Error Propagation in Gossip-Based Distributed Particle Filters

Syamantak Datta Gupta, Mark Coates, *Senior Member, IEEE*, and Michael Rabbat, *Senior Member, IEEE*

***Abstract*—This paper examines the impact of the gossip procedure on distributed particle filters that employ averaging to estimate the global likelihood function. We consider a model where a gossip-driven algorithm leads to the use of a slightly distorted version of the likelihood function, *in lieu* of its true value. Under standard regularity conditions, and a mild assumption on the true likelihood function, we derive a time-uniform bound on the weak-sense L_p error of the filter. Furthermore, we present an associated exponential inequality for the large deviations of the filter. These bounds capture the combined effects of sampling and consensus-based approximation. The results allow us to evaluate the impact of such approximations on the overall performance of the distributed particle filter, and analyze its stability. Finally, through numerical experiments, we demonstrate the practical implications of these results and explore the relationship of the performance of the filter with these theoretical error bounds.**

***Index Terms*—Communication overhead, consensus, distributed nonlinear filtering, Feynman-Kac models, gossip algorithms, stability.**

I. Introduction

PARTICLE filters [1]–[3] are extensively used as an efficient tool for addressing challenging problems of target tracking. In such problems, the requirement is to accurately estimate the probability distribution of the state of the system (e.g., position and/or velocity of a moving target), using a set of noisy observations. The problem can be represented by discrete time hidden Markov models when the present state of the system is assumed to evolve as a function of its state at the preceding time instant. Both the state evolution dynamics and the observation model are often substantially non-linear and non-Gaussian, which makes tracking difficult. In the particle filter approach, a set of point mass samples, known as "particles", are maintained at every time-step, each of which are candidate hypotheses of the state of the variable of interest. Particles are assigned weights based on how well they correspond with the observed measurements and system dynamics. The weight of a particle depends on the likelihood function associated with it, and these weights are updated at every time step, as new observations are available. The posterior probability distribution of the state is approximately represented by this "cloud" of weighted particles.

We consider a wireless sensor network setting where multiple sensor nodes make measurements at each time-step. In gossip-based distributed particle filter algorithms [4]–[17], each node runs an individual particle filter and approximately evaluates a global statistic through consensus. In some of these algorithms [4]–[7] each node samples from the same set of particles and maintains an identical particle representation of the previous global posterior by approximating the likelihood function. Under an assumption of conditional independence, the global likelihood function is equal to the product of the individual sensor likelihood functions [18], and so these algorithms strive to construct an approximate global log-likelihood at each node by summing the local log-likelihoods. When this summation is computed using a finite number of gossip iterations, there will be an error in the likelihood calculation since gossip algorithms converge asymptotically. In this paper, we analyze the error propagation in this class of algorithms, where the sensor nodes are synchronized. The overall computational complexity of the synchronized distributed particle filter increases compared to the centralized filter, but the focus is on reduced communication overhead, because communication can often be the dominant consumer of energy in a distributed sensing system.

In centralized particle filters, the sampling procedures at each time-step lead to discrepancies between the particle representation and the true posterior filtering distribution. Consequently, it is of interest to identify sufficient conditions that allow this error to remain bounded over time, so that the particle filter remains stable. Analysis of the stability of these filters has generated significant interest over the years, and bounds have been derived that characterize the stability of particle filters and how the sampling errors propagate [1], [19]–[25]. A more detailed discussion of these results is presented in Section VII.

In a distributed particle filter approach that approximates the likelihood function through consensus, the discrepancy between the true distribution of the state variables and that represented by the particle cloud is further aggravated due to the use of an approximate version of the log-likelihood function. In [26], a convergence result was presented for a specific distributed particle filter algorithm. However, there are not many results available in the literature on the propagation of error in distributed particle filters, especially for those using gossip algorithms.

In this paper, we quantify the extent to which the additional error arising from the likelihood approximation in distributed

Manuscript received January 20, 2015; revised June 22, 2015; accepted August 11, 2015. Date of publication August 24, 2015; date of current version September 25, 2015. This work was supported by PWGSC contract W7707-145675/001/HAL funded by Defence R&D Canada. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Franz Hlawatsch.

The authors are with the Department of Electrical and Computer Engineering, McGill University, Montreal H3A 0E9 QC, Canada. (e-mail: syamantak.dattagupta@mcgill.ca; mark.coates@mcgill.ca; michael.rabbat@mcgill.ca).

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/TSIPN.2015.2471846

particle filters has an impact on the performance bounds. We consider a model where the gossip-based approximation distorts the likelihood function by an exponent that lies between $(1+\delta)$ and $(1-\delta)$, where δ is a small positive fraction. Following the framework presented in [22], we use a Feynman-Kac model to analyze the problem.

The main contributions of this paper are as follows. We present a time-uniform bound on the weak-sense L_p error induced by the combined effects of sampling and an approximate evaluation of the likelihood function. We also present an exponential estimate of the large deviations of the filter. Our results show that for the class of consensus algorithms considered, the performance of the distributed particle filter remains stable over time.

A. Paper Organization

The rest of this paper is organized as follows. Basic notation is introduced in Subsection I-B. A detailed description of the problem is provided in Section II, along with an introduction of the mathematical tools used in the analysis. In Section III, we present a time-uniform bound on the error propagation and a large deviation result for consensus-based distributed particle filters. The relationship between the number of gossip iterations performed and the accuracy of the likelihood approximations is analyzed in Section IV. The proofs of the main results are provided in Section V. Section VI describes numerical experiments that illustrate the practical implications of the theoretical bounds derived in this paper. Section VII gives a detailed discussion of related work. Finally, Section VIII concludes with a summary of the contributions and comments on possible extensions of this work.

B. Notation

For any measurable space E and a function $h : E \to \mathbb{R}$, let

$$\|h\|_\infty \overset{\text{def}}{=} \sup_{x\in E} |h(x)|$$

denote the supremum norm, and let $\mathcal{B}_b(E)$ be the Banach space of bounded functions with respect to the supremum norm. Let $\operatorname{osc}(h)$ be the oscillation of a function $h \in \mathcal{B}_b(E)$:

$$\operatorname{osc}(h) \overset{\text{def}}{=} \sup_{x,y\in E} |h(x) - h(y)|.$$

Let $\mathcal{E}$ be the Borel σ-algebra on E, and let E_t, $t \in \mathbb{Z}_+$, be a sequence of measurable spaces associated with the Borel σ-algebras $\mathcal{E}_t$. Define the product space $E_{i:t} \overset{\text{def}}{=} \prod_{j=i}^{t} E_j$. Let $\|\cdot\|_2$ denote the Euclidean norm and $\|\cdot\|_{\text{tv}}$ denote the total variation norm. Let $\mathcal{P}(E)$ be the set of all probability measures defined on E. For a measure $P \in \mathcal{P}(E)$, let $\mathbb{E}_P\{\cdot\}$ denote the expectation with respect to P, and with a slight abuse of notation, let us denote the action of P on a test function $h \in \mathcal{B}_b(E)$ by $P(h) \overset{\text{def}}{=} \mathbb{E}_P\{h(X)\} = \int_E h(x)P(dx)$. Finally, we write $X_{1:t}$ as shorthand for the sequence $(X_1, X_2, \ldots, X_t)$.

II. Background and Problem Formulation

A. The Filtering Problem and a Feynman-Kac Model

We strive to derive bounds for sensor networks that address a discrete-time non-linear filtering task. We use the following non-homogeneous Markov state-space model to capture the target dynamics and observations:

$$X_t = f_t(X_{t-1}, \varrho_t) \tag{1}$$
$$Y_t = g_t(X_t, \zeta_t). \tag{2}$$

Here $X_t \in E_t$ is the target state vector, $Y_t \in F_t$ are the observations and ϱ_t and ζ_t are system excitation noise and measurement noise respectively, at time t. The non-linear function f_t expresses the target dynamics and the non-linear measurement map g_t relates the state to the measurements.

Typically, we are interested in finding an estimate of the expected value of a function $h_t(X_t)$ of the state, conditioned on a series of observed realizations $y_{1:t}$ of $Y_{1:t}$,

$$\mathbb{E}\{h_t(X_t)|y_{1:t}\} = \int_{E_t} h_t(x_t)\mathrm{d}\mathbb{P}_t(x_t|y_{1:t}),$$

where $\mathbb{P}_t(X_t|y_{1:t})$ is the conditional distribution of X_t, given $Y_{1:t} = y_{1:t}$, induced by the measurement noise ζ_t through the observation model (2). For example, in a tracking problem the function $h_t(\cdot)$ could be the displacement, velocity, distance, or angular position of the target. When a conditional probability density function exists, this expectation is given by:

$$\mathbb{E}\{h_t(X_t)|y_{1:t}\} = \int_{E_t} h_t(x_t)p_t(x_t|y_{1:t})\mathrm{d}x_t.$$

In the particle filter approach, this expectation is approximated by summing over a finite collection of weighted particles which each can be interpreted as a candidate value for the target state.

The error propagation results presented in this paper build on bounds developed by Del Moral [22], who models the evolution of distributions in a particle filter as a particle approximation of a Feynman-Kac model. In this approach, the predictive posterior conditional density $p_t(x_t|y_{1:t-1})$ is represented by a measure η_t, and the updated posterior conditional density $p_t(x_t|y_{1:t})$ is represented by a measure $\Psi_t(\eta_t)$.

The Markov chain transitions from E_{t-1} to E_t according to an integral operator, $M_t(x_{t-1}, \mathrm{d}x_t)$, which captures the evolution of the signal diffusion in (1). $M_t(\cdot,\cdot)$ is the Markov transition kernel, and it is related to the state evolution as follows:

$$M_t(x_{t-1}, \mathrm{d}x_t) \overset{\text{def}}{=} \mathbb{P}\{X_t \in \mathrm{d}x_t | X_{t-1} = x_{t-1}\},$$

where $\mathbb{P}\{X_t \in \mathrm{d}x_t | X_{t-1} = x_{t-1}\}$ is the distribution induced by the excitation noise ϱ_t through the dynamic model (1). When the state transition density exists, $M_t(\cdot,\cdot)$ can be expressed as

$$M_t(x_{t-1}, \mathrm{d}x_t) = p_t(x_t|x_{t-1})\mathrm{d}x_t.$$

The likelihood function $p_t(y_t|x_t)$ is modeled by a potential function $G_t : E_t \to (0, 1]$, such that $G_t(x_t) \propto p_t(y_t|x_t)$. We have $G_t(x_t) = \prod_{j\in\mathcal{S}_t} G_{j,t}(x_t)$, where $G_{j,t}(x_t)$ is the local potential function corresponding to the measurements made at sensor j, and $\mathcal{S}_t$ is the set of sensors that make measurements at time t. For the purpose of analysis, we assume that appropriate normalization can be conducted so that the potential functions take values in $(0, 1]$.

For any $h_t \in \mathcal{B}_b(E_t)$ we can define a measure $\eta_t \in \mathcal{P}(E_t)$:

$$\eta_t(h_t) \overset{\text{def}}{=} \gamma_t(h_t)/\gamma_t(1)$$
$$\text{with} \quad \gamma_t(h_t) \overset{\text{def}}{=} \mathbb{E}_{\eta_0}\left(h_t(X_t)\prod_{\tau=0}^{t-1} G_\tau(X_\tau)\right).$$

Based on the potential function, define the Boltzmann-Gibbs transformation operator $\Psi_t(\cdot)(\mathrm{d}x_t)$ that acts on any measure $\nu \in \mathcal{P}(E_t)$:

$$\Psi_t(\nu)(\mathrm{d}x_t) \overset{\text{def}}{=} \frac{1}{\nu(G_t)} G_t(x_t)\nu(\mathrm{d}x_t). \tag{3}$$

The operator $\Psi_t(\eta_t)(\mathrm{d}x_t)$ generates the normalized posterior distribution. The prediction step of the filter can be formulated by combining the Markov diffusion operator with Ψ_t, to define an operator:

$$\Phi_t(\eta_{t-1})(\mathrm{d}x_t) \overset{\text{def}}{=} \int_{E_{t-1}} \Psi_{t-1}(\eta_{t-1})(\mathrm{d}x_{t-1}) M_t(x_{t-1}, \mathrm{d}x_t). \tag{4}$$

This generates the normalized *predictive* posterior distribution $\eta_t(\mathrm{d}x_t) = \Phi_t(\eta_{t-1})(\mathrm{d}x_t)$.

