Weak sense $\mathbb{L}_p$ error bounds for leader–node distributed particle filters

Boris N. Oreshkin and Mark Coates
Department of Electrical and Computer Engineering
McGill University
Montreal, Canada
Email: boris.oreshkin@mail.mcgill.ca, mark.coates@mcgill.ca

***Abstract*—The leader node particle filter is a partially distributed approach to tracking in a sensor network, in which the node performing the particle filter computations (the leader node) changes over time. The primary advantage is that the position of the leader node can follow the target, improving the efficiency of data collection. When the leader node changes, the particle filter must be communicated to the new leader. Exchanging a complete representation can require many thousands of bits, so the filtering distribution is more coarsely approximated, either by transmitting only a subset of particles or by training a parametric model. The concern is that this approximation could lead to instability and eventually tracking failure. The major contribution of this paper is the development of bounds on the error of the leader node particle filter. In contrast to previous bounds, which grow exponentially over time, the bounds indicate convergent behaviour.**

Keywords: Tracking, particle filtering, distributed estimation.

I. Introduction

One of the major concerns in distributed sensor network tracking is the maintenance of the appropriate tradeoff between tracking performance and network lifetime. If a centralized approach is used to process measurements from the sensors and in scenarios where Gaussian approximation is justifiable, one of the following well–established tracking algorithms can be used to obtain acceptable performance guarantees: the extended Kalman Filter [1], [2], the Gaussian sum filter [3] or grid–based filters [4]. However, if better performance guarantees are required in the situation where the class of approximated dynamics and/or observation models is substantially non–linear and non–Gaussian, different particle filter based trackers can be used [5]. In sensor network applications, there are two major disadvantages: particle filters are generally more computationally demanding [6], and communication of a particle filter representation can require the transfer of many more bits than parametric alternatives.

A particle filter maintains a set of "particles" that are simply candidate state values of the system (for example, the position and velocity of the object). The filter evaluates how well individual particles correspond to the dynamic model and set of observations, and assigns weights accordingly. The set of weighted particles provides a pointwise approximation to the filtering distribution, which represents the posterior probability of the state. This approximation allows one to form estimates of the state values and hence track the state.

Tracking algorithms using particle filters in sensor networks frequently adopt a centralized approach, wherein the particle filter resides at a computation centre and measurements are collected at this centre. This approach has several disadvantages. Centralization introduces a single point of failure and can lead to high, unevenly distributed energy consumption because of the heavy communication cost involved in transmitting the data to the fusion centre.

Distributed algorithms, such as the distributed particle filtering algorithms proposed in [7], [8], address the aforementioned problems. These algorithms decentralize the computation or communication so that a single fusion centre is not required. Multiple particle filters run concurrently at different sensor nodes and compressed data or approximate filtering distributions are shared between them. These distributed algorithms, while mitigating some of the inherent problems of centralization, can be computationally expensive, because multiple nodes are required to perform computation throughout the entire tracking procedure.

This paper considers the partially distributed scenario, where the node performing the particle filtering (the *leader* node) changes over time, as proposed in [9] and refined and analyzed in [10]. A hand–off of information to a new leader node is therefore required whenever the leader changes. This involves either transmitting the particle filter in the form of raw particle values and weights or training and communicating a parametric approximation. For convenience, we will refer to this approach as the *leader node particle filter*. It has the benefit of sharing the computational burden amongst the nodes in the network. Perhaps more importantly, the choice of leader node is governed by a desire for the leader node to be close to the object's position. The leader node can then poll nearby sensor nodes that are in close proximity to the object and hence are likely to generate more accurate measurements. The ability to collect measurements through local, one–hop communication can be critical in cases where the data volume is high.

In attempting to alleviate the communication cost of transmitting all particle values when the leader node is exchanged (which can involve thousands of bits), the filtering distribution

is often more coarsely approximated, either by transmitting only a subset of the particles or by training a parametric model.

A. Related Work and Contributions

One of the concerns is that this approximation exercise can induce errors in the filtering process that lead to instability and eventual loss of the object's path. Although simulation (and to some extent, experimental) results indicate that such instability effects are rarely observed, the currently available theoretical bounds on estimation error for leader node particle filtering grow exponentially over time [10]. Previous analyses have used maximum log–error to model the approximation error propagation in a leader node particle filter associated with an arbitrary dynamic system. Because log–error does not take into account the structure of the system, the resulting bounds diverge. However, several authors have investigated the stability of centralized particle filtering algorithms and obtained stronger results that use the properties of the underlying dynamical system [11], [12]. In particular, these results indicate that in the case of an exponentially stable or "sufficiently mixing" system, the particle filter error is bounded and the filter itself is stable. While these results cannot be used to guarantee the stability of the leader node particle filter directly, they do suggest an approach that uses information about the dynamical system to obtain stable bounds in the distributed framework. The major contribution of this paper is the provision of weak sense $\mathbb{L}_p$ bounds for leader node particle filtering. We consider two settings. First, we analyze a leader node particle filter that uses non–parametric bootstrap approximation to the posterior distribution when changing leader nodes. Second, we examine an alternative filter that employs parametric posterior distribution approximation based on the greedy likelihood maximization algorithm developed in [13], [14]. Unlike [10], the bounds that we obtain indicate that under reasonable assumptions one should expect a convergent behaviour from a leader node particle filter in both settings.

We note that these results are relevant to any situation where a particle filtering representation of a distribution (or likelihood function) is approximated, either by subsampling or by a parametric approach. Such situations arise in non–parametric belief propagation [15], [16] and in the distributed particle filter [7], [8].

B. Paper Organization

The rest of the paper is organized as follows. The next section sets up the analysis framework and outlines relevant terminology. Section III provides the detailed description of the algorithms that we analyze and states the main results. Section IV outlines the proofs of these results, and Section V discusses the implications. Finally, Section VI concludes the paper.

