Analysis of error propagation in particle filters with approximation

Boris N. Oreshkin and Mark J. Coates
Telecommunications and Signal Processing–Computer Networks Laboratory
Department of Electrical and Computer Engineering
McGill University, Montreal, QC, Canada
Email: boris.oreshkin@mail.mcgill.ca, mark.coates@mcgill.ca

Abstract

This paper examines the impact of approximation steps that become necessary when particle filters are implemented on resource-constrained platforms. We consider particle filters that perform intermittent approximation, either by subsampling the particles or by generating a parametric approximation. For such algorithms, we derive time-uniform bounds on the weak-sense L_p error and present associated exponential inequalities. We motivate the theoretical analysis by considering the leader-node particle filter and present numerical experiments exploring its performance and the relationship to the error bounds.

Index Terms

AMS 2000 subject classifications: 62L12,65C35,65L20

Collaborative distributed tracking, particle filtering, error analysis

I. Introduction

Particle filters have proven to be an effective approach for addressing difficult tracking problems [9]. Since they are more computationally demanding and require more memory than most other filtering algorithms, they are really only a valid choice for challenging problems, for which other well-established techniques perform poorly. Such problems involve (approximated) dynamics and/or observation models that are substantially non-linear and non-Gaussian.

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

The analysis of approximation error propagation and stability of non-linear Markov filters has been an active research area for several decades [13], [21]. In the case of the particle filter, there has been interest in establishing what conditions must hold for the filter to remain stable (the error remaining bounded over time), despite the error that is introduced at every time-step of the algorithm by the pointwise approximation of the particle representation [3]–[6], [8], [16], [17]. The impact of modeling errors, including incorrect initial distributions and short-lived model mismatches, has also been investigated [17], [26].

In this paper, we focus on examining the impact of additional intermittent approximation steps which become necessary when particle filters are implemented on resource-constrained platforms. The approximations we consider include subsampling of the particle representation and the generation of parametric mixture models. The main results of the paper are time-uniform bounds on the weak-sense L_p-error induced by the combination of particle sampling error and the additional intermittent approximation error (subsampling or parametric). We employ the Feynman-Kac semigroup analysis methodology described in [4]; our investigation of parametric approximation is founded on error bounds for the greedy likelihood maximization algorithm, which was developed in [19] and analyzed in [23], [28].

This research was supported by the National Scientific and Engineering Research Council of Canada (NSERC) through the Discovery Grants program and the MITACS (Mathematics in Information Technology and Complex Systems) Networked Centres of Excellence.

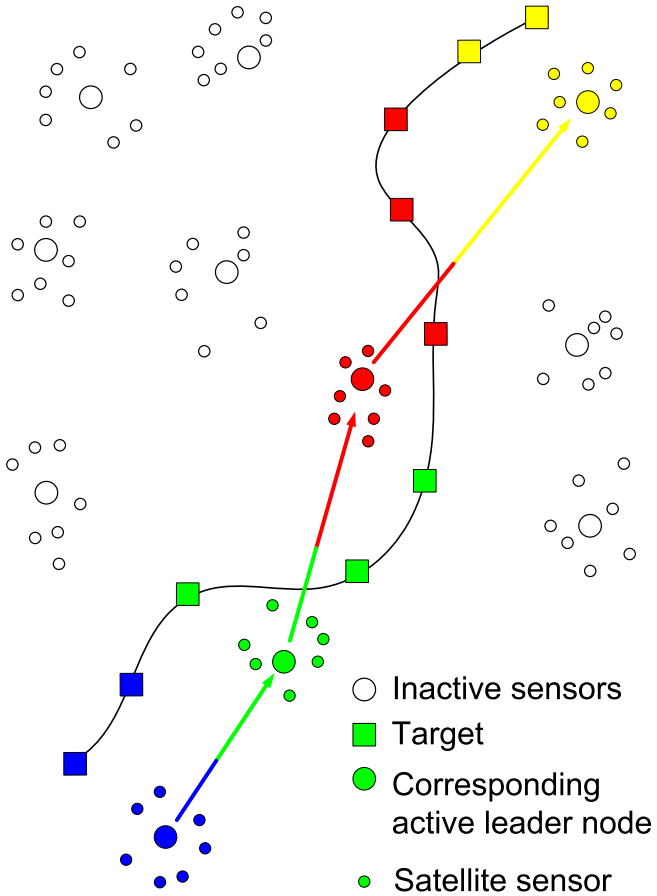


Fig. 1. The leader node distributed particle filtering setting

Throughout the paper, we will motivate the theoretical analysis by considering the concrete example of the "leader node" particle filter [20], an algorithm (described in detail below) that has been proposed for distributed tracking in sensor networks. We outline two other examples to illustrate that the analyzed problem arises in several practical settings.