A direct analogy between the Feynman-Kac model and the predict-update Bayesian recursion framework is apparent. The diffusion step defined in (4) corresponds to the prediction stage of the Bayesian model, since

$$p_t(x_t|y_{1:t-1}) = \int_{E_{t-1}} p_{t-1}(x_{t-1}|y_{1:t-1}) p_t(x_t|x_{t-1}) \mathrm{d}x_{t-1}.$$

The Boltzmann-Gibbs transformation $\Psi_t(\cdot)$ in (3) corresponds to the update stage of the Bayesian filter, as

$$p_t(x_t|y_{1:t}) = \frac{p_t(y_t|x_t)p_t(x_t|y_{1:t-1})}{\int_{E_t} p_t(y_t|x_t)p_t(x_t|y_{1:t-1})\mathrm{d}x_t}.$$

B. Particle Approximations

A particle filter can be defined by developing an N-particle approximation to the Feynman-Kac model, consisting of N path particles:

$$\{\xi_i^k\}_{0 \le i \le t} \in E_{0:t} \quad k \in 1, \ldots, N.$$

The posterior distribution is represented by the particle approximation of the prediction Feynman-Kac model as:

$$\eta_t^N \overset{\text{def}}{=} \frac{1}{N}\sum_{k=1}^{N} \delta_{\xi_t^k} \ ,$$

where δ is the Dirac delta function.

Let the N-tuple ξ_t denote the configuration at time t of N particles $\{\xi_t^k\}_{k=1}^N$. Thus $\{\xi_t^k\}_{k=1}^N$ is an element in the product space E_t^N. The particle filter can be represented by a two-step updating process:

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

In the selection stage of each time-step, N particles $\{\widehat{\xi}_t^k\}_{k=1}^N$ are randomly selected from the particle cloud. This random selection is achieved by setting, with probability $\alpha_t G_t(\xi_t^k)$, $\widehat{\xi}_t^k = \xi_t^k$; otherwise a random particle $\widetilde{\xi}_t^k$ is chosen 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 $\widehat{\xi}_t^k = \widetilde{\xi}_t^k$. Here α_t is a small positive parameter that controls the selection procedure. It determines how often particles are retained and how often they are replaced by random selection from the population. During the mutation stage, each particle $\widehat{\xi}_t^k$ evolves according to the Markov transition M_{t+1}.

For any function $h \in \mathcal{B}_b(E)$, let the sampling operator $S^N : \mathcal{P}(E) \to \mathcal{P}(E^N)$ be defined as:

$$S^N(\eta)(h) \overset{\text{def}}{=} \frac{1}{N}\sum_{k=1}^{N} h(\xi^k),$$

where $(\xi^1, \ldots, \xi^N)$ is an i.i.d. sample of particle locations from η. With this notation, the standard particle filter can be expressed using the recursion $\eta_t^N = S^N(\Phi_t(\eta_{t-1}^N))$.

C. Consensus-induced error

When we run a consensus algorithm in an attempt to evaluate the global likelihood function, there is an estimation error. If there are $|\mathcal{S}_t|$ sensors that take measurements at time t, a common strategy is to perform consensus over the normalized log-likelihoods, evaluated at each particle value, and then multiply by $|\mathcal{S}_t|$. In the situation where the communication network topology is fixed and an average consensus algorithm is executed for a fixed number of iterations, the evaluated global likelihood (or potential) function is a deterministic function of the particle representations ξ_t^k. We can thus identify a potential function, $\widehat{G}_t$, that corresponds to the estimated likelihood at time t. If a randomized gossip algorithm is used, then the error is random and dependent on the sequence of nodes that are selected to perform gossip operations. If we denote this random sequence by χ, then the (random) potential function associated with the estimated likelihood can be denoted by $\widehat{G}_{t,\chi}$. In the following discussion, we focus on $\widehat{G}_t$, but note that equivalent relationships can be specified for $\widehat{G}_{t,\chi}$.

With the potential function $\widehat{G}_t$ in hand, we can construct a Feynman-Kac model for the distributed particle filter. Let us define an operator $\widehat{\Psi}$:

$$\widehat{\Psi}_t(\nu)(\mathrm{d}x_t) \overset{\text{def}}{=} \frac{1}{\nu(\widehat{G}_t)} \widehat{G}_t(x_t)\nu(\mathrm{d}x_t).$$

Here $\widehat{G}_t$ is the potential function corresponding to the estimated likelihood at time t. Using this, we can define an operator $\widehat{\Phi}_t(\eta_{t-1}) \overset{\text{def}}{=} \widehat{\Psi}_{t-1}(\eta_{t-1})M_t$, and consequently express the particle approximation to the Feynman-Kac model associated with the potential functions $\widehat{G}$ as

$$\widehat{\eta}_t^N \overset{\text{def}}{=} S^N(\widehat{\Phi}_t(\widehat{\eta}_{t-1}^N)). \tag{5}$$

III. Bounds on the Propagated Error for a Gossip Based Distributed Particle filter

In this section, we present two theorems on the error propagation in a distributed particle filter using a gossip-driven algorithm. The first theorem provides a time-uniform bound on the weak-sense mean L_p error, while the second gives an associated exponential inequality. We begin by stating a few conditions imposed on the system.

A. Regularity Conditions and Assumptions

The performance bounds we build upon, as derived in [22], are expressed in terms of regularity and mixing conditions on the Markov operator M_t and the potential function G_t. In particular, Del Moral specifies the following condition in [22, Section 3.5.2].

- Condition $(M)^{(m)}$: Given $m > 0$, there exists a strictly positive number $\epsilon_M \in (0,1)$ such that for any $t \geq 0$ and $x_t, y_t \in E_t$,

$$\begin{aligned} M_{t,t+m}(x_t,\cdot) &\stackrel{\text{def}}{=} M_{t+1}M_{t+2}\dots M_{t+m}(x_t,\cdot) \\ &\geq \epsilon_M M_{t,t+m}(y_t,\cdot). \end{aligned}$$

Condition $(M)^{(m)}$ implies that the Markov chain associated with state evolution undergoes sufficient mixing within a finite number of time steps. Additionally, we make the following assumption on the potential function.

- Condition (G): The potential function G_t is bounded away from zero, and appropriately normalized, so that there exists some $\epsilon_G > 0$ such that for any t and for all $x_t \in E_t$:

$$\epsilon_G \leq G_t(x_t) \leq 1.$$

It follows immediately from Condition (G), that for any t and $x_t, y_t \in E_t$,

$$G_t(x_t) \geq \epsilon_G G_t(y_t). \tag{6}$$

Del Moral [22, Section 3.5.2] uses (6) as a regularity condition to analyze error propagation in particle filters. Since the potential function is proportional to the likelihood function $p_t(y_t|x_t)$, condition (G) implies that $p_t(y_t|x_t) > 0$ for all t, $x_t \in E_t$, $y_t \in F_t$, i.e., the likelihood function is sufficiently flat over its range. This is a reasonable assumption when the space of observations F_t is bounded.

- Condition (h): The test function h_t is such that $\|h_t\|_\infty \leq 1$ for all $t \geq 0$.

The above condition implies that $\sup_{x_t \in E_t, t\geq 0} |h_t(x_t)| \leq 1$. In existing literature on the asymptotic properties of the propagated error in particle filters, eg, [22], [27], the test function h_t is typically chosen such that $h_t \in \text{Osc}_1(E_t)$, where $\text{Osc}_1(E_t) \subset \mathcal{B}_b(E_t)$ denotes the set of $\mathcal{E}_t$-measurable test functions with oscillations at most unity,

$$\text{Osc}_1(E_t) \stackrel{\text{def}}{=} \{h \in \mathcal{B}_b(E_t) : \text{osc}(h) \leq 1\}.$$

For $h_t \in \text{Osc}_1(E_t)$, for any $x_t \in E_t$, $t \geq 0$, $h_t(x_t)$ can be expressed as

$$h_t(x_t) = \underline{h}_t + \tilde{h}_t(x_t),$$

where $\underline{h}_t = \inf_{x_t \in E_t} h_t(x_t)$ is constant for a given t, and $\tilde{h}_t(x_t) \in [0,1]$. The only variability in $h_t(x_t)$, in that case, arises from the bounded oscillations $\tilde{h}_t(x_t)$, and therefore the part $\underline{h}_t$ has no impact on the estimation error. Consequently, Condition (h) is a reasonable extension of the condition $h_t \in \text{Osc}_1(E_t)$, where we set the constant part $\underline{h}_t$ to zero. Conversely, it is easy to see that when Condition (h) is satisfied, $\text{osc}(h_t) \leq 2$.

Finally, the following assumption restricts the extent to which the approximate potential function deviates from its true value.

Assumption 1: The gossip or consensus algorithm can achieve an accuracy such that for any t and for all $x_t \in E_t$:

$$\frac{|\log \widehat{G_t}(x_t) - \log G_t(x_t)|}{|\log G_t(x_t)|} \leq \delta, \tag{7}$$

where $\delta < 1$ is a small non-negative constant.

The main results are presented below. The proofs of both theorems are given in Section V.

B. Main Results

Theorem 1: Suppose Conditions $(M)^{(m)}$ and (G) hold and Assumption 1 holds, and h_t satisfies Condition (h). Then for $p \geq 1$ and for any $N \in \mathbb{N}$, we have a time uniform estimate

$$\sup_{t\geq 0} \mathbb{E}\left\{|[\widehat{\eta}_t^N - \eta_t](h_t)|^p\right\}^{1/p} \leq \epsilon_0 \left(\frac{2c(p)^{\frac{1}{p}}}{\sqrt{N}} + \delta|\log \epsilon_G|\right),$$

where the constant ϵ_0 is given by:

$$\epsilon_0 = \frac{2m}{\epsilon_M^3 \epsilon_G^{(2m-1)}}, \tag{8}$$

and $c(p)$ is a constant depending only on p.

This result indicates that a distributed particle filter based on a consensus algorithm does not induce instability in terms of error propagation, and the error remains bounded over time. It is observed that the propagated error depends on the parameter δ, bounding the relative amount of distortion in the potential function approximation, and ϵ_G, the lowest value assumed by the potential function. It follows that the term $\epsilon_0 \delta \log|\epsilon_G|$ gives an upper bound of the L_p norm of the propagated error, as the number of particles $N \to \infty$.

It is instructive to compare this bound with that corresponding to a centralized particle filter. When the effect of approximation of the potential function due to gossip is absent, the corresponding bound for the propagated error is given by:

$$\sup_{t\geq 0} \mathbb{E}\left\{|[\eta_t^N - \eta_t](h_t)|^p\right\}^{1/p} \leq \frac{2\epsilon_0 c(p)^{\frac{1}{p}}}{\sqrt{N}}. \tag{9}$$

A derivation of this result is presented in Subsection V-B. A similar result appears in [22, Theorem 7.4.4]. This gives a bound on the weak-sense L_p error for the centralized particle filter, which arises solely due to sampling.

The next result provides the exponential estimate for the probability of large deviations of the approximate Feynman-Kac model associated with the gossip-based distributed particle filter.

Theorem 2: Suppose Conditions $(M)^{(m)}$ and (G) hold and Assumption 1 holds, and let h_t satisfy Condition (h). Then, for any number of samples $N \in \mathbb{N}$, and for any $\epsilon > 0$, we have

$$\sup_{t\geq 0} \mathbb{P}\{|[\widehat{\eta}_t^N - \eta_t](h_t)| \geq (1+\epsilon)\epsilon_0 \delta |\log \epsilon_G\|\} \leq \left(1 + \sqrt{2\pi N}\delta \log|\epsilon_G|\epsilon\right) e^{-\frac{(\delta|\log \epsilon_G|)^2 \epsilon^2 N}{8}}.$$

where ϵ_0 is defined in Theorem 1.

Theorem 2 gives an estimate of the probability that the L_1 estimation error exceeds the bound given by Theorem 1 by a small amount. This probability depends on the parameter of distortion δ and ϵ_G, the lower bound of the potential function.

Both theorems rely on Assumption 1. Recall that for the k-th particle ξ_t^k, the potential function $G_t(\xi_t^k)$ (corresponding to the observation $Y_t = y_t$) is approximated by the sensor $j \in \mathcal{S}_t$, as $\widehat{G}_{j,t}(\xi_t^k)$, where $\mathcal{S}_t$ is the set of sensor nodes active at time t. The final potential function is estimated as

$$\widehat{G}_t(\xi_t^k) = \prod_{j\in\mathcal{S}_t} \widehat{G}_{j,t}(\xi_t^k).$$

For $\widehat{G}_t(\xi_t^k)$ to satisfy Assumption 1, we must have

$$(1+\delta)\log(G_t(\xi_t^k)) \leq \log(\widehat{G}_t(\xi_t^k)) \leq (1-\delta)\log(G_t(\xi_t^k)),$$

which implies that

$$(1+\delta)\log(G_t(\xi_t^k)) \leq \sum_{j\in\mathcal{S}_t} \log(\widehat{G}_{j,t}(\xi_t^k)) \leq (1-\delta)\log(G_t(\xi_t^k)).$$

Therefore, a sufficient condition for Assumption 1 to hold is

$$\frac{(1+\delta)}{|\mathcal{S}_t|}\log(G_t(\xi_t^k)) \leq \log(\widehat{G}_{j,t}(\xi_t^k)) \leq \frac{(1-\delta)}{|\mathcal{S}_t|}\log(G_t(\xi_t^k)),$$

for all $k = 1, \ldots, N$ and for all $j \in \mathcal{S}_t$. The next section studies how this condition can be satisfied when gossip iterations are used for distributed (approximate) calculation of the joint log-likelihood.