II. Analysis Framework and Terminology

We adopt the following general signal model:

$$\mathbf{x}_t = f_t(\mathbf{x}_{t-1}, \varrho_t) \tag{1}$$

$$\mathbf{y}_t^j = g_t^j(\mathbf{x}_t, \zeta_t^j), \quad j = 1 \dots M \tag{2}$$

where $\mathbf{x}_t$ is the $d_x \times 1$ target state vector at time t, $\mathbf{y}_t^j$ is the $d_y^j \times 1$ measurement vector at sensor node j, ϱ_t and ζ_t^j are system excitation and measurement noises correspondingly, f_t is a nonlinear system map $f_t : \mathbb{R}^{d_x} \to \mathbb{R}^{d_x}$, and g_t^j is a nonlinear measurement map $g_t^j : \mathbb{R}^{d_x} \to \mathbb{R}^{d_y^j}$. Denote by y_t the complete vector of observations $\{y_t^j; j = 1, \dots, M\}$.

In order to conduct stability analysis, we need to introduce slightly more rigorous mathematical notations. Let $(E_t, \mathcal{E}_t)$, $t \in \mathbb{N}$ be a sequence of measurable spaces. The target state vector evolves according to a non–homogeneous (discrete–time) Markov chain X_t with transitions M_{t+1} from E_t into E_{t+1}. We denote by $X_t' = X_{[0:t]}$ the historical path process associated with X_t, and use M_t' to denote the Markov transitions of the path process. Note that in the proposed signal model, we constrain E_t to be $\mathbb{R}^{d_x \times 1}$, but in general it can be any measurable space.

Associated with a measurable space of the form $(E, \mathcal{E})$ is a set of probability measures $\mathcal{P}(E)$ and the Banach space of bounded functions $\mathcal{B}_b(E)$ with supremum norm:

$$||h|| = \sup_{x \in E} |h(x)|. \tag{3}$$

We define a convex set $\mathrm{Osc}_1(E)$ of $\mathcal{E}$–measurable test functions with finite oscillations:

$$\mathrm{osc}(h) = \sup(|h(x) - h(y)|;\ x, y \in E) \tag{4}$$

$$\mathrm{Osc}_1(E) = \{h : \mathrm{osc}(h) \leq 1\} \tag{5}$$

In order to simplify the representation of integral operators, we define for a measure $\mu \in \mathcal{P}(E)$, $\mu(h) \triangleq \int_E h(x)\mu(dx)$.

Through the rest of this paper we adopt the methodology developed in [11] to analyze the behaviour of filtering distributions arising from (1) and (2). This methodology involves representing the particle filter as an N–particle approximation of a Feynman–Kac model. In the remainder of this section, we describe how this representation is performed; for a much more detailed description and discussion, please refer to [11].

The evolution of the unconditional signal distribution in (1) is completely defined by the Markov transition kernel $M(\cdot, \cdot)$ and the initial signal distribution μ_0:

$$\Pr\{X_t \in d\mathbf{x}_t | X_{t-1} = \mathbf{x}_{t-1}\} = M_t(\mathbf{x}_{t-1}, d\mathbf{x}_t) \tag{6}$$

According to (6), the signal distribution at time t, with respect to the sequence of states $x_1, \dots, x_t$, can be written as follows

$$\mathbb{P}_{\mu,t}(d(\mathbf{x}_0, \dots, \mathbf{x}_t)) = \mu(d\mathbf{x}_0)M_1(\mathbf{x}_0, d\mathbf{x}_1) \dots M_t(\mathbf{x}_{t-1}, d\mathbf{x}_t) \tag{7}$$

We now introduce bounded and non–negative potential functions G_t on E_t, and set $G_t(x_t) = p_{Y_t|X_t}(y_t|x_t)$ so that G_t reflects the probabilistic relationship between state and signal in the measurement process (2). This leads to the following definition of the unnormalized *prediction* Feynman–Kac model, for $h_t \in \mathcal{B}_b(E_t)$ and $t \in \mathbb{N}$.

$$\gamma(h_t) \triangleq \mathbb{E}_{\eta_0}\left(h_t(X_t)\prod_{i=0}^{t-1} G_i(X_i)\right) \tag{8}$$

where $\mathbb{E}_{\eta_0}$ denotes expectation with respect to the distribution of an E_t–valued Markov chain X_t with transitions M_t. The normalized prediction Feynman–Kac model is then:

$$\eta_t(h_t) = \frac{\gamma(h_t)}{\gamma(1)} \tag{9}$$

Note that η_t is closely related to the predictive posterior probability density function in the Bayesian methodology.

We introduce the Boltzmann–Gibbs transformation Ψ_t to reflect the effect of the likelihood function at time t on the normalized prediction model. The transformation Ψ_t maps the set of probability measures on E_t onto itself, i.e. $\Psi_t : \nu \in \mathcal{P}_t(E_t) \mapsto \Psi_t(\nu) \in \mathcal{P}_t(E_t)$. For a particular measure ν,

$$\Psi_t(\nu)(d\mathbf{x}_t) = \frac{1}{\nu(G_t)} G_t(\mathbf{x}_t)\nu(d\mathbf{x}_t). \tag{10}$$

This transformation is used to construct the key operator $\Phi_t : \mathcal{P}(E_{t-1}) \rightarrow \mathcal{P}(E_t)$, which is used to update the predictive posterior distribution from time step $t-1$ to time step t. This operator uses the fitness assessment described by the likelihood function G_{t-1} and the diffusion step described by the Markov kernel M_t.

$$\Phi_t(\eta) = \Psi_{t-1}(\eta)M_t \tag{11}$$

$$\eta_t = \Phi_t(\eta_{t-1}) \tag{12}$$

Let $Q_{i,t}$ and $\Phi_{i,t}$, $i \leq t$ be the semigroups associated respectively with the Feynman–Kac distribution flows γ_t and η_t.

$$\Phi_{i,t} = \Phi_t \circ \Phi_{t-1} \circ \ldots \circ \Phi_{i+1} \tag{13}$$

$$Q_{i,t} = Q_t \circ Q_{t-1} \circ \ldots \circ Q_{i+1} \tag{14}$$

Here $Q_t(x_{t-1}, dx_t) = G_{t-1}(x_{t-1})M_t(x_{t-1}, dx_t)$ and $\Phi_{i,t}$ is a nonlinear integral operator from $\mathcal{P}(E_i)$ to $\mathcal{P}(E_t)$, defined for $\mu_i \in \mathcal{P}(E_i)$ and $h_t \in \mathcal{B}_b(E_t)$ as:

$$\Phi_{i,t}(\mu_i)(h_t) = \mu_i(G_{i,t}P_{i,t}(h_t))/\mu_i(G_{i,t})$$

where $G_{i,t} = Q_{i,t}(1)$ and $P_{i,t}(h_t) = Q_{i,t}(h_t)/Q_{i,t}(1)$.

