(v) = 100$ for $v = 1, 2, 3, 4$. As time advances, $N(v)$ changes because each particle is moved between the nodes at random (according to the Markov chain random walk on the square). The RMSE was estimated from an ensemble of 2.5×10^5 simulations (see [27] for simulation details).

We can see that the decay rate of the RMSE in the figure approaches $O(\frac{1}{\sqrt{k}})$. As it is known [34] that the CPF converges to the optimal filter at rate $O(\frac{1}{\sqrt{N}})$, this confirms the theoretical result in Theorem 12.

B. Performance Comparison With EKF and DEKF

We compare the performance of the six filters outlined above for the vehicle tracking problem. The performance is compared with respect to the system complexity and measurement frequency and it shows that the MCDPF method is more robust and more accurate with complex systems or low measurement frequency, compared to the DEKF and RDEKF. To quantify performance, we consider the RMSE between the true system trajectory and the mean of the state position distribution estimated by the different filters, normalized by the number of vehicles and scaled by the timestep.

An estimator is said to diverge when the RMSE is higher than the static estimator that always returns the initial condition. The filters were run with $N = 100$ particles for CPF and MCPDF, $k = 10$ Markov chain iteration steps for MCDPF, and 2 consensus iteration steps for Regularized DEKF or DEKF. The position RMSE computed with $\Delta t = 0.01$ is plotted versus time in Fig. 3.

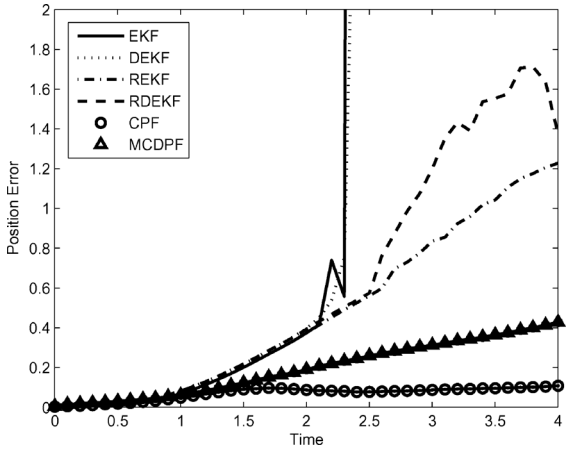

Fig. 3. RMSE in position versus time for the EKF, DEKF, REKF, RDEKF, CPF and MCDPF filters. We see that the non-regularized filters diverge rapidly, while the particle filters perform well.

TABLE I
ALGORITHM PERFORMANCE: RMSE VALUES AND THE FRACTION OF DIVERGENT RUNS (AVERAGED OVER 1000 MONTE CARLO RUNS)

4 vehicles	$\Delta t = 0.1$		$\Delta t = 0.05$	
Algorithm	RMSE	Divergence	RMSE	Divergence
EKF	N/A	49.7%	0.1236	0%
DEKF	N/A	62.3%	0.1733	0.1%
Regularized EKF	1.1986	2.1%	0.1164	0%
Regularized DEKF	2.5427	2.3%	0.1579	0%
CPF	0.2871	0%	0.1921	0%
MCDPF	0.8011	0%	0.7453	0%

10 vehicles	$\Delta t = 0.1$		$\Delta t = 0.05$	
Algorithm	RMSE	Divergence	RMSE	Divergence
EKF	N/A	100%	N/A	100%
DEKF	N/A	100%	N/A	100%
Regularized EKF	1.7108	8%	1.5674	2.9%
Regularized DEKF	12.7356	16.7%	5.8275	6.1%
CPF	0.4987	0%	0.2434	0%
MCDPF	0.7416	0%	0.4807	0%

The filter results are given in Table I with varying numbers of vehicles (changing system complexity) and Δt (changing measurement frequency). We see that for sufficiently small Δt the Kalman-filter-based approaches are superior, whereas for large Δt with significant nonlinearities the particle filters are better. This is due to the fact that the accuracy of the linearization varies inversely with Δt. Secondly, we see that increased system complexity dramatically reduces the performance of the Kalman filter based approaches, which no longer give acceptable performance even for the smaller time step while the particle filters (including the MCDPF) are robust with respect to the increase in system complexity, with very little loss of accuracy. The distributed filters (DEKF, RDEKF, and MCDPF) all perform worse than their centralized equivalents (EKF, REKF, and CPF, respectively), as expected, but converge as the iteration frequency increases (see Fig. 2 and [27]).