A. Example 1: leader node particle filter for sensor network tracking

One of the major concerns in distributed sensor network tracking is balancing the tradeoff between tracking performance and network lifetime. Sensor nodes are most commonly battery-powered devices, so it is important to limit the energy consumption, which is dominated (if the sensors are passive) by communication. Particle filter tracking algorithms in sensor networks frequently adopt a centralized approach, wherein the particle filter resides at a computation centre and measurements from remote sensors are transported across the network to this centre. This approach has several disadvantages. Centralization introduces a single point of failure and can lead to high, unevenly distributed energy consumption because of the high communication cost involved in transmitting the data.

Distributed algorithms, such as the distributed particle filtering algorithms proposed in [1], [24], address the aforementioned problems. These algorithms decentralize the computation or communication so that a single fusion centre is not required. Multiple particle filters run concurrently at different sensor nodes and compressed data or approximate filtering distributions are shared between them. These distributed algorithms, while mitigating some of the inherent problems of centralization, can be computationally expensive, because multiple nodes are required to perform computation throughout the entire tracking procedure.

The *leader node particle filter*, proposed in [20], [29] and refined and analyzed in [11], [27], represents a compromise; it is partially distributed, in the sense that at any time-step only one node performs the particle filtering (the *leader* node), but this node changes over time. The setting corresponding to this filtering paradigm is

depicted in Fig. 1. A leader node (depicted by the large circles) is responsible for performing local tracking based on the data acquired by the *satellite* sensor nodes (depicted by small circles). The satellite nodes have sensing capabilities and can locally transmit the acquired data to the nearest leader node. The leader node fuses the data gathered by the satellite nodes in its neigbourhood, incorporating them into its particle filter.

Significant communication energy savings can be achieved if the leader node follows the target, because sensor nodes only have to relay their measurements to a nearby location. This is illustrated in Figure 1; the leader node approximately tracks the target trajectory (depicted by the squares). Sensor management strategies are used to determine when to change leader node [27]. When this occurs, information must be exchanged so that the new leader node can reconstruct the particle filter. In attempting to alleviate the communication cost of transmitting all particle values when the leader node is exchanged (which can involve thousands of bits), the filtering distribution is more coarsely approximated, either by transmitting only a subset of the particles or by training a parametric model.

B. Example 2: Tracking with delayed measurements

In wireless sensor networks, packet losses can lead to measurements arriving out-of-order to a node performing tracking. Incorporating delayed measurements into a particle filter is important, because they can be highly informative and improve tracking performance. One of the simplest, and most effective, strategies is to run the particle filter again from the time-step corresponding to the delayed measurement. This strategy can be hampered by the limited memory of most sensor network devices, which means it is impossible to store full particle representations for multiple time-steps. The alternative is to store an approximation, either a subsampled set of particles, or a parametric representation, for previous time-steps. When the particle filter is run again, it is initialized by sampling from the approximated distribution. The effect is equivalent to injecting intermittent approximations (subsampling or parametric) into the particle filter.

C. Example 3: Real-time tracking with computational constraints

When real-time tracking is performed on an embedded processor with computational limitations, it can be important to adjust the time devoted to particle filter computation. For example, consider a mobile robot that employs a particle filter to track its position and at the same time conducts iterative strategic planning of its motion in order to reach a target location. The goal can be achieved more efficiently (in less time and with less energy expenditure) if there is an adjustment of the computational time devoted to each of these two tasks. The adaptive particle filter proposed in [10] and the real-time particle filter of [14] adjust the number of particles at each time-step based on an estimate of the complexity of the filtering distribution (assigning fewer particles for simple distributions). Through these schemes, the accuracy of the position estimation can be preserved, but more time can be devoted to motion planning. The adaptation of the number of particles is an example of the subsampling approximation that we analyze in this paper.

D. Paper Organization

The rest of the paper is organized as follows. Tables 1 and 2 summarize the notation that is used in the paper. Section II sets up the analysis framework and outlines relevant terminology. Section III presents some foundational results that serve as the basis for our analysis. In Section IV we present error bounds and exponential inequalities for particle filters that perform intermittent subsampling, and in Section V we analyze the performance of particle filters that employ parametric approximation. Section VI describes numerical experiments that illustrate the performance of the algorithms we analyze and the relationship to the bounds. Section VII discusses related work and Section VIII summarizes the contribution and makes concluding remarks.