The semigroup $\Phi_{i,t}$ describes the evolution of the normalized prediction Feynman–Kac model η_t. We can define an associated particle filter by developing an N–particle approximation to this model. This consists of N path particles:

$$\xi_t'^k = (\xi_{i,t}^k)_{0\leq i\leq t} \in E_t' = E_{[0,t]} \qquad i \in 1, \ldots, N$$

The particle approximation of the prediction Feynman–Kac model is defined as:

$$\eta_n^N = \frac{1}{N}\sum_{k=1}^{N} \delta_{\xi_{t,t}^k}$$

The N–tuple ξ_t represents the configuration at time t of N particles ξ_t^k, and resides in the product space E_t^N. The particle filter then involves a two–step updating process:

$$\xi_t \in E_t^N \overset{\text{selection}}{\longrightarrow} \widehat{\xi}_t \in E_t^N \overset{\text{mutation}}{\longrightarrow} \xi_{t+1} \in E_{t+1}^N$$

The selection stage consists of selecting randomly N particles $\widehat{\xi}_t^k$. This random selection is achieved by setting, with probability $\epsilon_t G_t(\xi_t^k)$, $\widehat{\xi}_t^k = \xi_t^k$; otherwise we choose a random particle $\widetilde{\xi}_t^k$ with distribution $\sum_{k=1}^N \frac{G_t(\xi_t^k)}{\sum_{j=1}^N G_t(\xi_t^j)}\delta_{\xi_t^k}$, and we set $\widehat{\xi}_t^k = \widetilde{\xi}_t^k$. During the mutation phase, each particle $\widehat{\xi}_t^k$ evolves according to the Markov transition M_{t+1}.

III. Leader Node Particle Filter Error Bounds

This section presents our main results. Initially we define two versions of the leader node particle filter, and subsequently we state weak sense $\mathbb{L}_p$ error bounds for these two filters. In both filters, we assume that there is a change of leader node at time t with probability q, and communication of the particle filter occurs at this time. The value of q is related to the rate of leader node change. Thus the approximation error assessed using this scheme is clearly related to the allowed frequency of the leader node exchange. We consider this simplified communication scheme because our interest is in analyzing the upper–bounds for the propagation of approximation error. The total tracking error can be decomposed into the target state estimation error, which is due to the measurement errors and randomness of the dynamic system and approximation error, which is due to the particle or parametric approximation performed at every step. The bounds on the expected error of the second kind do not depend on particular sequences of leader nodes allocated. For the analysis of the first kind of error please refer to [17], which addresses the sensor selection problem.

A. Particle Filter Descriptions

PF1: Bootstrap Approximation Leader Node Particle Filter:

$$\text{with probability } q, \quad \Phi_t(\eta_{t-1}^{N_1}) \Rightarrow \eta_t^{N_2} \longrightarrow \eta_t^{N_2} \Rightarrow \eta_t^{N_1} \tag{15}$$

$$\text{with probability } 1-q, \quad \Phi_t(\eta_{t-1}^{N_1}) \Rightarrow \eta_t^{N_1} \tag{16}$$

Here the implication sign $\Rightarrow$ represents a sampling operation and the right arrow $\longrightarrow$ denotes the communication process. Note that $\eta_t^{N_1}$ in (15) can be regarded as a bootstrap approximation of size N_1 to an empirical distribution $\eta_t^{N_2}$. Let us introduce the following sampling operator $S^N : \mathcal{P}(E) \rightarrow \mathcal{P}(E^N)$ to facilitate the analysis of (15)–(16):

$$S^N(\eta) = \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi^i} \ . \tag{17}$$

Here $(\xi^1, \ldots, \xi^N)$ is the i.i.d sample from η. Using this operator we can rewrite (15) in a more concise form:

$$\text{with probability } q, \quad \eta_t^{N_1} = S^{N_1} \circ S^{N_2} \circ \Phi_t(\eta_{t-1}^{N_1}) \tag{18}$$

The second particle filter we define relies upon a parametric approximation of the distribution. Although the parametric density approximation can be accomplished in many ways, in order to develop error bounds, we require that the algorithm has approximation and estimation errors that are controlled in some well–defined sense. For this reason, we consider the

following family of bounded parametric densities:

$$\mathcal{C} = \{\phi_\theta(x) : \theta \in \Theta \subset \mathbb{R}^d, a \leq \phi_\theta \leq b\} \tag{19}$$

where $0 < a < b < \infty$ and Θ defines parameter space. We are looking for an approximation g_k to a true density $\frac{d\eta_t}{dx}$. The approximation is restricted to a class of discrete k–component convex combinations of the form:

$$\mathcal{G}_k = \mathrm{conv}_k(\mathcal{C}) = \left\{ g^k : g^k(x) = \sum_{i=1}^{k} \alpha_i \phi_{\theta_i}(x), \phi_\theta \in \mathcal{C}, \sum_{i=1}^{k} \alpha_i = 1,\ \alpha_i \geq 0,\ \theta_i \in \Theta \right\} \tag{20}$$

To analyze the potential performance of the algorithm based on parametric particle communication we consider the following greedy approximation procedure that is based on the minimization of some loss function $\Upsilon : \mathcal{B}_b(E_t) \to \mathbb{R}$ [14], [18].