VI. CONCLUSION

In this paper we proved convergence of the Markov Chain Distributed Particle Filter (MCDPF) in the sense of mean square to the optimal filter as the number of particles N and the number of Markov chain steps per measurement k both go to infinity, extending the previously known almost sure convergence of the MCDPF [24]. In proving mean square convergence, we gave an explicit error bound for the MCDPF, showing that the convergence rate to the optimal filter is of order $O(\frac{1}{\sqrt{k}} + \frac{1}{\sqrt{N}})$. This theoretical result confirms earlier numerical results [24], [27] that suggested convergence at this rate.

We are also able to see that the error of the MCDPF increases at rate $O(\frac{e^{\frac{1}{\delta^2}}}{\delta})$ as the spectral gap $\delta = 1 - \nu$ goes to zero, where ν is the modulus of the second-largest eigenvalue (SLEM) of the Markov chain on the sensor network. This shows that fast-mixing Markov chains (with large spectral gaps) will produce better estimates, as we would expect, giving guidance as to how the mixing process could be optimized. We have not explored the question of how the particle exchange process should be chosen to improve the filter performance, although it seems likely that large sensor networks will require hierarchical or multiscale communication (e.g., [44], [45]) for good performance. Particle filters typically experience significant degradation in accuracy as the system dimension increases, leading to a classical curse of dimensionality [46]–[48], and the MCDPF will also exhibit this behavior.

The MCDPF provides a distributed particle filter algorithm that exchanges only particles (state estimates) between sensor nodes. It thus makes no assumptions on the estimated distribution or dynamics over a regular centralized particle filter. While these properties are very attractive, and the simplicity of the MCPDF makes implementation easy, this paper does not consider how the efficiency of the MCDPF compares to other distribution methods, such as those based on Gaussian Mixture Models or similar approximations. We expect that the MCDPF will be less efficient for low-dimensional systems, but will scale well as the dimension of the state space increases. Some initial comparison results [27] have started to investigate this. The per-step computational cost of the MCDPF is $O\left(\frac{knN}{m}\right)$, so it scales linearly in the state dimension n, the number of Markov chain steps per measurement k, and the number $\frac{N}{m}$ of particles per sensor node. See [27] for further details regarding the computational complexity of the method and comparisons with centralized and distributed Kalman filters.

The bounds for the MCDPF error derived in this paper are not uniform in time. This could be addressed by introducing a mixing assumption on the transition kernel, in the spirit of [49] or [50] (see [34] for a discussion in the centralized case). We have also not proved here that the MCDPF converges in mean square to the Centralized Particle Filter (CPF) for a fixed number of particles N, as k goes to infinity. The numerical example in this paper suggests that this is indeed true, and it seems likely that a modification of the our techniques would also give such a result.

REFERENCES

[1] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.
[2] H. Akashi and H. Kumamoto, "Construction of discrete-time nonlinear filter by Monte Carlo methods with variance-reducing techniques," *Syst. Contr.*, vol. 19, pp. 211–221, 1975.
[3] J. E. Handschin and D. Q. Mayne, "Monte Carlo techniques to estimate the conditional expectation in multi-stage non-linear filtering," *Int. J. Contr.*, vol. 9, pp. 547–559, 1969.
[4] J. E. Handschin, "Monte Carlo techniques for prediction and filtering of non-linear stochastic processes," *Automatica*, vol. 6, pp. 555–563, 1970.
[5] V. S. Zaritskii, V. B. Svetnik, and L. Shimelevich, "Monte Carlo technique in problems of optimal data processing," *Automat. Remote Contr.*, vol. 12, pp. 95–103, 1975.