IV. Gossip Communication Overhead and Error Bounds

The particle filter error bounds stated in the previous section depend on the accuracy $0 \leq \delta < 1$ of the approximate log-potential values used. In gossip-based distributed particle filters [28], the accuracy δ can be related to the number of gossip iterations and the particular sort of likelihood approximation used. To illustrate this relationship we consider a simple method which uses gossip-based synchronous distributed averaging iterations to fuse the potential (weight) values associated with each particle.

Let $\mathcal{S}$ denote the set of all sensors (including those that do not gather a measurement at time t), and let $z_j(0)$ denote the initial value at sensor j. Synchronous gossip [29]–[32] for distributed averaging is a message passing implementation of linear iterations: sensor j updates

$$z_j(\ell) = \sum_{j'\in\mathcal{S}} a_{j,j'} z_{j'}(\ell-1), \tag{10}$$

where $\ell = 1, 2, \ldots,$ is the gossip iteration index. Suppose that there are $n = |\mathcal{S}|$ sensors. Let A denote the $n \times n$ matrix with $a_{j,j'}$ as its (j,j')'th element. We assume that $a_{j,j} > 0$ for all $j \in \mathcal{S}$ and that for $j \neq j'$, $a_{j,j'} > 0$ if and only if sensors j and j' communicate directly. We assume that communication relationships are symmetric: $a_{j,j'} > 0$ implies that $a_{j',j} > 0$. We also assume that the network is connected. Finally, we assume that the matrix A is doubly-stochastic: its has row-sums and column-sums equal to 1. These assumptions, which are standard in the literature [31], [32], are made here for convenience, although they can be relaxed in many situations [33] without significantly changing the results.

Under the assumptions mentioned above, it is well-known that A can be viewed as the probability transition matrix of an ergodic Markov chain, and the stationary distribution is uniform over the state-space $\mathcal{S}$. It follows that for all sensors $j \in \mathcal{S}$, the value $z_j(\ell)$ asymptotically converges to the average of the initial values across the network: $\lim_{\ell\to\infty} z_j(\ell) = \bar{z}$, where

$$\bar{z} \stackrel{\text{def}}{=} \frac{1}{|\mathcal{S}|}\sum_{j\in\mathcal{S}} z_j(0).$$

It is also well-known that the rate of convergence depends on the connectivity of the communication network, as captured by spectral properties of the matrix A. We express such a convergence result in the following lemma, which follows from arguments similar to those used in the proof of [33, Corollary 5.2].

Lemma 1: Let A be doubly-stochastic and correspond to a connected communication network. Let $1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n$ denote the eigenvalues of A sorted in descending order, and let $\rho_A = \max\{|\lambda_2|, |\lambda_n|\}$ denote the second largest eigenvalue in modulus. For all initial conditions $\{z_j(0)\}_{j\in\mathcal{S}}$ and for $0 < C < 1$, it holds that

$$\max_{j\in\mathcal{S}} |z_j(\ell) - \bar{z}| \leq C \cdot \max_{j\in\mathcal{S}} |z_j(0) - \bar{z}|$$

if the number of gossip iterations ℓ satisfies

$$\ell \geq \frac{(3/2)\log(n) + \log(1/C)}{\log(1/\rho_A)}. \tag{11}$$

In the context of distributed particle filtering, the sensors execute N instances of gossip-based distributed averaging in parallel, one instance for each particle. For the k-th particle ξ_t^k at time step t, $k = 1, \ldots, N$, sensor $j \in \mathcal{S}$ initializes a gossip value to

$$z_j^k(0) = \begin{cases} |\mathcal{S}|\log\left(G_{j,t}(\xi_t^k)\right) & \text{if } j \in \mathcal{S}_t \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathcal{S}_t$ is the set of sensors that take a measurement at time step t. Then the limiting gossip value is

$$\bar{z}^k = \sum_{j\in\mathcal{S}_t} \log\left(G_{j,t}(\xi_t^k)\right) \equiv \log\left(G_t(\xi_t^k)\right),$$

the exact log-potential value for the particle ξ_t^k.

The following lemma characterizes how large the initial relative error may be, given that the initial values at each sensor are bounded and non-positive.

Lemma 2: Suppose that there exists a positive constant a such that the initial values are in the interval $-a \le z_j(0) \le 0$ for all $j = 1,\ldots,n$, and suppose that the average of the initial values is equal to $\bar{z} \in [-a, 0]$. Then

$$\max_{j=1,\ldots,n} \frac{|z_j(0) - \bar{z}|}{|\bar{z}|} \le (n-1).$$

Let $\delta \in [0,1)$ be given, and suppose that the approximate log-potential values $\log \widehat{G}_t(\xi^k)$ are obtained by running ℓ iterations of distributed averaging, followed by a max-gossip procedure to ensure that the values at all sensors are identical. Then taking $C = \delta/(n-1)$ in Lemma 1, combined with Lemma 2, we obtain that the condition from Assumption 1,

$$\frac{\left|\log\left(\widehat{G}_t(\xi^k)\right) - \log\left(G_t(\xi^k)\right)\right|}{\left|\log\left(G_t(\xi^k)\right)\right|} \le \delta,$$

holds as long as

$$\ell \ge \frac{(3/2)\log(n) + \log(\frac{n-1}{\delta})}{\log(1/\rho_A)} = \mathcal{O}\left(\frac{\log(n/\delta)}{\log(1/\rho_A)}\right)$$

gossip iterations are performed at each time step of the distributed particle filter. This directly illustrates the connection between the number of gossip iterations ℓ required to obtain a desired accuracy δ as a function of the network size n and topology (as captured by ρ_A).

To conclude this section we remark that a similar approach as that used in this section could be used to analyze more sophisticated gossip-based distributed particle filters, such as those described in [4], [5], [15], [17], [34]. The derivations above are for a naïve distributed particle filter which runs one instance of gossip on the weight associated with each particle. The methods described in [4], [5], [15], [17], [34] introduce more sophisticated approaches to reduce the communication overhead by incorporating additional approximations (e.g., gossiping more on particles with the largest weights). To obtain similar bounds for these algorithms using the steps outlined above requires characterizing and controlling the additional error introduced by these approximations, which is beyond the scope of this work.

V. Analysis

The proof of Theorem 1 builds upon the framework developed in [22], where the role of the sampling operation and the propagation of the discrepancies between the true and estimated distributions is characterized using a Feynman-Kac model. In this section, we present the derivations, after reviewing some fundamental results from [22], [27] and [35] that are relevant to our analysis.

A. Preliminary Results

Recall the operators $\Psi_t(\cdot)$ and $\Phi_t(\cdot)$ defined in (3) and (4) respectively. The repeated application of the operator $\Phi_t(\eta_{t-1})$, $t \ge 1$, results in the operators $\Phi_{i:t}$, $i \le t$, that govern the evolution of the predictive posterior distribution η from time i to time t:

$$\Phi_{i:t} \stackrel{\text{def}}{=} \Phi_t \circ \Phi_{t-1} \circ \ldots \circ \Phi_{i+1}.$$

We relate the operator $\Phi_{i:t}$ to a composite potential function, $G_{i:t}$, and to a Markov kernel, $P_{i:t}$, that specifies the transition from E_i to E_t. The composite potential function $G_{i:t} : E_i \to (0,\infty)$ is defined as

$$G_{i:t}(x_i) \stackrel{\text{def}}{=} \int_{E_{i+1:t}} \prod_{j=i}^{t-1} G_j(x_j) M_{i+1}(x_i, \mathrm{d}x_{i+1}) \ldots M_t(x_{t-1}, \mathrm{d}x_t).$$

The Markov kernel $P_{i:t} : \mathcal{P}(E_i) \to \mathcal{P}(E_t)$ is defined as:

$$P_{i:t}(h_t) \stackrel{\text{def}}{=} \frac{\int_{E_{i+1:t}} h_t(x_t) \prod_{j=i}^{t-1} G_j(x_j) M_{i+1}(x_i, \mathrm{d}x_{i+1}) \ldots M_t(x_{t-1}, \mathrm{d}x_t)}{\int_{E_{i+1:t}} \prod_{j=i}^{t-1} G_j(x_j) M_{i+1}(x_i, \mathrm{d}x_{i+1}) \ldots M_t(x_{t-1}, \mathrm{d}x_t)}.$$

By introducing the composite operator $\Psi_{i:t}(\eta)(h_i) = \eta(G_{i:t}h_i)/\eta(G_{i:t})$, we can write $\Phi_{i:t}(\eta) = \Psi_{i:t}(\eta)P_{i:t}$. These composite operators describe the evolution of the Feynman-Kac model over time, from i to t.

The Dobrushin contraction coefficient [22], also known as the Dobrushin ergodic coefficient, is defined as follows:

$$\beta(P_{i:t}) \stackrel{\text{def}}{=} \sup\{\|P_{i:t}(x_i,\cdot) - P_{i:t}(y_i,\cdot)\|_{\text{tv}}; x_i, y_i \in E_i\},$$

where $\|\cdot\|_{\text{tv}}$ denotes the total variation norm. The Dobrushin coefficient lies between 0 and 1, and can be interpreted as a distance between the transition probability measures at the points x_i and y_i. It plays an important role in the analysis of asymptotic behaviour of Markov chains. Furthermore, if Conditions $(M)^{(m)}$ and (G) hold, then according to [22, Proposition 4.3.3] we have the following estimate for the Dobrushin contraction coefficient:

$$\beta(P_{i:t}) \le \left(1 - \epsilon_M^2 \epsilon_G^{(m-1)}\right)^{\lfloor (t-i)/m \rfloor}.$$

The following proposition from Del Moral [22, Proposition 4.3.7] underpins our analysis of the stability of the semigroups $\Phi_{i:t}$.

Proposition 1: For any $0 \le i \le t$, $\nu_i \in \mathcal{P}(E_i)$, and $h_t \in \mathcal{B}_b(E_t)$ with $\|h_t\|_\infty \le 1$, there exists a function h_i in $\mathcal{B}_b(E_i)$ with $\|h_i\|_\infty \le 1$ such that for any $\eta_i \in \mathcal{P}(E_i)$,

$$|[\Phi_{i:t}(\eta_i) - \Phi_{i:t}(\nu_i)](h_t)| \le \beta(P_{i:t}) \frac{2\|G_{i:t}\|}{\eta_i(G_{i:t})} |(\eta_i - \nu_i)(h_i)|.$$

This result describes the propagation of the one-step approximation error through the non-linear operator $\Phi_{i:t}$. It reveals the link between the initial error at time i and the propagated error at time t through the properties of the composite potential function $G_{i:t}$ and the Dobrushin contraction coefficient $\beta(P_{i:t})$.

According to Proposition 4.3.3 of [22], the oscillations of the composite potential functions can be bounded as follows under conditions $(M)^{(m)}$ and (G):

$$\frac{\|G_{i:t}\|}{\eta_i(G_{i:t})} \leq \epsilon_M^{-1}\epsilon_G^{-m}.$$

With this bound, Proposition 1 implies that under the regularity assumptions (G) and $(M)^{(m)}$, the error propagation in the sequential Feynman-Kac filter can be characterized as follows:

$$|[\Phi_{i:t}(\eta_i) - \Phi_{i:t}(\mu_i)](h_t)| \leq \frac{2\left(1-\epsilon_M^2\epsilon_G^{(m-1)}\right)^{\lfloor (t-i)/m \rfloor}}{\epsilon_M \epsilon_G^m}|(\eta_i - \nu_i)(h_i)|. \quad (12)$$

For a set of N particles $\{\xi_i^k\} \in E_i^N$, $i = 1, \ldots, N$ and a function $h_i \in \mathcal{B}_b(E_i)$, such that $u_k \leq h_i(\xi_i^k) \leq v_k$, define

$$\sigma^2(h_i) \stackrel{\text{def}}{=} \frac{1}{N}\sum_{k=1}^{N}(v_k - u_k)^2.$$

Clearly, $\sigma(h_i) \leq \operatorname{osc}(h_i)$. The following result from [27] bounds the weak-sense L_p error induced by the sampling operator for functions with finite oscillations.

Lemma 3 ([27]): Suppose $P \in \mathcal{P}(E)$, then for any $p \geq 1$ and an $\mathcal{E}$-measurable function h with finite oscillations we have

$$\mathbb{E}\{|[P - S^N(P)](h)|^p\}^{\frac{1}{p}} \leq c(p)^{\frac{1}{p}}\frac{\sigma(h)}{\sqrt{N}},$$

where $c(p)$ is defined as follows:

$$c(p) = \begin{cases} 1 & \text{if } p = 1 \\ 2^{-p/2}p\Gamma(p/2) & \text{if } p > 1 \end{cases}$$

and $\Gamma(\cdot)$ is the Gamma function.