Algorithm 1: Sequential greedy approximation

Given $g_{\theta_1} \in \mathcal{C}$
for $i = 2$ *to* k **do**
  Find $\phi_{\theta_i} \in \mathcal{C}$ and $0 \leq \alpha_i \leq 1$ to minimize the function:
  $(\phi^*_{\theta_i}, \alpha^*_i) = \arg\min_{\alpha_i, \phi_{\theta_i}} \Upsilon((1-\alpha_i) g_{\theta_{i-1}} + \alpha_i \phi_{\theta_i})$
  Let $g_i = (1-\alpha^*_i) g_{\theta_{i-1}} + \alpha^*_i \phi^*_{\theta_i}$
endfor

Algorithm 1 is clearly suboptimal. However, the results on greedy approximation in non–Hilbert spaces [19] indicate that approximation rates achieved by this procedure are close or equal to those of the optimum approach. Thus the analysis of the error bounds of a leader node particle filter utilizing parametric greedy approximation provides a good approximation to error bounds that can be achieved by an equivalent algorithm that uses an optimal distribution approximation algorithm. In this sense, we hope that this analysis provides achievable worst–case performance upper–bounds for any leader node particle filter using a "sufficiently good" parametric distribution approximation algorithm.

We note that Algorithm 1 can be used to calculate an estimate g of an unknown density w based on the minimization of Kullback–Leibler (KL) divergence, which can be defined for two measures ν and μ:

$$D(\nu||\mu) = \mathbb{E}_\nu \log\left(\frac{d\nu}{d\mu}\right), \tag{21}$$

if the loss function of the form $\Upsilon(g) = -\mathbb{E}_w \log(g)$ is used [14]:

$$\begin{aligned} g &= \arg\min_g D(w||g) = \arg\min_g \mathbb{E}_w \log(w) - \mathbb{E}_w \log(g) \\ &= \arg\min_g -\mathbb{E}_w \log(g). \end{aligned} \tag{22}$$

In practice w is unknown, but its approximation by an empirical equivalent leads to the Maximum Likelihood density estimation algorithm. Accordingly, there are several papers [13], [14] that develop bounds on approximation and estimation error of Algorithm 1 in terms of KL–divergence.

The minimization of KL–divergence does not directly lead to the minimization of $\mathbb{L}_p$ error, but it does correspond to the minimization of a bound on $\mathbb{L}_1$ error [10].

PF2: Parametric Approximation Leader Node Particle Filter:

$$\text{with probability } q, \quad \Phi_t(\eta^N_{t-1}) \Rrightarrow \widehat{\eta}^k_t \longrightarrow \widehat{\eta}^k_t \Rightarrow \eta^N_t \tag{23}$$

$$\text{with probability } 1-q, \quad \Phi_t(\eta^N_{t-1}) \Rightarrow \eta^N_t \tag{24}$$

Here the $\Rrightarrow$ represents the local distribution approximation process. Thus instead of a bootstrap approximation, we transmit parameters of the distribution estimated from the particle approximation available at the current time step. The particle filter *PF2* employs the greedy density approximation algorithm of [18], detailed above in Algorithm 1.

B. Main Results

Our results are stated in terms of the quantities $r_{i,t}$, which measures the relative oscillations of the potential functions $G_{i,t}$, and $\beta_{i,t}$, which measures the contraction properties of the Markov transition $P_{i,t}$. These quantities are defined as:

$$r_{i,t} = \sup_{x_i, y_i \in E_i} (G_{i,t}(x_i)/G_{i,t}(y_i)) \tag{25}$$

$$\beta(P_{i,t}) = \sup_{x_i, y_i \in E_i} ||P_{i,t}(x_i, \cdot) - P_{i,t}(y_i, \cdot)||_{\mathrm{tv}} \tag{26}$$

Theorem 1 specifies a weak sense $\mathbb{L}_p$ bound on the error incurred by the bootstrap approximation particle filter, *PF1*.

Theorem 1. *For any $t \geq 0$, $p \geq 1$, and $h_t \in \mathrm{Osc}_1(E_t)$ the $\mathbb{L}_p$ error of the distributed particle filter algorithm (15)–(16) is upper bounded as follows, for a finite constant $d(p)$:*

$$\mathbb{E}\left\{ \left| \left[\eta^{N_1}_t - \eta_t\right](h_t) \right|^p \right\}^{\frac{1}{p}} \leq d(p)^{\frac{1}{p}} \left(\frac{1}{\sqrt{N_1}} + \frac{q}{\sqrt{N_2}} \right) \sum_{i=0}^{t} r_{i,t} \beta(P_{i,t}).$$

Theorem 2 specifies a weak sense $\mathbb{L}_p$ bound on the error incurred by the parametric approximation particle filter, *PF2*.

Theorem 2. *Let η^N_i be the N–particle approximation of η_i at each particle filtering step and $\widehat{\eta}^k_i \in \mathcal{G}_k$ be the k–component mixture approximation constructed by the greedy likelihood approximation algorithm of [18] applied to $\Phi_i(\eta^N_{i-1})$. Then for any $t \geq 0$, $p \geq 1$, and $h_t \in \mathrm{Osc}_1(E_t)$, $||h_t|| \leq 1$, the $\mathbb{L}_p$ error of the distributed particle filter algorithm (23)–(24) is upper bounded by the following:*

$$\begin{aligned} \mathbb{E}\left\{ \left| \left[\eta^N_t - \eta_t\right](h_t) \right|^p \right\}^{\frac{1}{p}} &\leq d(p)^{\frac{1}{p}} \frac{1}{\sqrt{N}} \sum_{i=0}^{t} r_{i,t} \beta(P_{i,t}) \\ &+ \sqrt{2} q \sum_{i=0}^{t} r_{i,t} \beta(P_{i,t}) \left(\frac{c}{k} + \mathbb{E}\left\{ D(\eta^N_i || \mathcal{G})^p \right\}^{\frac{1}{p}} \right)^{1/2}. \end{aligned}$$

where c is a constant that depends only on the class of

approximating densities, and

$$D(\eta_i^N||\mathcal{G}) = \inf_{\bar{\eta}\in\mathcal{G}} D(\eta_i^N||\bar{\eta}) \tag{27}$$