[6] J. M. Hammersley and K. W. Morton, "Symposium on Monte Carlo methods: Poor man's Monte Carlo," *J. R. Stat. Soc., Ser. B*, vol. 16, pp. 23–38, 1954.
[7] M. N. Rosenbluth and A. Rosenbluth, "Monte Carlo calculation of the average extension of molecular chains," *J. Chem. Phys.*, vol. 23, pp. 356–359, 1956.
[8] P. M. Djurić, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Míguez, "Particle filtering," *IEEE Signal Process. Mag.*, vol. 20, no. 5, pp. 19–38, Sep. 2003.
[9] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," in *Proc. IEEE Radar Signal Process.*, Apr. 1993, vol. 140, pp. 107–113.
[10] S. Maskell and N. Gordon, "A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, pp. 174–188, 2001.
[11] J. S. Liu and R. Chen, "Sequential Monte Carlo methods for dynamic systems," *J. Amer. Statist. Assoc.*, vol. 93, pp. 1032–1044, 1998.
[12] G. Kitagawa, "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models," *J. Computat. Graph. Statist.*, vol. 5, no. 1, pp. 1–25, 1996.
[13] M. K. Pitt and N. Shephard, "Filtering via simulation: Auxiliary particle filters," *J. Amer. Statist. Associat.*, vol. 94, no. 446, pp. 590–599, 1999.
[14] C. Musso, N. Oudjane, and F. Legland, "Improving regularized particle filters," in *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds. New York: Springer, 2001, Statist. Eng. Inf. Sci., pp. 247–271.
[15] M. Rosencrantz, G. Gordon, and S. Thrun, "Decentralized sensor fusion with distributed particle filters," in *Proc. UAI*, 2003, pp. 493–500.
[16] D. Gu, "Distributed particle filter for target tracking," in *Proc. IEEE Int. Conf. Robot. Autom.*, Apr. 2007, pp. 3856–3861.
[17] B. S. Rao and H. F. Durrant-Whyte, "Fully decentralised algorithm for multisensor Kalman filtering," in *Proc. D IEEE Contr. Theory and Applicat. [see also Proc. IEEE Contr. Theory Applicat.]*, Sep. 1991, vol. 138, pp. 413–420.
[18] R. Olfati-Saber, "Distributed Kalman filter with embedded consensus filters," in *Proc. IEEE Conf. Dec. Contr., Eur. Contr. Conf.*, Dec. 2005, pp. 8179–8184.
[19] R. Olfati-Saber, "Distributed Kalman filtering for sensor networks," in *Proc. IEEE Conf. Dec. Contr.*, Dec. 2007, pp. 5492–5498.
[20] L. Ong, B. Upcroft, M. Ridley, T. Bailey, S. Sukkarieh, and H. D.-. Whyte, "Consistent methods for decentralised data fusion using particle filters," in *Proc. IEEE Int. Conf. Multisens. Fusion and Integr. Intell. Syst.*, Sep. 2006, pp. 85–91.
[21] M. Coates, "Distributed particle filters for sensor networks," in *Proc. 3rd Wkshp Inf. Process. Sens. Netw. (IPSN)*, 2004, pp. 99–107.
[22] X. Sheng, Y. Hu, and P. Ramanathan, "Distributed particle filter with GMM approximation for multiple targets localization and tracking in wireless sensor network," in *Proc. 4th Int. Symp. Inf. Process. Sens. Netw.*, Piscataway, NJ, 2005, p. 24.
[23] D. Gu, J. Sun, Z. Hu, and H. Li, "Consensus based distributed particle filter in sensor networks," in *Proc. Int. Conf. Inf. Autom.*, Jun. 2008, pp. 302–307.
[24] S. H. Lee and M. West, "Markov chain distributed particle filters (MCDPF)," in *Proc. 48th IEEE Conf. Dec. Contr.*, 2009, pp. 5496–5501.
[25] N. D. Freitas, "Rao-Blackwellised particle filtering for fault diagnosis," in *Proc. IEEE Aerosp.*, 2001, pp. 1767–1772.
[26] J.-H. Kim, M. West, E. Scholte, and S. Narayanan, "Multiscale consensus for decentralized estimation and its application to building systems," in *Proc. Amer. Contr. Conf.*, Jun. 2008, pp. 888–893.
[27] S. H. Lee and M. West, "Performance comparison of the distributed extended Kalman filter and Markov Chain distributed particle filter (MCDPF)," in *Proc. 2nd IFAC Wkshp Distrib. Estimat. Contr. Netw. Syst. (NecSys'10)*, 2010, pp. 151–156.
[28] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag, 2001.
[29] D. Crisan, P. D. Moral, and T. Lyons, "Discrete filtering using branching and interacting particle systems," *Markov Proc. Rel. Fields*, vol. 5, pp. 293–318, 1998.
[30] F. Le Gland and N. Oudjane, "Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters," in *Proc. Ann. Appl. Probabil.*, 2004, vol. 14, pp. 144–187.
[31] P. Del Moral, *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. New York: Springer, 2004.
[32] A. Bain and D. Crisan, *Fundamentals of Stochastic Filtering*. New York: Springer, 2008.