Lemma 4: Let Condition (G) and Assumption 1 hold, and let $h_i : E_i \to \mathbb{R}$ be a measurable function such that $\|h_i\|_\infty \leq 1$. Then, $(\widehat{\Phi}_i(\widehat{\eta}_{i-1}^N) - \Phi_i(\widehat{\eta}_{i-1}^N))(h_i) \leq \delta|\log\epsilon_G|$.

The lemma is similar to one presented in [36], and the proof uses a similar argument.

Proof: It follows from Assumption 1 and Condition (G) that for all i,

$$\sup_{x_i \in E_i}|\log G_i(x_i) - \log\widehat{G}_i(x_i)| \leq \delta\sup_{x_i\in E_i}|\log(G_i(x_i))| = \delta|\log\epsilon_G|. \quad (13)$$

For $h_i \in \mathcal{B}(E_i)$, let μ_{h_i} denote the conditional expectation of $h_i(x_i)$ given x_{i-1},

$$\mu_{h_i}(x_{i-1}) \stackrel{\text{def}}{=} \int_{E_i} h_i(x_i)M_i(x_{i-1}, \mathrm{d}x_i). \quad (14)$$

Since $\|h_i\|_\infty \leq 1$ and $M_i(x_{i-1}, \mathrm{d}x_i)$ is a Markov kernel, $\|\mu_{h_i}\|_\infty \leq 1$. For any $\lambda \in [0, 1]$, define the measure

$$\phi_\lambda(\mathrm{d}x_i) \stackrel{\text{def}}{=} \int_{E_{i-1}}\frac{1}{\widehat{\eta}_{i-1}^N(G_\lambda)}G_\lambda(x_{i-1})\widehat{\eta}_{i-1}^N(\mathrm{d}x_{i-1})M_i(x_{i-1}, \mathrm{d}x_i), \quad (15)$$

where

$$\log G_\lambda(x) = (1-\lambda)\log G_{i-1}(x) + \lambda\log\widehat{G}_{i-1}(x).$$

Then $\phi_0 = \Phi_i(\widehat{\eta}_{i-1}^N)$ and $\phi_1 = \widehat{\Phi}_i(\widehat{\eta}_{i-1}^N)$. Furthermore, define the probability measure π_λ on E_{i-1} as

$$\pi_\lambda(\mathrm{d}x_{i-1}) \stackrel{\text{def}}{=} \frac{G_\lambda(x_{i-1})\widehat{\eta}_{i-1}^N(\mathrm{d}x_{i-1})}{\widehat{\eta}_{i-1}^N(G_\lambda)}. \quad (16)$$

Let $\Delta g(x) = \log\widehat{G}_{i-1}(x) - \log G_{i-1}(x)$. Then, we can write

$$\log G_\lambda(x) = \log G_{i-1}(x) + \lambda\Delta g(x). \quad (17)$$

Differentiating both sides of (17) with respect to λ, and rearranging, we obtain

$$\frac{\mathrm{d}G_\lambda(x)}{\mathrm{d}\lambda} = \Delta g(x)G_\lambda(x). \quad (18)$$

Using (16) and (18), for any f defined over E_{i-1} we have,

$$\begin{aligned}\frac{\mathrm{d}\pi_\lambda(f)}{\mathrm{d}\lambda} &= \frac{\widehat{\eta}_{i-1}^N(G_\lambda\Delta g f)\widehat{\eta}_{i-1}^N(G_\lambda) - \widehat{\eta}_{i-1}^N(G_\lambda f)\widehat{\eta}_{i-1}^N(G_\lambda\Delta g)}{\{\widehat{\eta}_{i-1}^N(G_\lambda)\}^2} \\ &= \pi_\lambda(\Delta g f) - \pi_\lambda(f)\pi_\lambda(\Delta g) \\ &= \pi_\lambda(\{\Delta g - \pi_\lambda(\Delta g)\}f).\end{aligned}$$

Substituting $f(\cdot) = M_i(\cdot, \mathrm{d}x_i)$, we have $\frac{\mathrm{d}\pi_\lambda(M_i(\cdot,\mathrm{d}x_i))}{\mathrm{d}\lambda} = \pi_\lambda(\{\Delta g - \pi_\lambda(\Delta g)\}M_i(\cdot, \mathrm{d}x_i))$. Recalling the definition of the action of a measure $\pi_\lambda(\cdot)$ on a function $M(\cdot, \mathrm{d}x_i)$ introduced in Sec. I-B and using this relationship gives

$$\begin{aligned}\frac{\mathrm{d}\phi_\lambda(h_i)}{\mathrm{d}\lambda} &= \frac{\mathrm{d}}{\mathrm{d}\lambda}\int_{E_i}h_i(x_i)\int_{E_{i-1}}\pi_\lambda(\mathrm{d}x_{i-1})M_i(x_{i-1}, \mathrm{d}x_i) \\ &= \int_{E_i}h_i(x_i)\frac{\mathrm{d}}{\mathrm{d}\lambda}\pi_\lambda(M_i(\cdot, \mathrm{d}x_i)) \\ &= \int_{E_i}h_i(x_i)\pi_\lambda(\{\Delta g - \pi_\lambda(\Delta g)\}M_i(\cdot, \mathrm{d}x_i)).\end{aligned}$$

Using linearity of measures and rearranging the order of integration (which is justified via Fubini's theorem, since all the measures considered here are finite), we obtain

$$\begin{aligned}\frac{\mathrm{d}\phi_\lambda(h_i)}{\mathrm{d}\lambda} &= \int_{E_i}h_i(x_i)\int_{E_{i-1}}\pi_\lambda(\mathrm{d}x_{i-1})\Delta g(x_{i-1})M_i(x_{i-1}, \mathrm{d}x_i) \\ &\quad - \int_{E_i}h_i(x_i)\int_{E_{i-1}}\pi_\lambda(\mathrm{d}x_{i-1}) \\ &\quad \times \int_{E_{i-1}}\pi_\lambda(\mathrm{d}x_{i-1})\Delta g(x_{i-1})M_i(x_{i-1}, \mathrm{d}x_i)\end{aligned}$$

$$
\begin{aligned}
&= \int_{E_{i-1}} \pi_\lambda(\mathrm{d}x_{i-1})\Delta g(x_{i-1}) \int_{E_i} h_i(x_i) M_i(x_{i-1}, \mathrm{d}x_i) \\
&\quad - \int_{E_{i-1}} \pi_\lambda(\mathrm{d}x_{i-1}) \int_{E_{i-1}} \pi_\lambda(\mathrm{d}x_{i-1})\Delta g(x_{i-1}) \\
&\quad \times \int_{E_i} h_i(x_i) M_i(x_{i-1}, \mathrm{d}x_i) \\
&= \pi_\lambda(\{\Delta g - \pi_\lambda(\Delta g)\}\mu_{h_i}),
\end{aligned}
$$

where the last equality uses the definition of μ_{h_i}.

We would like to bound $\phi_1(h_i) - \phi_0(h_i) = \int_0^1 \mathrm{d}\phi_\lambda(h_i)$. Since $\|\mu_{h_i}\|_\infty \leq 1$, applying Hölder's inequality and using that $L^1(E_{i-1}, \pi_\lambda) \subset L^2(E_{i-1}, \pi_\lambda)$ gives

$$
\begin{aligned}
\phi_1(h_i) - \phi_0(h_i) &= \int_0^1 \pi_\lambda(\{\Delta g - \pi_\lambda(\Delta g)\}\mu_{h_i})\mathrm{d}\lambda \\
&\leq \int_0^1 \pi_\lambda(|\Delta g - \pi_\lambda(\Delta g)|)\mathrm{d}\lambda \\
&\leq \int_0^1 \{\pi_\lambda\left(|\Delta g - \pi_\lambda(\Delta g)|^2\right)\}^{1/2}\mathrm{d}\lambda.
\end{aligned}
$$

Moreover, since the second moment is an upper-bound for the variance and Δg is bounded [see (13)], we get

$$
\begin{aligned}
\phi_1(h_i) - \phi_0(h_i) &\leq \int_0^1 \{\pi_\lambda\left(|\Delta g - \pi_\lambda(\Delta g))|^2\right)\}^{1/2}\mathrm{d}\lambda \\
&\leq \int_0^1 \{\pi_\lambda\left(|\Delta g|^2\right)\}^{1/2}\mathrm{d}\lambda \\
&\leq \int_0^1 \|\Delta g\|_\infty \mathrm{d}\lambda \\
&= \int_0^1 \|\log \widehat{G}_{i-1} - \log G_{i-1}\|_\infty \mathrm{d}\lambda \\
&\leq \delta|\log \epsilon_G|,
\end{aligned}
$$

where the last step follows from (13). ■

B. Proof of Theorem 1

We observe that $\Phi_0(\widehat{\eta}_{-1}^N) = \eta_0$. Furthermore, for all $i \in \{0, 1, \ldots, t\}$, $\Phi_{i:t}(\Phi_i(\widehat{\eta}_{i-1}^N)) = \Phi_{i-1:t}(\widehat{\eta}_{i-1}^N)$.

Then $\widehat{\eta}_t^N - \eta_t$ can be expressed as a telescopic sum of operators as below:

$$
\begin{aligned}
\widehat{\eta}_t^N - \eta_t &= \Phi_{t:t}(\widehat{\eta}_t^N) - \Phi_{0:t}(\Phi_0(\widehat{\eta}_{t-1}^N)) \\
&= \sum_{i=0}^t [\Phi_{i:t}(\widehat{\eta}_i^N) - \Phi_{i:t}(\Phi_i(\widehat{\eta}_{i-1}^N))] \\
&= \sum_{i=0}^t [\Phi_{i:t}(\widehat{\eta}_i^N) - \Phi_{i:t}(\widehat{\Phi}_i(\widehat{\eta}_{i-1}^N))] \\
&\quad + \sum_{i=0}^t [\Phi_{i:t}(\widehat{\Phi}_i(\widehat{\eta}_{i-1}^N)) - \Phi_{i:t}(\Phi_i(\widehat{\eta}_{i-1}^N))]. \qquad (19)
\end{aligned}
$$

By applying Minkowski's inequality to (19) for $t \geq 0, p \geq 1$, and $\|h_t\|_\infty \leq 1$, we obtain

$$
\begin{aligned}
&\mathbb{E}\left\{|[\widehat{\eta}_t^N - \eta_t](h_t)|^p\right\}^{1/p} \\
&\leq \sum_{i=0}^t \mathbb{E}\{\left|[\Phi_{i:t}(\widehat{\eta}_i^N) - \Phi_{i:t}(\widehat{\Phi}_i(\widehat{\eta}_{i-1}^N))](h_t)\right|^p\}^{1/p} \\
&+ \sum_{i=0}^t \mathbb{E}\{\left|[\Phi_{i:t}(\widehat{\Phi}_i(\widehat{\eta}_{i-1}^N)) - \Phi_{i:t}(\Phi_i(\widehat{\eta}_{i-1}^N))](h_t)\right|^p\}^{1/p}. \qquad (20)
\end{aligned}
$$

In the right-hand side of the above inequality, the first sum corresponds to the error introduced due to sampling (cf., equation (5), whereas the second corresponds to that due to the use of an approximate potential function. To prove the theorem, we shall now obtain bounds for the two sums appearing in (20). For the first sum, applying Proposition 1, (12), and Lemma 3 we have

$$
\begin{aligned}
&\sum_{i=0}^t \left[\mathbb{E}\{\left|[\Phi_{i:t}(\widehat{\eta}_i^N) - \Phi_{i:t}(\widehat{\Phi}_i(\widehat{\eta}_{i-1}^N))](h_t)\right|^p\}^{1/p}\right] \\
&\quad \leq \sum_{i=0}^t \left[\frac{2\left(1 - \epsilon_M^2 \epsilon_G^{(m-1)}\right)^{\lfloor (t-i)/m \rfloor}}{\epsilon_M \epsilon_G^m} \times \mathbb{E}\{|(\widehat{\eta}_i^N - \widehat{\Phi}_i(\widehat{\eta}_{i-1}^N))(h_i)|^p\}^{1/p}\right] \qquad (21) \\
&= \sum_{i=0}^t \left[\frac{2\left(1 - \epsilon_M^2 \epsilon_G^{(m-1)}\right)^{\lfloor (t-i)/m \rfloor}}{\epsilon_M \epsilon_G^m} \times \mathbb{E}\{|(S^N(\widehat{\Phi}_i(\widehat{\eta}_{i-1}^N)) - \widehat{\Phi}_i(\widehat{\eta}_{i-1}^N))(h_i)|^p\}^{1/p}\right] \qquad (22) \\
&\leq \sum_{i=0}^t \left[\frac{2\left(1 - \epsilon_M^2 \epsilon_G^{(m-1)}\right)^{\lfloor (t-i)/m \rfloor}}{\epsilon_M \epsilon_G^m} c(p)^{\frac{1}{p}} \frac{\sigma(h_i)}{\sqrt{N}}\right] \qquad (23) \\
&\leq \frac{4m}{\epsilon_M^3 \epsilon_G^{(2m-1)}} \frac{c(p)^{\frac{1}{p}}}{\sqrt{N}}, \qquad (24)
\end{aligned}
$$

where the last inequality is obtained by noting that when $\|h_i\|_\infty \leq 1$, $\sigma(h_i) \leq \mathrm{osc}(h_i) \leq 2$, and moreover

$$
\sum_{i=0}^t \left(1 - \epsilon_M^2 \epsilon_G^{(m-1)}\right)^{\lfloor (t-i)/m \rfloor} \leq \frac{m}{\epsilon_M^2 \epsilon_G^{(m-1)}}. \qquad (25)
$$