IV. Stability Analysis and Proofs

To analyze the convergence behaviour of the leader node particle filter we use the following decomposition of sampling error into the sum of local sampling errors [11]:

$$\eta_t^N - \eta_t = \sum_{i=0}^{t} \left[\Phi_{i,t}(\eta_i^N) - \Phi_{i,t}(\Phi_i(\eta_{i-1}^N))\right]. \tag{28}$$

Here η_t^N is the N particle approximation of η_t at time t. Using the triangle inequality and (28) the expected $\mathbb{L}_p$ error of the particle filter at time t can be upper bounded in the following manner:

$$\mathbb{E}\left\{\left|\left[\eta_t^N - \eta_t\right](h_t)\right|^p\right\}^{\frac{1}{p}} \leq \sum_{i=0}^{t} \mathbb{E}\left\{\left|\left[\Phi_{i,t}(\eta_i^N) - \Phi_{i,t}(\Phi_i(\eta_{i-1}^N))\right](h_t)\right|^p\right\}^{\frac{1}{p}}. \tag{29}$$

A. Bootstrap Approximation Leader Node Particle Filter

We begin by observing that Φ_t always operates on the N_1 particle approximation in (15)–(16). Following the lines of analysis in [11], we introduce the N–particle equivalents of $Q_{i,t}$, $P_{i,t}$, and $G_{i,t}$. The random potential function $G_{i,t}^{N_1} : E_i \to (0,\infty)$ is defined as:

$$G_{i,t}^{N_1}(\mathbf{x}_i) = \frac{G_{i,t}}{\Phi_i(\eta_{i-1}^{N_1})(G_{i,t})}(\mathbf{x}_i). \tag{30}$$

The random Markov kernel $P_{i,t}^{N_1} : E_i \to E_t$ is defined for any $(h_t, \mathbf{x}_i) \in (\mathfrak{B}_b(E_t) \times E_i)$ as:

$$P_{i,t}^{N_1}(h_t)(\mathbf{x}_i) = \int (P_{i,t}h_t(\mathbf{x}_i) - P_{i,t}h_t(\mathbf{y}_i))G_{i,t}^{N_1}(\mathbf{y}_i)\Phi_i(\eta_{i-1}^{N_1})(d\mathbf{y}_i), \tag{31}$$

and

$$Q_{i,t}^{N_1}(h_t)(\mathbf{x}_i) = G_{i,t}^{N_1}(\mathbf{x}_i) \times P_{i,t}^{N_1}(h_t)(\mathbf{x}_i). \tag{32}$$

Each summand in (29) can be rewritten in terms of $Q_{i,t}^{N_1}$:

$$\Phi_{i,t}(\eta_i^{N_1}) - \Phi_{i,t}(\Phi_i(\eta_{i-1}^{N_1})) = \frac{1}{\eta_i^{N_1}(G_{i,t}^{N_1})}\left[\eta_i^{N_1} - \Phi_i(\eta_{i-1}^{N_1})\right] Q_{i,t}^{N_1} \tag{33}$$

We now commence the proof of Theorem 1 concerning the $\mathbb{L}_p$ bounds on the error of *PF1*. We will need the following result proven by Del Moral [11] (part of Lemma 7.3.3) that is cited here without a proof.

Lemma 1. *(Del Moral [11], Lemma 7.3.3)*

For any $p \geq 1$ and sequence of $\mathcal{E}$–measurable functions $(h_i)_{i\geq 1}$ with finite oscillations such that $\mu_i(h_i) = 0$ for all $i \geq 1$ we have

$$\sqrt{N}\mathbb{E}\{|m(X)(h)^p|\}^{\frac{1}{p}} \leq d(p)^{\frac{1}{p}}\sigma(h) \tag{34}$$

where the following definitions are used

$$m(x)(h) = \frac{1}{N}\sum_{i=1}^{N} h_i(x^i) \quad \textit{and} \quad \sigma^2(h) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{osc}^2(h_i) \tag{35}$$

and finite constants $d(p)$ are given by the following:

$$d(2p) = \frac{(2p)!}{p!}2^{-p}, \tag{36}$$

$$d(2p-1) = \frac{(2p-1)!}{(p-1)!\sqrt{p-1/2}}2^{-(p-1/2)} \tag{37}$$

Before deriving Theorem 1 we investigate an important property of bootstrap approximation.

Lemma 2. *For any $p \geq 1$, and $h \in \mathrm{Osc}_1(E)$, the unconditional expected $\mathbb{L}_p$ error of a bootstrap approximation $\widehat{\mathbb{P}}_n$ of size n sampled from an empirical distribution $\mathbb{P}_N$ of size N is upper bounded by the following:*

$$\mathbb{E}\left\{\left|\left[\widehat{\mathbb{P}}_n - P\right](h)\right|^p\right\}^{\frac{1}{p}} \leq d(p)^{\frac{1}{p}}\left(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{n}}\right) \tag{38}$$

Proof: Let us introduce $\mathbf{X} = [X_1, \ldots, X_N]^T$, a sample from probability distribution P, and $\mathbf{X}^* = [X_1^*, \ldots, X_n^*]^T$, a bootstrap sample of size n from the empirical distribution $\mathbb{P}_N$. By $\mathbb{E}_{\mathbf{X}}$ and $\mathbb{E}_{\mathbf{X}^*}$ we denote the expectations with respect to distributions P and $\mathbb{P}_N$ respectively. Using the triangle inequality an obvious result follows:

$$\mathbb{E}\left\{\left|\left[\widehat{\mathbb{P}}_n - P\right](h)\right|^p\right\}^{\frac{1}{p}} \leq \mathbb{E}_{\mathbf{X},\mathbf{X}^*}\left\{\left|\left[\widehat{\mathbb{P}}_n - \mathbb{P}_N\right](h)\right|^p\right\}^{\frac{1}{p}} + \mathbb{E}_{\mathbf{X}}\left\{\left|\left[\mathbb{P}_N - P\right](h)\right|^p\right\}^{\frac{1}{p}} \tag{39}$$