[33] X.-L. Hu, T. B. Schön, and L. Ljung, "A basic convergence result for particle filtering," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1337–1348, 2008.
[34] D. Crisan and A. Doucet, "A survey of convergence results on particle filtering methods for practitioners," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 736–746, Mar. 2002.
[35] V. Romanovsky, *Discrete Markov Chains*. New York: Wolters-Noordhoff, 1970.
[36] D. Crisan, "Particle filters — A theoretical perspective," in *Sequential Monte Carlo Methods in Practice*, N. Doucet, A. de Freitas, and N. Gordon, Eds. New York: Springer, 2001, ch. 2, pp. 17–42.
[37] R. Douc, O. Cappé, and E. Moulines, "Comparison of resampling schemes for particle filtering," in *Proc. 4th Int. Symp. Image Signal Process. Anal.*, 2005, pp. 64–69.
[38] X.-L. Hu, T. B. Schon, and L. Ljung, "A basic convergence result for particle filtering," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1337–1348, 2008.
[39] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*. New York: Springer–Verlag, 1993 [Online]. Available: http://probability.ca/MT/
[40] D. Steinsaltz, "Convergence of moments in a Markov-chain central limit theorem," *Indagationes Math.*, vol. 12, pp. 533–555, 2001.
[41] A. Lasota and M. C. Mackey, *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*. New York: Springer-Verlag, 1994.
[42] H. Tanner, A. Jadbabaie, and G. J. Pappas, "Coordination of multiple autonomous vehicles," in *Proc. IEEE Medit. Conf. Contr. Autom.*, Jun. 2003, pp. 3448–3453.
[43] R. Olfati-Saber and J. S. Shamma, "Consensus filters for sensor networks and distributed sensor fusion," in *Proc. IEEE Conf. Dec. Contr., Eur. Contr. Conf.*, Dec. 2005, pp. 6698–6703.
[44] J.-H. Kim, M. West, E. Scholte, and S. Narayanan, "Multiscale consensus for decentralized estimation and its application to building systems," in *Proc. Amer. Contr. Conf.*, 2008, pp. 888–893.
[45] C. Selle and M. West, "Multiscale networks for distributed consensus algorithms," in *Proc. 48th IEEE Conf. Dec. Contr.*, 2009, pp. 4753–4758.
[46] T. Bengtsson, P. Bickel, and B. Li, "Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems," *Probabil. Statist.: Essays in Honor of David A. Freedman*, vol. 2, pp. 316–334, 2008.
[47] F. Daum, R. Co, and J. Huang, "Curse of dimensionality and particle filters," in *Proc. IEEE Aerosp. Conf.*, 2003, vol. 4, pp. 1979–1993.
[48] C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson, "Obstacles to high-dimensional particle filtering," *Month. Weather Rev.*, vol. 136, no. 12, pp. 4629–4640, 2008.
[49] P. Del Moral and A. Guionnet, "On the stability of interacting processes with applications to filtering and genetic algorithms," in *Proc. Ann. l'Inst. Henri Poincare (B) Probabil. Statist.*, 2001, vol. 37, pp. 155–194.
[50] F. Le Gland and N. Oudjane, "Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters," in *Proc. Ann. Appl. Probabil.*, 2004, pp. 144–187.

Sun Hwan Lee received the B.S. degree in aerospace and mechanical engineering from Seoul National University, Seoul, Korea, in 2004, and the M.S. and Ph.D. degrees in aeronautics and astronautics from Stanford University, Stanford, CA, in 2006 and 2010, respectively.

He joined Lattice Engines Inc., San Mateo, CA, in 2010 and currently holds the position of Senior Analytics Engineer. His research interests are in statistical signal processing, machine learning, predictive analysis, and the application of big data to practical fields.

Dr. Lee was awarded a Samsung scholarship.

Matthew West received the B.Sc. degree in pure and applied mathematics from the University of Western Australia, in 1996, and the Ph.D. degree in control and dynamical systems from the California Institute of Technology, Pasadena, in 2004.

He was an Assistant Professor with the Department of Mathematics, University of California, Davis, from 2003 to 2004, and an Assistant Professor with the Department of Aeronautics and Astronautics, Stanford University, CA, from 2004 to 2007. He joined the Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, as an Assistant Professor in 2008, and is currently an Associate Professor. His research interests include distributed computation, time integration of stochastic systems, and particle methods for simulation and estimation.