Next, using Proposition 1, (12) and Lemma 4 on the second sum in (20), we have

$$\sum_{i=0}^{t} \mathbb{E}\left\{\left|[\Phi_{i,t}(\widehat{\Phi}_i(\widehat{\eta}_{i-1}^N)) - \Phi_{i,t}(\Phi_i(\widehat{\eta}_{i-1}^N))](h_t)\right|^p\right\}^{1/p}$$

$$\le \sum_{i=0}^{t} \left[\frac{2\left(1-\epsilon_M^2 \epsilon_G^{(m-1)}\right)^{\lfloor (t-i)/m \rfloor}}{\epsilon_M \epsilon_G^m} \times \mathbb{E}\{|(\widehat{\Phi}_i(\widehat{\eta}_{i-1}^N) - \Phi_i(\widehat{\eta}_{i-1}^N))(h_i)|^p\}^{1/p} \right] \quad (26)$$

$$\le \frac{2m}{\epsilon_M^3 \epsilon_G^{(2m-1)}} \delta |\log \epsilon_G|. \quad (27)$$

Finally, combining (24) and (27) in (20), we get

$$\mathbb{E}\left\{|[\widehat{\eta}_t^N - \eta_t](h_t)|^p\right\}^{1/p} \le \frac{2m}{\epsilon_M^3 \epsilon_G^{(2m-1)}} \left(\frac{2c(p)^{\frac{1}{p}}}{\sqrt{N}} + \delta |\log \epsilon_G| \right), \quad (28)$$

For a centralized particle filter, where the effect of approximation of the potential function due to gossip is absent, the global approximation error between the true filtering distribution and its N-particle approximation, $\eta_t^N - \eta_t$, can be related to the sequence of local approximation errors $\eta_i^N - \Phi_i(\eta_{i-1}^N), i = 0, \ldots, t$, [22, equation 7.24]:

$$\eta_t^N - \eta_t = \sum_{i=0}^{t} \left[\Phi_{i:t}(\eta_i^N) - \Phi_{i:t}(\Phi_i(\eta_{i-1}^N)) \right]. \quad (29)$$

The corresponding bound (9) for the propagated error can be derived using equations (12), (29) and Lemma 3, following the same methodology used to obtain (24).

C. Proof of Theorem 2

Using the triangle inequality in (19), we obtain

$$|[\widehat{\eta}_t^N - \eta_t](h_t)| \le Z_1 + Z_2, \quad (30)$$

where

$$Z_1 = \left| \sum_{i=0}^{t} [\Phi_{i,t}(\widehat{\eta}_i^N) - \Phi_{i,t}(\widehat{\Phi}_i(\widehat{\eta}_{i-1}^N))](h_t) \right|,$$

$$Z_2 = \left| \sum_{i=0}^{t} [\Phi_{i,t}(\widehat{\Phi}_i(\widehat{\eta}_{i-1}^N)) - \Phi_{i,t}(\Phi_i(\widehat{\eta}_{i-1}^N))](h_t) \right|.$$

Let $(1+\epsilon)\epsilon_0 \delta |\log \epsilon_G| = \kappa$. Then, $\mathbb{P}\{|[\widehat{\eta}_t^N - \eta_t](h_t)| \ge \kappa\}$ can be expressed as

$$\mathbb{P}\{|[\widehat{\eta}_t^N - \eta_t](h_t)| \ge \kappa\} \le \mathbb{P}\{Z_1 + Z_2 \ge \kappa\}, \quad (31)$$

For a pair of random variables Z_1, Z_2, if $Z_1 + Z_2 \ge \kappa$ then for any choice of κ_1, κ_2, such that $\kappa_1 + \kappa_2 = \kappa$, either $Z_1 \ge \kappa_1$ or $Z_2 \ge \kappa_2$, and therefore

$$\sup_{t\ge 0} \mathbb{P}\{Z_1 + Z_2 \ge \kappa\} \le \sup_{t\ge 0} \mathbb{P}\{Z_1 \ge \kappa_1\} + \sup_{t\ge 0} \mathbb{P}\{Z_2 \ge \kappa_2\}. \quad (32)$$

It follows from Markov's inequality, that if $\psi(\cdot)$ is a strictly monotonically increasing non-negative valued function, then for any random variable Z, and real number κ, we have

$$\mathbb{P}\{Z \ge \kappa\} \le \frac{\mathbb{E}\{\psi(Z)\}}{\psi(\kappa)}. \quad (33)$$

Let $\tau_1 > 0$, $\tau_2 > 0$. Following Chernoff's method, choosing $\psi_1(Z_1) = e^{\tau_1 Z_1}$ and $\psi_2(Z_2) = e^{\tau_2 Z_2}$, from (32), we obtain

$$\sup_{t\ge 0} \mathbb{P}\{|[\widehat{\eta}_t^N - \eta_t](h_t)| \ge \kappa\} \le \sup_{t\ge 0} e^{-\tau_1\kappa_1} \mathbb{E}\{e^{\tau_1 Z_1}\} + \sup_{t\ge 0} e^{-\tau_2\kappa_2} \mathbb{E}\{e^{\tau_2 Z_2}\}.$$

Let us now consider the exponential series expansion

$$\mathbb{E}e^{\tau_1} Z_1 = 1 + \sum_{n\ge 1} \frac{\tau_1^n}{n!} \mathbb{E}Z_1^n. \quad (34)$$

Observe that, using Minkowski's inequality and (24), for any $n \ge 1$, we have

$$\mathbb{E}\{Z_1^n\}^{1/n} = \mathbb{E}\left\{ \left| \sum_{i=0}^{t} [\Phi_{i,t}(\widehat{\eta}_i^N) - \Phi_{i,t}(\widehat{\Phi}_i(\widehat{\eta}_{i-1}^N))](h_t) \right|^n \right\}^{1/n} \quad (35)$$

$$\le \sum_{i=0}^{t} \mathbb{E}\left\{ \left| [\Phi_{i,t}(\widehat{\eta}_i^N) - \Phi_{i,t}(\widehat{\Phi}_i(\widehat{\eta}_{i-1}^N))](h_t) \right|^n \right\}^{1/n} \quad (36)$$

$$\le \frac{4m}{\epsilon_M^3 \epsilon_G^{(2m-1)}} \frac{c(n)^{\frac{1}{n}}}{\sqrt{N}}, \quad (37)$$

and therefore,

$$\mathbb{E}\{Z_1^n\} \le 2^n \left(\frac{2m}{\epsilon_M^3 \epsilon_G^{(2m-1)}} \right)^n \frac{c(n)}{N^{n/2}}. \quad (38)$$

Using (34), noting that $\frac{2m}{\epsilon_M^3 \epsilon_G^{(2m-1)}} = \epsilon_0$, and recalling the definition of $c(n)$ in Lemma 3, we get

$$e^{-\tau_1\kappa_1} \mathbb{E}\{e^{\tau_1 Z_1}\} \le e^{-\tau_1\kappa_1} \left(1 + \sum_{n\ge 1} \frac{\tau_1^n}{n!} \frac{(2\epsilon_0)^n c(n)}{N^{n/2}} \right) \quad (39)$$

$$\le e^{-\tau_1\kappa_1} \left(1 + \frac{2\epsilon_0\tau_1}{\sqrt{N}} + \sum_{n\ge 2} \left[\frac{2\epsilon_0\tau_1}{\sqrt{2N}} \right]^n \frac{\Gamma(n/2)}{(n-1)!} \right). \quad (40)$$

Using a power series expansion of the error function $\mathrm{Erf}(\cdot)$, we obtain the following identity for any real α

$$\sum_{n\ge 1} \alpha^n \frac{\Gamma(n/2)}{(n-1)!} = \alpha\sqrt{\pi} e^{\alpha^2/4} \left[1 + \mathrm{Erf}\left(\frac{\alpha}{2}\right) \right]. \quad (41)$$

Then,

$$e^{-\tau_1\kappa_1}\left(1+\frac{2\epsilon_0\tau_1}{\sqrt{N}}+\sum_{n\geq 2}\left[\frac{2\epsilon_0\tau_1}{\sqrt{2N}}\right]^n\frac{\Gamma(n/2)}{(n-1)!}\right)$$
$$=e^{-\tau_1\kappa_1}\left(1+\frac{2\epsilon_0\tau_1}{\sqrt{N}}-\frac{2\epsilon_0\tau_1\sqrt{\pi}}{\sqrt{2N}}+\sum_{n\geq 1}\left[\frac{2\epsilon_0\tau_1}{\sqrt{2N}}\right]^n\frac{\Gamma(n/2)}{(n-1)!}\right) \quad (42)$$
$$=e^{-\tau_1\kappa_1}\left(1+\frac{2\epsilon_0\tau_1}{\sqrt{N}}-\frac{2\epsilon_0\tau_1\sqrt{\pi}}{\sqrt{2N}}+\frac{2\epsilon_0\tau_1\sqrt{\pi}}{\sqrt{2N}}e^{\epsilon_0^2\tau_1^2/(2N)}\left[1+\mathrm{Erf}\left(\frac{\epsilon_0\tau_1}{\sqrt{2N}}\right)\right]\right). \quad (43)$$

Finally, noting that $\sup_x \mathrm{Erf}(x)=1$ and $1-\sqrt{\pi/2}<0$, from (42) we obtain

$$e^{-\tau_1\kappa_1}\mathbb{E}\{e^{\tau_1 Z_1}\}\leq\left(1+2\sqrt{2\pi}\frac{\epsilon_0\tau_1}{\sqrt{N}}\right)e^{\epsilon_0^2\tau_1^2/(2N)-\kappa_1\tau_1}. \quad (44)$$

Applying Markov's inequality for the second term in (32), we obtain

$$\mathbb{P}\{Z_2\geq\kappa_2\}\leq\sup_{t\geq 0}e^{-\tau_2\kappa_2}\mathbb{E}\{e^{\tau_2 Z_2}\} \quad (45)$$
$$=\sup_{t\geq 0}e^{-\tau_2\kappa_2}\left(1+\sum_{n\geq 1}\frac{\tau_2^n}{n!}\mathbb{E}\{Z_2^n\}\right) \quad (46)$$
$$\leq\sup_{t\geq 0}e^{-\tau_2\kappa_2}\left(1+\sum_{n\geq 1}\frac{\tau_2^n}{n!}\left(\epsilon_0\delta|\log\epsilon_G|\right)^n\right) \quad (47)$$
$$=e^{-\tau_2\kappa_2}e^{\tau_2\epsilon_0\delta|\log\epsilon_G|} \quad (48)$$
$$=e^{\tau_2(\epsilon_0\delta|\log\epsilon_G|-\kappa_2)}. \quad (49)$$

The inequalities in (44) and (45) hold for any $\kappa_1>0$, $\kappa_2>0$ such that $\kappa_1+\kappa_2=\kappa=(1+\epsilon)\epsilon_0\delta|\log\epsilon_G|$. We choose $\kappa_1=\epsilon_0\delta|\log\epsilon_G|\epsilon/2$ and $\kappa_2=\epsilon_0\delta|\log\epsilon_G|(1+\epsilon/2)$. The proof is completed by choosing $\tau_1=\frac{\kappa_1 N}{\epsilon_0^2}$ and letting $\tau_2\to\infty$. For these choices of κ_1, κ_2, τ_1 and τ_2, the bound in (44) becomes $\left(1+\sqrt{2\pi N}\delta\log|\epsilon_G|\epsilon\right)e^{-\frac{(\delta|\log\epsilon_G|)^2\epsilon^2 N}{8}}$, and the bound in (45) becomes 0.

The value of τ_1 is chosen to obtain a simple expression for the bound. The bound obtained above can be made tighter by optimally choosing the parameters κ_1, κ_2 and τ_1.

Note that Z_2 quantifies the error due to the mismatch between the true and approximate log-likelihood functions. As this mismatch is bounded by Assumption 1, the moments of Z_2 are bounded by a geometric series, as shown in (26); and consequently, the probability $\mathbb{P}\{Z_2\geq\kappa_2\}$ goes to zero when κ_2 exceeds the threshold $\epsilon_0\delta|\log\epsilon_G|$. The term Z_1, on the other hand, represents the impact of sampling, and approaches 0 as N goes to infinity. However, it is also worth noting that splitting the error term into Z_1 and Z_2 does not decouple the impact of sampling and that of using an approximate log-likelihood function, as the final bound is still dependent on the parameter of distortion δ.