Application of Lemma 1 to the second summand gives:

$$\mathbb{E}_{\mathbf{X}}\left\{\left|\left[\mathbb{P}_N - P\right](h)\right|^p\right\}^{\frac{1}{p}} \leq d(p)^{\frac{1}{p}}\frac{1}{\sqrt{N}}\sigma^2(h) \tag{40}$$

$$\leq d(p)^{\frac{1}{p}}\frac{1}{\sqrt{N}} \tag{41}$$

On the other hand, the first summand can be written as follows:

$$\mathbb{E}_{\mathbf{X},\mathbf{X}^*}\left\{\left|\left[\widehat{\mathbb{P}}_n - \mathbb{P}_N\right](h)\right|^p\right\}^{\frac{1}{p}} = \mathbb{E}_{\mathbf{X}}\left\{\mathbb{E}_{\mathbf{X}^*}\left\{\left|\left[\widehat{\mathbb{P}}_n - \mathbb{P}_N\right](h)\right|^p \middle| \mathbf{X}\right\}\right\}^{\frac{1}{p}} \tag{42}$$

It follows that for almost all sequences $\mathbf{X}$ [20] Lemma 1 holds and we have the following result:

$$\mathbb{E}_{\mathbf{X}}\left\{\mathbb{E}_{\mathbf{X}^*}\left\{\left|\left[\widehat{\mathbb{P}}_n - \mathbb{P}_N\right](h)\right|^p \middle| \mathbf{X}\right\}\right\}^{\frac{1}{p}} \leq \mathbb{E}_{\mathbf{X}}\left\{d(p)\left(\frac{1}{\sqrt{n}}\sigma^2(h)\right)^p\right\}^{\frac{1}{p}} \leq \mathbb{E}_{\mathbf{X}}\left\{d(p)\left(\frac{1}{\sqrt{n}}\right)^p\right\}^{\frac{1}{p}} = d(p)^{\frac{1}{p}}\frac{1}{\sqrt{n}} \tag{43}$$

Combining (39), (40) and (43) we obtain the desired result ■

Proof of Theorem 1:

Taking into account (33) and using the Hölder inequality, each summand in (29) can be upper bounded as follows:

$$\mathbb{E}\left\{\left|\left[\Phi_{i,t}(\eta_i^{N_1}) - \Phi_{i,t}(\Phi_i(\eta_{i-1}^{N_1}))\right](h_t)\right|^p\right\}^{\frac{1}{p}} \leq \left|\left|\frac{Q_{i,t}^{N_1}(h_t)}{\eta_i^{N_1}(G_{i,t}^{N_1})}\right|\right| \mathbb{E}\left\{\left|\left[\eta_i^{N_1} - \Phi_i(\eta_{i-1}^{N_1})\right](h_t)\right|^p\right\}^{\frac{1}{p}} \quad (44)$$

The second factor in the above formula can be analyzed in the following way. Given the independence of the events (15) and (16) from the quantities under the expectation in (44) we can conclude the following:

$$\mathbb{E}\left\{\left|\left[\eta_i^{N_1} - \Phi_i(\eta_{i-1}^{N_1})\right](h_t)\right|^p\right\}^{\frac{1}{p}} = \mathbb{E}\left\{q\left|\left[S^{N_1}\circ S^{N_2}\circ\Phi_i(\eta_{i-1}^{N_1}) - \Phi_i(\eta_{i-1}^{N_1})\right](h_t)\right|^p + (1-q)\left|\left[S^{N_1}\circ\Phi_i(\eta_{i-1}^{N_1}) - \Phi_i(\eta_{i-1}^{N_1})\right](h_t)\right|^p\right\}^{\frac{1}{p}} \quad (45)$$

Using Jensen's inequality yields:

$$\mathbb{E}\left\{\left|\left[\eta_i^{N_1} - \Phi_i(\eta_{i-1}^{N_1})\right](h_t)\right|^p\right\}^{\frac{1}{p}} \leq q\mathbb{E}\left\{\left|\left[S^{N_1}\circ S^{N_2}\circ\Phi_i(\eta_{i-1}^{N_1}) - \Phi_i(\eta_{i-1}^{N_1})\right](h_t)\right|^p\right\}^{\frac{1}{p}} + (1-q)\mathbb{E}\left\{\left|\left[S^{N_1}\circ\Phi_i(\eta_{i-1}^{N_1}) - \Phi_i(\eta_{i-1}^{N_1})\right](h_t)\right|^p\right\}^{\frac{1}{p}} \quad (46)$$

Using Lemma 1 we can see that the second summand in (46) is upper bounded by the following:

$$\mathbb{E}\left\{\left|\left[S^{N_1}\circ\Phi_i(\eta_{i-1}^{N_1}) - \Phi_i(\eta_{i-1}^{N_1})\right](h_t)\right|^p\right\}^{\frac{1}{p}} \leq d(p)^{\frac{1}{p}}\frac{1}{\sqrt{N_1}} \quad (47)$$

The application of Lemma 2 to the first summand in (46) gives:

$$\mathbb{E}\left\{\left|\left[S^{N_1}\circ S^{N_2}\circ\Phi_i(\eta_{i-1}^{N_1}) - \Phi_i(\eta_{i-1}^{N_1})\right](h_t)\right|^p\right\}^{\frac{1}{p}} \leq d(p)^{\frac{1}{p}}\left(\frac{1}{\sqrt{N_1}} + \frac{1}{\sqrt{N_2}}\right) \quad (48)$$

Combining (46)–(48) results in

$$\mathbb{E}\left\{\left|\left[\eta_i^{N_1} - \Phi_i(\eta_{i-1}^{N_1})\right](h_t)\right|^p\right\}^{\frac{1}{p}} \leq d(p)^{\frac{1}{p}}\left(\frac{1}{\sqrt{N_1}} + \frac{1}{\sqrt{N_2}}q\right) \quad (49)$$

Using the properties of the Dobrushin contraction coefficient [11]:

$$\left|\left|\frac{Q_{i,t}^{N_1}(h_t)}{\eta_i^{N_1}(G_{i,t}^{N_1})}\right|\right| \leq r_{i,t}\beta(P_{i,t}) \quad (50)$$

and combining (44), (49), and (29) gives the desired result. ■

The following remark discusses a special case in which the upper bound presented in Theorem 1 becomes slightly tighter.