The condition $\sup h\leq 1$ is specifically required when we apply Proposition 1 from [22]. It may be possible to relax condition (h) to $\sup h\leq K_h<\infty$ where K_h is any real number, in which case the bound established in Lemma 4 would be scaled by a factor of K_h, and Proposition 1 would have to be extended to bounded functions. The bounds in Theorem 1 and 2 would be modified accordingly.

VI. Numerical Results

In this section, we present results of numerical experiments where the particle filter uses an approximate version of the potential function, obtained through a gossip-based algorithm. These results provide an insight into how the use of an approximate potential function affects the performance of the filter in a practical tracking problem. They also allow us to compare the weak-sense L_p error evaluated during the experiment with its theoretical bound set by Theorem 1.

The following model of state evolution and information acquisition is adopted. The variable of interest is X_t, the position of the target that moves in a two dimensional space $E_t=[-k,k]\times[-k,k]$, $k>0$. The target starts at $X_1=[0,0]^\top$ and the state of X_t evolves dynamically according to the following equation:

$$X_t=X_{t-1}+\Delta+v_t,$$

where Δ is a deterministic step taken towards a fixed direction by the target at every time instant, and $v_t\sim\mathcal{N}(0,\Sigma_1)$ is a process noise that distorts the otherwise linear trajectory of the target, with covariance matrix $\Sigma_1=\sigma_1^2 I$. To ensure that the target remains within the space E_t, the simulation scheme is devised such that X_t is reflected back whenever it hits the boundaries of E_t. For the present simulations, the following values are used $k=100$, $\Delta=[0.5,0.5]^\top$, and $\sigma_1^2=0.64$. The dynamics of X_t are simulated over $T=50$ time steps.

The observations Y_t are assumed to be a linear function HX_t of X_t ($H\in\mathbb{R}^{2\times 2}$), contaminated with noise. For the simulations, we choose $H=[1,2;2,1]$.

We recall that for Theorem 1 to hold, the potential function $G_t(x_t)\propto p_t(y_t|x_t)$ must satisfy Condition (G). Here we assume the constant of proportionality to be 1. The bounds on $G_t(x_t)$ impose additional conditions on the measurement equation which are implemented as follows. For all t, let the range space of Y_t be given by $F_t=[-3k,3k]\times[-3k,3k]$. This is the range of Y_t when $Y_t=HX_t$ and $X_t\in E_t$. The likelihood function $p_t(y_t|x_t)$ is taken to be a mixture of a Gaussian distribution and a uniform distribution: for all $t\geq 0$,

$$p_t(y_t|x_t)=C_N f_N(y_t,Hx_t,\Sigma_2)+C_U f_U(F_t), \quad (50)$$

where $f_N(Hx_t,\Sigma_2)$ is the density function of a Gaussian random variable with mean Hx_t and covariance matrix Σ_2, while $f_U(F_t)$ is that of a uniformly distributed random variable defined over F_t. Then $f_U(F_t) = \frac{1}{(6k)^2}$ and $f_N(Hx_t, \Sigma_2) = \frac{1}{2\pi|\Sigma_2|^{1/2}} e^{\left((y_t - Hx_t)^\top \Sigma^{-1}(y_t - Hx_t)\right)}$.[1] Let $\Sigma_2 = \sigma_2^2 I$.

To satisfy the conditions on the extreme values of $p_t(y_t|x_t)$ given by Condition (G), it is required that

$$C_N \frac{1}{2\pi\sigma_2^2} + C_U \frac{1}{(6k)^2} = 1, \quad C_U \frac{1}{(6k)^2} = \epsilon_G$$

and finally, to ensure that $p_t(y_t|x_t)$ is a valid probability density function, we require that for all $x_t \in E_t$, $C_N + C_U = 1$.

The values satisfying the constraints above are given by

$$C_U = \epsilon_G (6k)^2, \quad C_N = 1 - \epsilon_G (6k)^2,$$
$$\text{and } \sigma_2^2 = \frac{1}{2\pi}\left(\frac{1 - \epsilon_G (6k)^2}{1 - \epsilon_G}\right).$$

For the simulations, we set $\epsilon_G = \frac{0.1}{(6k)^2}$, which ensures that on average, 10% of the observations are generated from the uniform part of the distribution and the rest from the Gaussian part.

For the chosen value of $k = 100$, the above parameters have the following values: $\epsilon_G = 2.7778 \times 10^{-7}$, $C_N = 0.9$, $C_U = 0.1$, and $\sigma_2^2 = 0.1432$. The measurements are generated from the density given by (50) at each time step t, for $t \in \{1, \ldots, T\}$.

The above model is carefully constructed so that the performance of the filter can be analyzed in detail in the light of Theorem 1, while having control over the parameter ϵ_G. We note that our results from Section III remain valid for more general settings as well.

The target trajectory is estimated using a centralized bootstrap particle filter, with an approximate version of the potential function being used in lieu of its true value, to simulate the performance of a consensus-based distributed particle filter. A scenario involving N particle trajectories is considered. Recall that ξ_t^k is the state of the k-th particle at time t, for $k = 1, \ldots, N$. The "correct" potential function corresponding to the k-th particle at time t is given by $G_t(\xi_t^k) = p_t(y_t|\xi_t^k)$.

In the bootstrap particle filter algorithm, the potential function associated with a particle indicates its weight at the resampling step. An increase in the value of the potential function will increase the probability of a particle to be sampled, and vice versa. For the model discussed in this paper, $\widehat{G}_t(\xi_t^k)$ can be between $\left(G_t(\xi_t^k)\right)^{1+\delta}$ and $\left(G_t(\xi_t^k)\right)^{1-\delta}$, with $\delta > 0$, which means that the numerical value of $\widehat{G}_t(\xi_t^k)$ may be greater or less than the corresponding true potential function $G_t(\xi_t^k)$. Thus, for a given value of δ, the performance of the filter will depend on exactly *how* the potential function is approximated for the individual particle trajectories.

The weak-sense L_p error will approach the bound given by Theorem 1 only under a worst case scenario. In order to compare the errors observed in the numerical experiments to the error bound of Theorem 1, we therefore artificially construct a setting where filter estimation is deliberately *deteriorated* as follows. For each individual particle, the distortion of the potential function is defined such that the further a particle ξ_t^k is from x_t the larger is the magnification of the corresponding $\widehat{G}_t(\xi_t^k)$, and the closer it is to x_t, the larger the reduction of the weight. This means that the weights of "poor" trajectories are increased and those of "good" trajectories are reduced. For every ξ_t^k, we define

$$\widehat{G}_t(\xi_t^k) = \left(G_t(\xi_t^k)\right)^{\left(1+\lambda_t^k\delta\right)}, \tag{51}$$

where $\lambda_t^k = \left(1 - 2\frac{\|\xi_t^k - x_t\|_2 - \min_k \|\xi_t^k - x_t\|_2}{\max_k \|\xi_t^k - x_t\|_2 - \min_k \|\xi_t^k - x_t\|_2}\right)$ and $\|\cdot\|_2$ denotes the Euclidean norm. Clearly, thus defined, when $\|\xi_t^k - x_t\|_2 = \min_k \|\xi_t^k - x_t\|_2$ and the particle is closest to the true state at time t, $\widehat{G}_t(\xi_t^k) = \left(G_t(\xi_t^k)\right)^{1+\delta}$ and the weight is *reduced* to the largest extent. On the contrary, when $\|\xi_t^k - x_t\|_2 = \max_k \|\xi_t^k - x_t\|_2$, $\widehat{G}_t(\xi_t^k) = \left(G_t(\xi_t^k)\right)^{1-\delta}$ and the weight is increased to the largest extent. This is one of many possible ways to guarantee that the weights are distorted in such a way as to inflate the weight of the worst particle to the greatest extent and deflate that of the best particle to the greatest extent.

The target is tracked using $N = 500$ particles at each step. We are interested in evaluating the performance of this distorted particle filter in terms of the weak-sense L_p error defined in Theorem 1. The requirement, then, is to look at the difference between $\widehat{\eta}_t(h_t(x_t))$ and $\eta_t(h_t(x_t))$ for a test function h_t. Here, $\widehat{\eta}_t(h_t(x_t))$ is simply the particle filter estimate of $h_t(x_t)$, obtained by averaging the values of $h_t(\xi_t^k)$ over the distribution represented by the particle cloud. The quantity $\eta_t(h_t(x_t))$, on the other hand, is the expected value of $h_t(x_t)$ over the true distribution of x_t. In our simulations, the true distribution is estimated by a robust Kalman filter.

The Kalman filter estimate is obtained by ignoring the uniform distribution present in the measurement dynamics and assuming the measurement noise to be Gaussian, i.e., by assuming $p_t(y_t|x_t) = f_N(y_t, Hx_t, \Sigma_2)$. Whenever the innovation or measurement residual $(y_t - Hx_{t|t-1})$ is higher than a certain threshold, it is inferred that the corresponding observation is generated from the uniform part of the distribution. In those cases the filter ignores the observation and continues with the a priori state estimate. Since ϵ_G is small, this estimation approach does not lead to significant discrepancies.

Sample trajectories of the target, along with those estimated by the simulated distributed particle filter and the robust Kalman filter for a small and large value of δ are presented in Fig. 1. The target trajectory is indicated by the blue line, and the robust Kalman filter estimate is indicated by the red line. The green line represents the estimated trajectory by a distributed particle filter under the worst case scenario described by (51) for $\delta = 0.02$, while the magenta line indicates that for $\delta = 0.3$. It is seen that for the same trajectory, for the smaller value of δ, the performance of the consensus-based distributed particle filter remains stable and close to the approximately optimal solution provided by the robust Kalman filter. For a significantly high value of δ, on the other hand, the distributed particle filter tends to perform poorly. For $\delta = 0.3$, the position estimated by the particle filter is observed to have deviated significantly from the true position of the target. Nonetheless, even under this high

[1]Ideally, the Gaussian part of the distribution should be truncated to ensure that Y_t does not assume a value beyond its domain F_t. However, for this experiment, due to the choice of Δ and σ_1^2, X_t remains sufficiently away from its boundaries, and therefore almost all of the mass of $f_N(y_t, Hx_t, \Sigma_2)$ remains within F_t.

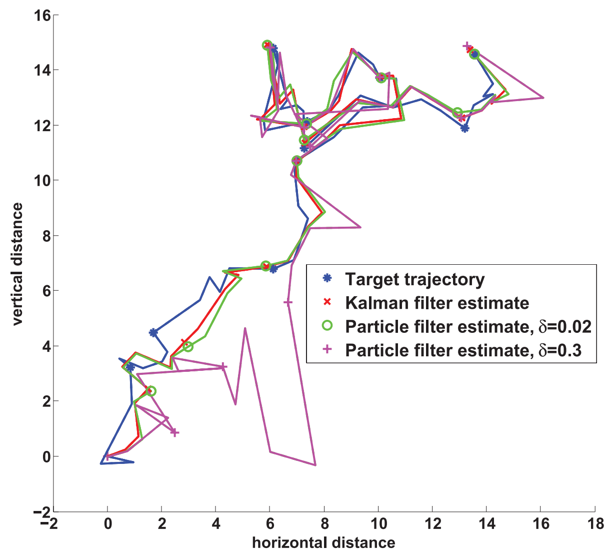

Fig. 1. A sample trajectory of the target and its estimates.

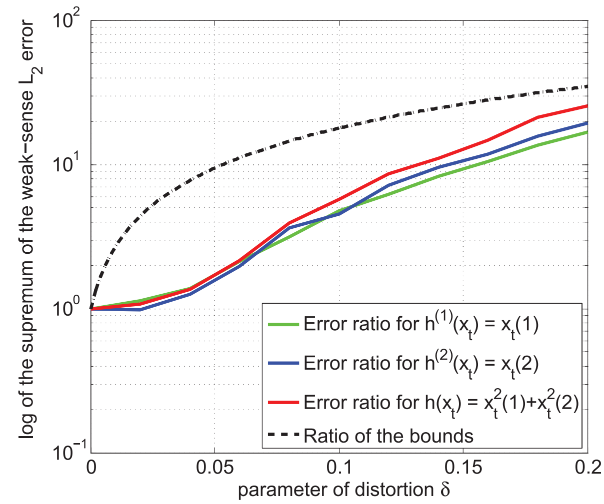

Fig. 2. Ratio of the weak-sense L_2 errors for the distributed and centralized particle filters, for different values of δ.

level of distortion, the filter remains stable, and the estimated trajectory repeatedly returns to the true trajectory of the target.