Remark 1. *If the number of particles N_1 used in Monte–Carlo filter calculations is a multiple of the number of particles N_2 used for communication between leader nodes, i.e., $N_1 = kN_2$ for some $k \in \mathbb{N}$, the upper bound of Theorem 1 becomes:*

$$\mathbb{E}\left\{\left|\left[\eta_t^{N_1} - \eta_t\right](h_t)\right|^p\right\}^{\frac{1}{p}} \leq d(p)^{\frac{1}{p}}\left(\frac{1-q}{\sqrt{N_1}} + \frac{q}{\sqrt{N_2}}\right)\sum_{i=0}^{t} r_{i,t}\beta(P_{i,t}), \quad (51)$$

Proof: It is clear that in the case where $N_1 = kN_2$, a deterministic sampling approach consisting of creating the N_1 particle approximation by just copying the N_2 particle approximation k times can be used in (15). Thus the sampling error of the S^{N_1} operator in (48) is zero. ■

B. Parametric Approximation Leader Node Particle Filter

Proof of Theorem 2:

We begin with an error decomposition similar to that used in Theorem 1:

$$\mathbb{E}\left\{\left|\left[\Phi_{i,t}(\eta_i^N) - \Phi_{i,t}(\Phi_i(\eta_{i-1}^N))\right](h_t)\right|^p\right\}^{\frac{1}{p}} \leq \left|\left|\frac{Q_{i,t}^N(h_t)}{\eta_i^N(G_{i,t}^N)}\right|\right| \times \left(q\mathbb{E}\left\{\left|\left[\Phi_i(\eta_{i-1}^N) - \widehat{\eta}_i^k\right](h_i)\right|^p\right\}^{\frac{1}{p}} + q\mathbb{E}\left\{\left|\left[\widehat{\eta}_i^k - \eta_i^N\right](h_i)\right|^p\right\}^{\frac{1}{p}} + (1-q)\mathbb{E}\left\{\left|\left[S^N\circ\Phi_i(\eta_{i-1}^N) - \Phi_i(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{\frac{1}{p}}\right) \quad (52)$$

First, Lemma 1 and Lemma 2 can be used to upper bound the second and third terms in the bracketed expression. Second, we observe that:

$$\mathbb{E}\left\{\left|\left[\widehat{\eta}_i^k - \eta_i^N\right](h_i)\right|^p\right\}^{\frac{1}{p}} = \quad (53)$$

$$= \mathbb{E}\left\{\left|\int_{E_i}\left(\frac{d\widehat{\eta}_i^k}{dx}(x) - \frac{d\eta_i^N}{dx}(x)\right)h_i(x)dx\right|^p\right\}^{\frac{1}{p}} \leq \mathbb{E}\left\{\left(\int_{E_i}\left|\frac{d\widehat{\eta}_i^k}{dx}(x) - \frac{d\eta_i^N}{dx}(x)\right|dx\right)^p\right\}^{\frac{1}{p}}||h_i|| \quad (54)$$

Furthermore, the following relationship between the $\mathbb{L}_1$ norm and KL–divergence [10] can be used:

$$\left(\int_{E_i}\left|\frac{d\widehat{\eta}_i^k}{dx}(x) - \frac{d\eta_i^N}{dx}(x)\right|dx\right)^2 \leq 2D(\eta_i^N||\widehat{\eta}_i^k) \quad (55)$$

The results of Li and Barron [13] and Rakhlin [14] demonstrate that the following holds:

$$D(\eta_i^N||\widehat{\eta}_i^k) - D(\eta_i^N||\mathcal{G}) = -\frac{1}{N}\sum_{i=1}^{N}\log\left(\frac{d\widehat{\eta}_i^k}{dx}(x_i)\right) + \frac{1}{N}\sum_{i=1}^{N}\log\left(\frac{d\eta^*}{dx}(x_i)\right) \leq \frac{\gamma c_{F_N,P}^2}{k} \quad (56)$$

where for densities in class $\mathcal{C}$ upper–bounded by b and lower–bounded by a, $\gamma c_{F_N,P}^2 < c = 4\frac{b^2}{a^2}(2+\log(b/a))$ and

$$\frac{d\eta^*}{dx} = \arg\min_{\frac{d\eta}{dx}\in\mathcal{G}} D(\eta_i^N||\eta) \tag{57}$$

Finally, exploiting (55) and (56) and using the Cauchy–Schwarz inequality one can deduce for any $p \geq 1$:

$$\mathbb{E}\left\{\left(\int_{E_i}\left|\frac{d\widehat{\eta}_i^k}{dx}(x) - \frac{d\eta_i^N}{dx}(x)\right|dx\right)^p\right\}^{\frac{1}{p}} \tag{58}$$

$$\leq \sqrt{2}\mathbb{E}\left\{\left(\frac{c}{k} + D(\eta_i^N||\mathcal{G})\right)^{p/2}\right\}^{\frac{1}{p}} \tag{59}$$

$$\leq \sqrt{2}\left(\mathbb{E}\left\{\left(\frac{c}{k} + D(\eta_i^N||\mathcal{G})\right)^{2p/2}\right\}^{\frac{1}{2}}\right)^{1/p} \tag{60}$$

$$\leq \sqrt{2}\left(\frac{c}{k} + \mathbb{E}\left\{D(\eta_i^N||\mathcal{G})^p\right\}^{\frac{1}{p}}\right)^{1/2} \tag{61}$$