Recall that for any test function $h_t(X_t)$, Theorem 1 provides a time uniform bound for the following error:

$$\overline{\Xi}_t = \sup_{t=1,\dots,T} \left(\mathbb{E}\left\{|[\widehat{\eta}_t^N - \eta_t](h_t)|^p\right\}\right)^{1/p}.$$

In order to obtain estimates for $\mathbb{E}\left\{|[\widehat{\eta}_t^N - \eta_t](h_t)|^p\right\}^{1/p}$, we generate $M = 1500$ realizations of X_t, and approximate the expected value by averaging over these realizations,

$$\hat{\Xi}_t = \sup_{t=1,\dots,T} \left(\frac{1}{M}\sum_{m=1}^{M}\left\{|[\widehat{\eta}_{t,m}^N - \eta_{t,m}](h_t)|^p\right\}\right)^{1/p}$$

where $\widehat{\eta}_{t,m}^N$, $\eta_{t,m}$ represent the measures corresponding to the estimated distributions for the m-th realization of X_t.

To make an exact comparison of the errors measured in experiment with the expression of error derived in Theorem 1, we need to know the parameter ϵ_M from Condition $(M)^{(m)}$ which is difficult to estimate. Instead, here we analyze the ratio of the upper bound of the weak-sense L_2 error incurred using a distorted version of the potential function, to that corresponding to the case when the true potential function is used ($\delta = 0$). Recall that by Theorem 1, this is given by

$$\epsilon_0\left(\frac{2c(2)^{\frac{1}{2}}}{\sqrt{N}} + \delta|\log\epsilon_G|\right)\Bigg/\left(\frac{2\epsilon_0 c(2)^{\frac{1}{2}}}{\sqrt{N}}\right) = \left(1 + \frac{\sqrt{N}}{2}\delta|\log\epsilon_G|\right).$$

Fig. 2 plots the ratio of the weak-sense L_2 error for the worst case scenario defined in (51) to that for the undistorted particle filter ($\delta = 0$), for each of the test functions:

$$h_t^{(1)}(X_t) = C^{(1)}X_t(1),$$
$$h_t^{(2)}(X_t) = C^{(2)}X_t(2),$$
$$\text{and } h_t(X_t) = C\left(X_t^2(1) + X_t^2(2)\right),$$

where $X_t = [X_t(1)X_t(2)]^\top$ and $C^{(1)}$, $C^{(2)}$ and C are scaling constants chosen to ensure that $h_t^{(1)}$, $h_t^{(2)}$ and h_t satisfy Condition (h), so that the requirements of Theorem 1 are fulfilled. For each of the test functions, the ratio of the errors remain well within that of the corresponding theoretical bounds. For low values of δ ($\delta < 0.04$), the effect of using a distorted version of the potential function remains insignificant, since the uncertainty in observations dominates the error in likelihood evaluation. As δ increases, the impact of distortion becomes more prominent and the ratio of the errors grows monotonically.

In summary, the numerical experiments in this section indicate that when a consensus-based approximation leads to the use of slightly distorted potential function, then as long as the distortion is within a reasonable range, the performance of the distributed filter remains stable and comparable to a centralized filter, even under a worst case scenario. These results complement the theoretical findings of Section III.

VII. Discussion and Related Work

For several decades, the analysis of the error propagation and its role in the stability of non-linear Markov filters has garnered a considerable amount of research interest, and continues to be an active area of research. In [37], Kunita analyzed the asymptotic behaviour of the error and stability of the filter that has an ergodic signal transition semigroup with respect to the initial distribution. The stability of linear filters in the context of a non-Gaussian initial condition was explored in [38], along with an analysis of the stability of non-linear filters in the case where the signal diffusion is convergent. It was shown that a stable signal diffusion with respect to its initial condition is sufficient to guarantee the stability of the optimal filter with respect to the same initial condition. While the results presented in [37] and [38] are important they mainly address an optimal filtering scenario, and not approximate methods like the particle filter.

Several interesting results concerning the stability of particle filters have been developed over the past decade [20]–[25],

[39]. Del Moral [22] presented a detailed analysis of the various properties of particle filters using a Feynman-Kac semigroup approach. The methodology and results developed in the book [22] form the foundation of the analysis presented in our paper. In [20]–[22] Del Moral, Miclo and Guionnet analyzed the stability of particle filters using a general non-linear Feynman-Kac semigroup approach. In [40], uniform convergence results, functional central limit theorems, and exponential concentration estimates were derived for a new algorithm for particle filters. A lognormal central limit theorem for the normalizing constants involved in the approximations associated with particle filters was presented in [41]. Similar convergence results related to various particle filter algorithms were presented in [42], [43] and [44].

Among recent works, in [45], Handel derives a uniform, time-average convergence result for bounded continuous functions, under assumptions of tightness and ergodicity. A uniform convergence result for a class of particle filters is proved by Heine and Crisan in [46] under certain restrictions on the dynamic and measurement models and noise distributions, but without mixing assumptions. Time-average uniform convergence results for distributed particle filters could be derived by incorporating the methodology of [45] to our model. The ideas developed in [46] may provide avenues to relax the regularity assumption $M^{(m)}$ and extend our results to a more general framework, although this may come at the cost of additional assumptions on system dynamics.

Similar to our work, Jasra *et al.* [47] analyze the impact of using an approximating version of the likelihood function, although the nature of approximation is different. In their paper, the likelihood function is estimated through approximate Bayesian computation using auxiliary data, and it is shown that the propagated error remains bounded under certain regularity assumptions. However, the result is derived under a Lipschitz continuity assumption on the likelihood, and unlike our paper, no large deviation results are provided that quantify the tightness of the bound. The ideas developed in [47] could be used to analyze error propagation in other distributed filtering schemes where the approximation is more akin to that considered in [47].

The convergence properties of the asymptotic variance associated with the particle filter have garnered considerable interest in recent years [48]–[50]. Favetto [48] proves results on the tightness of asymptotic variances for bounded test functions, while Douc *et al.* [50] and Whitley [49] derive time-uniform bounds. It would be interesting to incorporate these convergence results into our model of gossip-based distributed particle filters and explore the implications. Using the ideas of [48]–[50], it may be possible to extend our results to quantify the asymptotic variance of distributed particle filters.

Much of the existing literature imposes conditions on the test function h_t while analyzing the propagation of error. Hu, Schön and Ljung [51], [52] have relaxed such conditions. In [51] a bound for the mean L_4 error was derived for unbounded real-valued functions under a general framework that incorporates several standard particle filter algorithms. This was extended to an L_p convergence result for an arbitrary $p \geq 2$ for unbounded functions in [52], although additional restrictions were imposed on the potential function. In [53], employing a method slightly different from those that are standard in kernel density estimation, Crisan and Míguez derive explicit bounds for the supremum of the propagated error, and prove pointwise and almost sure convergence results for unbounded test functions, relying only on the integrability of the test function. Our results may potentially be extended to generalized test functions by combining the analysis presented here with the methods employed in [53].

In this paper, we considered particle filters where resampling is performed at every time step. Alternatively, the implementation of a particle filter may employ *adaptive* resampling, where the resampling step is included not at each time step, but only at those times when a certain criterion is satisfied. In [54], under a regularity condition similar to (G) (6), Del Moral *et al.* derived a non-asymptotic exponential concentration estimate and an exponential coupling theorem associated with a centralized particle filter with adaptive resampling. Using these ideas, the results derived in our paper could be extended to include distributed particle filter algorithms that employ adaptive resampling.

Our results depend on the regularity conditions $(M)^{(m)}$ and (G). While assumptions such as these on the mixing and ergodicity properties of the underlying Markov operator or the potential function are standard in the literature, in some papers the analysis has been carried out under slightly relaxed conditions. In [24], [25], Le Gland and Oudjane studied the stability and convergence rates for particle filters using the Hilbert projective metric. In [24], the signal mixing assumptions were relaxed through the introduction of a specific, robust particle filter architecture with truncated likelihood functions. In [25], the mixing assumption was applied directly to the non-negative kernel that governs the evolution of the particle filter instead of the Markov kernel $M_t(\cdot, \cdot)$ that governs signal diffusion. The kernel introduced includes the effects of both the Markov transitions and the likelihood potentials.

The Dobrushin contraction coefficient plays an important role in the stability analysis presented in this paper, as well as in those developed in many of the references discussed above. Del Moral and Miclo [21] formulated the conditions for the exponential asymptotic stability of the Feynman-Kac semigroup and bounded the Lyapunov constant and Dobrushin coefficient. A time-uniform upper bound on the propagated error in the case of interacting particle systems can be derived using these results.

The error incurred in a particle filter due to sampling depends on the resampling scheme. Additional error could be contributed by other sources. Vaswani *et al.* [39] analyzed the stability of particle filters for the case where the model describing state evolution dynamics is incorrect. Using the same assumptions as in [25], it was shown in their paper, that even when the true Markov transition kernel is different from that used to update the filter, as long as this mismatch persists for a finite number of time-steps, the particle filter remains stable. Our results in this paper complement their work, as we show that the filter remains stable over time, when there is a mismatch between the true *potential function* and that used by the filter, as long as the mismatch is within a certain range.

Most of the convergence results discussed above were derived for centralized particle filters. Distributed particle filters, on the other hand, involve various heuristic algorithms and approximations [28], and therefore, results for the centralized filter are usually not directly applicable to them. Consequently, there are only a few results available on the error propagation in particle filters under a decentralized setting.

Orekshin and Coates [27], [35] analyzed the error propagation and stability of leader node particle filters for distributed tracking in sensor networks. In this setting, one of the nodes, known as the "leader node", performs particle filtering by fusing data recorded by the nodes in its vicinity, known as the "satellite nodes". When the leader node is changed, the communication cost of transmitting the required information is reduced by approximation, either through subsampling or by training a parametric model. The impact of such intermittent approximations was analyzed through the derivation of a time uniform bound on the error and an exponential inequality.

In [26], Míguez considered a distributed particle filter algorithm where the particles are grouped in several disjoint sets, and each set is assigned to a processing element that performs the task of particle filtering. These elements operate independently of each other, except at certain intermittent steps where they exchange subsets of particles and weights. For this model, it was shown that the L_1 estimation error converges uniformly over time. This result, however, is limited only to the specific algorithm considered. The theorems presented in this paper, on the other hand, are applicable to a large class of consensus-based distributed particle filter algorithms.

In consensus-based distributed particle filters, all agents simultaneously participate in filtering and maintain a local approximate particle representation of the global posterior distribution. This is achieved through a decentralized algorithm that establishes consensus among certain global quantities across agents. Depending on the nature of the quantities computed through consensus, these filters can be classified as follows [28]: those computing particle weights through a factorization of the global likelihood function [4]–[7], those computing the posterior distribution through a parametric approximation [8]–[13] and those computing parameters of the global likelihood function [14]–[16]. The results presented in our paper are applicable to the first class of algorithms, where the sensor nodes are synchronized.

VIII. Concluding Remarks

In this paper we have presented analytical results on the error propagation in gossip-based distributed particle filters. In a consensus-based approach, through a finite number of gossip iterations, the true potential function is replaced by an approximated version of itself. Here, we have considered a model where the approximation process distorts the potential function by an exponent that lies between the range $(1-\delta, 1+\delta)$. Under mild assumptions on the system, we have derived a time-uniform upper bound on the expected value of the weak-sense L_p error associated with the filter.

The results indicate that even when a slightly distorted version of the potential function is used, the error remains bounded over time. As long as the parameter δ is small, the error bounds remain close to those corresponding to the centralized filter. A distortion in the potential function can also result from a mismatch in the model used by the filter. Our results are applicable in that scenario as well.

The results presented in this paper are the first of their kind for gossip-driven distributed particle filters. These results were derived under some regularity assumptions on the Markov kernel and the potential function, and under certain restrictions on the function of interest, h_t. As possible extensions of this work, we would like to explore the behaviour of the error propagation under slightly relaxed conditions, and for more general test functions h_t. Ideas presented in some of the recent works [45]–[50], [53] on the convergence of particle filters, discussed in the preceding section, could be incorporated with those presented in this paper to analyze the performance of distributed filters in more detail. It would also be interesting to characterize the stability of distributed particle filters where different consensus algorithms are used. Analyzing more sophisticated gossip-based distributed particle filters within the framework presented in this paper would also be an interesting direction for future work.

Acknowledgments

The authors thank the anonymous reviewers and Dr. Stephane Blouin for feedback which helped to improve the content and presentation presentation of this article.

References

[1] P. Del Moral, "Measure valued processes and interacting particle systems. Application to non linear filtering problems," *Ann. Appl. Probab.*, vol. 8, no. 2, pp. 438–495, May 1996.

[2] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statist. Comput.*, vol. 10, no. 3, pp. 197–208, Jul. 2000.

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

[4] D. Üstebay, M. Coates, and M. Rabbat, "Distributed auxiliary particle filters using selective gossip," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 3296–3299.

[5] S. Farahmand, S. Roumeliotis, and G. Giannakis, "Set-membership constrained particle filter: Distributed adaptation for sensor networks," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4122–4138, Sep. 2011.

[6] C. Bordin and M. Bruno, "Consensus-based distributed particle filtering algorithms for cooperative blind equalization in receiver networks," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 3968–3971.

[7] V. Savic, H. Wymeersch, and S. Zazo, "Belief consensus algorithms for fast distributed target tracking in wireless sensor networks," *Signal Process.*, vol. 95, pp. 149–160, Feb. 2014.

[8] D. Gu, "Distributed particle filter for target tracking," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Rome, Italy, Apr. 2007, pp. 3856–3861.

[9] D. Gu, J. Sun, Z. Hu, and H. Li, "Consensus based distributed particle filter in sensor networks," in *Proc. IEEE Int. Conf. Inf. Autom. (ICIA)*, Zhangjiajie, China, Jun. 2008, pp. 302–307.

[10] D. Gu, "Distributed EM algorithm for Gaussian mixtures in sensor networks," *IEEE Trans. Neural Netw.*, vol. 19, no. 7, pp. 1154–1166, Jul. 2008.

[11] B. Oreshkin and M. Coates, "Asynchronous distributed particle filter via decentralized evaluation of Gaussian products," in *Proc. Int. Conf. Inf. Fusion (Fusion)*, Edinburgh, U.K., Jul. 2010, pp. 1–8.

[12] M. Mohammadi and A. Asif, "Consensus-based distributed unscented particle filter," in *Proc. IEEE Workshop Statist. Signal Process. (SSP)*, Nice, France, Jun. 2011, pp. 237–240.

[13] A. Simonetto, T. Keviczky, and R. Babuska, "Distributed nonlinear estimation for robot localization using weighted consensus," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Anchorage, AK, USA, May 2010, pp. 3026–3031.

[14] O. Slučiak, O. Hlinka, M. Rupp, F. Hlawatsch, and P. Djurić, "Sequential likelihood consensus and its application to distributed particle filtering with reduced communications and latency," in *Proc. Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2011, pp. 1766–1770.

[15] O. Hlinka, O. Slučiak, F. Hlawatsch, P. Djurić, and M. Rupp, "Likelihood consensus and its application to distributed particle filtering," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4334–4349, Aug. 2012.

[16] O. Hlinka, F. Hlawatsch, and P. Djurić, "Likelihood consensus-based distributed particle filtering with distributed proposal density adaptation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 3869–3872.

[17] A. Mohammadi and A. Asif, "Distributed consensus + innovation particle filtering for bearing/range tracking with communication constraints," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 620–635, Feb. 2015.

[18] M. Coates, "Distributed particle filtering for sensor networks," in *Proc. IEEE/ACM Int. Symp. Inf. Process. Sensor Netw. (IPSN)*, Berkeley, CA, USA, Apr. 2004, pp. 99–107.

[19] 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.

[20] P. Del Moral and A. Guionnet, "On the stability of interacting processes with applications to filtering and genetic algorithms," *Ann. l'Inst. H. Poincaré*, vol. 37, no. 2, pp. 155–194, Mar. 2001.

[21] P. Del Moral and L. Miclo, "On the stability of nonlinear Feynman-Kac semigroups," *Ann. Fac. Sci. Toulouse*, vol. 11, no. 2, pp. 135–175, 2002.

[22] P. Del Moral, *Feynman–Kac Formulae: Genealogical and Interacting Particle Approximations*. New York, NY, USA: Springer, 2004.

[23] R. Douc and O. Cappe, "Comparison of resampling schemes for particle filtering," in *Proc. Int. Symp. Image Signal Process. Anal. (ISPA)*, Zagreb, Croatia, Sep. 2005, pp. 64–69.

[24] F. Le Gland and N. Oudjane, "A robustification approach to stability and to uniform particle approximation of nonlinear filters: The example of pseudo-mixing signals," *Stochastic Process. Appl.*, vol. 38, no. 2, pp. 279–316, Aug. 2003.

[25] F. Le Gland and N. Oudjane, "Stability and uniform approximation of nonlinear filters using the Hilbert metric, and application to particle filters," *Ann. Appl. Probab.*, vol. 14, no. 1, pp. 144–187, Feb. 2004.

[26] J. Míguez, "On the uniform asymptotic convergence of a distributed particle filter," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop (SAM)*, A Coruña, Spain, Jun. 2014, pp. 241–244.

[27] B. Oreshkin and M. Coates, "Analysis of error propagation in particle filters with approximation," *Ann. Appl. Probab.*, vol. 21, no. 6, pp. 2343–2378, Dec. 2011.

[28] O. Hlinka, F. Hlawatsch, and P. Djurić, "Distributed particle filtering in agent networks: A survey, classification, and comparison," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 61–81, Jan. 2013.

[29] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. 31, no. 9, pp. 803–812, Sep. 1986.

[30] N. Tsitsiklis, "Problems in decentralized decision making and computation," Ph.D. dissertation, Dept. EECS, MIT, Boston, MA, USA, Nov. 1984.

[31] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Control Lett.*, vol. 53, no. 1, pp. 65–78, Sep. 2004.

[32] A. Dimakis, S. Kar, J. Moura, M. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proc. IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov. 2010.

[33] A. Olshevsky and J. Tsitsiklis, "Convergence speed in distributed consensus and averaging," *SIAM J. Control Optim.*, vol. 48, no. 1, pp. 33–55, Feb. 2009.

[34] A. Mohammadi and A. Asif, "A consensus/fusion based distributed implementation of the particle filter," in *Proc. IEEE Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process. (CAMSAP)*, San Juan, PR, USA, Dec. 2011, pp. 285–288.

[35] B. Oreshkin and M. Coates, "Analysis of error propagation in particle filters with approximation," McGill Univ., Montreal, QC, Canada, Tech. Rep., Aug. 2009 [Online]. Available: http://www.tsp.ece.mcgill.ca/Networks/projects/pdf/oreshkin_TechReport09a.pdf

[36] A. Doucet, "A note on the sensitivity of the posterior distribution to the likelihood function," Dept. Statistics, Univ. Oxford, Oxford, U.K., Tech. Rep., 2014 [Online]. Available: http://www.stats.ox.ac.uk/doucet/doucet_notesensitivityposteriortolikelihood.pdf

[37] H. Kunita, "Asymptotic behavior of the nonlinear filtering errors of Markov processes," *J. Multivariate Anal.*, vol. 1, no. 4, pp. 365–393, Dec. 1971.

[38] D. Ocone and E. Pardoux, "Asymptotic stability of the optimal filter with respect to its initial condition," *SIAM J. Control Optim.*, vol. 34, no. 1, pp. 226–243, Jan. 1996.

[39] N. Vaswani, "Change detection in partially observed nonlinear dynamic systems with unknown change parameters," in *Proc. Amer. Control Conf. (ACC)*, Boston, MA, USA, July 2004, pp. 5387–5393.

[40] P. Del Moral, A. Doucet, and S. Singh, "A backward particle interpretation of Feynman-Kac formulae," *ESAIM Math. Model. Numer. Anal.*, vol. 44, pp. 947–975, Sep. 2010.

[41] J. Bérrard, P. Del Moral, and A. Doucet, "A lognormal central limit theorem for particle approximations of normalizing constants," *arXiv:1307.0181*, 2013.

[42] P. Del Moral and A. Doucet, "Interacting Markov chain Monte Carlo methods for solving nonlinear measure-valued equations," *Ann. Appl. Probab.*, vol. 20, no. 2, pp. 593–639, Apr. 2010.

[43] F. Caron, P. Del Moral, M. Pace, and B.-N. Vo, "On the stability and the approximation of branching distribution flows, with applications to nonlinear multiple target filtering," *Stochastic Anal. Appl.*, vol. 29, no. 6, pp. 951–997, Oct. 2011.

[44] F. Caron, P. Del Moral, A. Doucet, and M. Pace, "Particle approximation of the intensity measures of a spatial branching point process arising in multitarget tracking," *SIAM J. Control Optim.*, vol. 49, no. 4, pp. 1766–1792, Jan. 2011.

[45] R. van Handel, "Uniform time average consistency of Monte Carlo particle filters," *Stochastic Process. Appl.*, vol. 119, no. 11, pp. 3835–3861, Nov. 2009.

[46] K. Heine and D. Crisan, "Uniform approximations of discrete-time filters," *Adv. Appl. Probab.*, vol. 40, no. 4, pp. 979–1001, Dec. 2008.

[47] A. Jasra, S. Singh, J. Martin, and E. McCoy, "Filtering via approximate Bayesian computation," *Statist. Comput.*, vol. 22, no. 6, pp. 1223–1237, 2012.

[48] B. Favetto, "On the asymptotic variance in the central limit theorem for particle filters," *ESAIM Probab. Statist.*, vol. 16, pp. 151–164, Jul. 2012.

[49] N. Whiteley, "Stability properties of some particle filters," *Ann. Appl. Probab.*, vol. 23, no. 6, pp. 2500–2537, Dec. 2013.

[50] R. Douc, E. Moulines, and J. Olsson, "Long-term stability of sequential Monte Carlo methods under verifiable conditions," *Ann. Appl. Probab.*, vol. 24, no. 5, pp. 1767–1802, Oct. 2014.

[51] H. Xiao-Li, T. Schön, and L. Ljung, "A basic convergence result for particle filtering," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1337–1348, Apr. 2008.

[52] H. Xiao-Li, T. Schön, and L. Ljung, "A general convergence result for particle filtering," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3424–3429, Jul. 2011.

[53] D. Crisan and J. Míguez, "Particle-kernel estimation of the filter density in state-space models," *Bernoulli*, vol. 20, no. 4, pp. 1879–1929, Nov. 2014.

[54] P. Del Moral, A. Doucet, and A. Jasra, "On adaptive resampling strategies for sequential Monte Carlo methods," *Bernoulli*, vol. 18, no. 1, pp. 252–278, Feb. 2012.

Syamantak Datta Gupta received the B.E. degree from Jadavpur University, Kolkata, India, and the M.A.Sc. and Ph.D. degrees from the University of Waterloo, Waterloo, Canada, all in electrical engineering, in 2006, 2009, and 2014, respectively. From 2006 to 2007, he was with Siemens Limited where he worked in industrial automation. In 2014, he was a Postdoctoral Fellow with the University of Waterloo, Waterloo, ON, Canada. Since 2015, he has been with the Department of Electrical and Computer Engineering, McGill University, as a Postdoctoral Fellow. His research interests include statistical signal processing, time series analysis, and nonlinear filtering.

Mark Coates (M'99–SM'07) received the B.E. degree in computer systems engineering from the University of Adelaide, Adelaide, SA, Australia, and the Ph.D. degree in information engineering from the University of Cambridge, Cambridge, U.K., in 1995 and 1999, respectively. He joined McGill University, Montreal, QC, Canada, in 2002, where he is currently an Associate Professor of the Department of Electrical and Computer Engineering. From 1999 to 2001, he was a Research Associate and Lecturer with the Rice University, Houston, TX, USA. From 2012 to 2013, he worked as a Senior Scientist with the Winton Capital Management, Oxford, U.K. His research interests include communication and sensor networks, statistical signal processing, and Bayesian and Monte Carlo inference. He was an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2007 to 2011, and a Senior Area Editor for IEEE SIGNAL PROCESSING LETTERS from 2012 to 2015. His most influential and widely cited contributions have been on the topics of network tomography and distributed particle filtering. His contributions on the latter topic received awards at the International Conference on Information Fusion, in 2008 and 2010. In 2006, his research team received the NSERC Synergy Award in recognition of their successful collaboration with Canadian industry, which has resulted in the licensing of software for anomaly detection and video-on-demand network optimization.

Michael Rabbat (S'02–M'07–SM'15) received the B.Sc. degree from the University of Illinois, Urbana-Champaign, Champaign, IL, USA, and the M.Sc. degree from Rice University, Houston, TX, USA, and the Ph.D. degree from the University of Wisconsin, Madison, WI, USA, all in electrical engineering, in 2001, 2003, and 2006. He joined McGill University, Montréal, QC, Canada, in 2007, and he is currently an Associate Professor. From 2013 to 2014, he held Visiting Positions at the Télécom Bretegne, Brest, France, the Inria Bretagne-Atlantique Reserch Centre, Rennes, France, and the KTH Royal Institute of Technology, Stockholm, Sweden. He was a Visiting Researcher at Applied Signal Technology, Inc., Sunnyvale, CA, USA, during the summer of 2003. His research interests include distributed algorithms for optimization and inference, consensus algorithms, and network modelling and analysis, with applications in distributed sensor systems, large-scale machine learning, statistical signal processing, and social networks. He currently serves as Senior Area Editor for the IEEE SIGNAL PROCESSING LETTERS and as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS, and the IEEE TRANSACTIONS ON CONTROL OF NETWORK SYSTEMS. He has coauthored the paper which received the Best Paper Award (Signal Processing and Information Theory Track) at the 2010 IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS) and was the recipient of Honorable Mention for Outstanding Student Paper Award at the 2006 Conference on Neural Information Processing Systems (NIPS), a Best Student Paper Award at the 2004 ACM/IEEE International Symposium on Information Processing in Sensor Networks (IPSN), and the 2014 McGill University Principal's Prize for Excellence in Teaching at the level of Assistant professor.