Combining this result with the bounds on the other terms in the error decomposition leads directly to the bound presented in the theorem. ■

V. Discussion

Theorems 1 and 2 provide $\mathbb{L}_p$ bounds for leader node particle filtering. Although these bounds are not time–uniform, they do help us understand the behaviour of leader node particle filtering. In particular, these bounds show that approximation errors decrease at the rate $1/\sqrt{N_2}$ and $1/\sqrt{k}$ in the non–parametric and parametric cases respectively. We can also see that the approximation error in the parametric case contains a bias term that has the potential to accumulate over time even if k is very large. However, if we impose reasonable conditions on the error terms, we can conclude that this does not necessarily lead to divergence. The necessary assumption is analogous to the assumption of exponential stability of the underlying semigroups that is often made to prove uniform stability of centralized particle filters. Specifically we assume that the relative oscillations $r_{i,t}$ of the potential functions are bounded and that the Markov transitions $P_{i,t}$ are sufficiently contracting, so that the following condition holds:

$$\lim_{t\to\infty}\sum_{i=0}^{t} r_{i,t}\beta(P_{i,t}) = D < \infty. \tag{62}$$

If the bias introduced at every approximation step can be bounded uniformly in time $\sup_i \mathbb{E}\{|D(\eta_i^N||\mathcal{G})|^p\}^{1/p} = \sup_i \mathbb{E}\{|\inf_{\bar{\eta}\in\mathcal{G}} D(\eta_i^N||\bar{\eta})|^p\}^{1/p} = F < \infty$, then the limiting error at every time step t, $\lim_{k,N\to\infty}\mathbb{E}\{\mathbb{L}_p(t)\}$, is bounded by a constant $\sqrt{2}qFD$. A formal development of results of this type is the subject of future work.

VI. Conclusions

We have presented the analysis of the leader node distributed particle filter. Our results have the form of upper bounds on the expected $\mathbb{L}_p$ error of this filter in two scenarios. First, we considered the case where bootstrap approximation to a posterior distribution is used to transmit particles during leader node exchange. Second, we analyzed the situation where parametric approximation is used for this purpose. Unlike previous results, our bounds indicate that, under reasonable assumptions about the underlying state–space model, convergent particle filter behaviour is expected in both scenarios.

References

[1] A. Jazwinski, *Stochastic processes and filtering theory*. New York: Academic Press, 1970.

[2] R. R. Brooks, P. Ramanathan, and A. Sayeed, "Distributed target classification and tracking in sensor networks," *Proc. IEEE*, vol. 91, pp. 1163–1171, Aug. 2003.

[3] D. Alspach and H. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *IEEE Trans. Automatic Control*, vol. 17, pp. 439–448, Aug. 1972.

[4] F. Martinerie, "Data fusion and tracking using HMMs in a distributed sensor network," *IEEE Trans. Aerospace and Electronic Systems*, vol. 33, pp. 11–28, Jan. 1997.

[5] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. Berlin: Springer–Verlag, 2001.

[6] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Sig. Processing*, vol. 50, pp. 174–188, Feb. 2002.

[7] M. Coates, "Distributed particle filtering for sensor networks," in *Proc. IEEE/ACM Int. Symp. Information Processing in Sensor Networks*, Berkeley, CA, Apr. 2004.

[8] X. Sheng and Y.-H. Hu, "Distributed particle filter with GMM approximation for multiple target localization and tracking in wireless sensor network," in *Proc. IEEE/ACM Int. Symp. Information Processing in Sensor Networks*, Los Angeles, CA, Apr. 2005.

[9] F. Zhao, J. Shin, and J. Reich, "Information-driven dynamic sensor collaboration," *IEEE Signal Processing Magazine*, vol. 19, pp. 61–72, Mar. 2002.

[10] A. Ihler, J. Fisher, and A. Willsky, "Particle filtering under communication constraints," in *Proc. IEEE Workshop on Statistical Signal Processing*, Bordeaux, France, May 2005.

[11] P. Del Moral, *Feynman–Kac formulae. Genealogical and interacting particle approximations*. New York: Springer, 2004.

[12] F. L. Gland and N. Oudjane, "A robustification approach to stability and to uniform particle approximation of nonlinear filters : the example of pseudo-mixing signals," *Stochastic Processes and their Applications*, vol. 38, no. 2, pp. 279–316, Aug. 2003.

[13] J. Li and A. Barron, "Mixture density estimation," in *Advances in Neural Information Processing Systems 12*, S. A. Solla, T. K. Leen, and K.-R. Muller, Eds. San Mateo, CA.: Morgan Kaufmann Publishers, 1999.

[14] A. Rakhlin, D. Panchenko, and S. Mukherjee, "Risk bounds for mixture density estimation," *ESAIM: Probab. and Statist.*, vol. 9, pp. 220–229, 2005.

[15] A. Ihler, J. Fisher, R. Moses, and A. Willsky, "Nonparametric belief propagation for self-calibration in sensor networks," in *Proc. IEEE/ACM Int. Symp. Information Processing in Sensor Networks*, Berkeley, CA, Apr. 2004.

[16] M. Briers, A. Doucet, and R. Evans, "Sequential auxiliary belief propagation," in *Int. Conf. Information Fusion*, Philadelphia, PA, July 2005.

[17] C. M. Kreucher, A. O. Hero, K. D. Kastella, and M. R. Morelande, "An information based approach to sensor management in large dynamic networks," *Proc. IEEE*, vol. 95, no. 5, pp. 978–999, May 2007.

[18] T. Zhang, "Sequential greedy approximation for certain convex optimization problems," *IEEE Trans. Info. Theory*, vol. 49, no. 3, pp. 682–691, 2003.

[19] M. Donahue, L. Gurvits, C. Darken, and E. Sontag, "Rates of convex approximation in non-Hilbert spaces," *Constructive Approximation*, vol. 13, pp. 187–220, 1997.

[20] A. van der Vaart and J. Wellner, *Weak Convergence and Empirical Processes*. New York: Springer, 1996.