II. Analysis Framework and Terminology

We consider a discrete-time non-linear filtering task in which the target dynamics and observations can be described by the following general state-space signal model:

$$X_t = f_t(X_{t-1}, \varrho_t) \tag{1}$$

$$Y_t = g_t(X_t, \zeta_t). \tag{2}$$

Here $X_t \in \mathbb{R}^{d_x}$ is the target state vector at time t, $Y_t \in \mathbb{R}^{d_y}$ is the measurement vector, ϱ_t and ζ_t are system excitation and measurement noises respectively, f_t is a nonlinear system map $f_t : \mathbb{R}^{d_x} \to \mathbb{R}^{d_x}$, and g_t is a nonlinear measurement map $g_t : \mathbb{R}^{d_x} \to \mathbb{R}^{d_y}$.

In order to conduct stability (error propagation) analysis, we need to introduce slightly more rigorous mathematical notation. Let $(E_t, \mathcal{E}_t)$, $t \in \mathbb{N}$ be a sequence of measurable spaces. The target state vector evolves according to a non-homogeneous discrete-time Markov chain X_t with transitions M_{t+1} from E_t into E_{t+1}. We denote by $X'_t = X_{[0:t]}$ the historical path process associated with X_t. Associated with a measurable space of the form $(E, \mathcal{E})$ is a set of probability measures $\mathcal{P}(E)$ and the Banach space of bounded functions $\mathcal{B}_b(E)$ with supremum norm:

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

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

$$\begin{aligned} \mathrm{osc}(h) &= \sup(|h(x) - h(y)|;\ x, y \in E) \\ \mathrm{Osc}_1(E) &= \{h : \mathrm{osc}(h) \leq 1\} \end{aligned}$$

For any $h \in \mathcal{B}_b(E)$ it is also possible to define the following:

$$||h||_{\mathrm{osc}} = ||h|| + \mathrm{osc}(h).$$

In order to simplify the representation, we define for a measure $\mu \in \mathcal{P}(E)$,

$$\mu(h) = \int_E h(x)\mu(dx)$$

and for the Markov kernel from $(E_{i-1}, \mathcal{E}_{i-1})$ to $(E_i, \mathcal{E}_i)$:

$$(\mu_{i-1}M_i)(A_i) = \int_{E_{i-1}} \mu_{i-1}(\mathrm{d}x_{i-1})M_i(x_{i-1}, A_i).$$

Thus the composite integral operator from $(E_i, \mathcal{E}_i)$ to $(E_t, \mathcal{E}_t)$, $M_{i,t} = M_{i+1} \dots M_t$, has the form:

$$(M_{i+1} \dots M_t)(x_i, \mathrm{d}x_t) = \int_{E_{[i+1:t-1]}} M_{i+1}(x_i, \mathrm{d}x_{i+1}) \dots M_t(x_{t-1}, \mathrm{d}x_t).$$

A. Feynman-Kac models

Throughout the rest of this paper we adopt the methodology developed in [4] to analyze the behaviour of filtering distributions arising from (1) and (2). This methodology involves representing the particle filter as an N-particle approximation of a Feynman-Kac model. We now discuss the Feynman-Kac representation; for a much more detailed description and discussion, please refer to [4].

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

$$\Pr\{X_t \in \mathrm{d}x_t | X_{t-1} = x_{t-1}\} = M_t(x_{t-1}, \mathrm{d}x_t) \tag{3}$$

According to (3), the signal distribution at time t, with respect to the sequence of random variables $X_1, \dots, X_t$, can be written as follows

$$\mathbb{P}_{\mu,t}(\mathrm{d}(x_0, \dots, x_t)) = \mu(\mathrm{d}x_0)M_1(x_0, \mathrm{d}x_1) \dots M_t(x_{t-1}, \mathrm{d}x_t)$$

defining the filtered probability space

$$\left(\Omega = \prod_{i=0}^{t} E_i, \mathcal{F}_t, \mathcal{F}_\infty, \mathbb{P}_\mu\right),$$

where the family of σ-algebras has the following property: $\mathcal{F}_i \subset \mathcal{F}_j \subset \mathcal{F}_\infty$ for any $i \leq j$ and $\mathcal{F}_\infty = \sigma(\cup_{i \geq 0}\mathcal{F}_i)$ We now introduce bounded and non-negative potential functions $G_t : E_t \to [0, \infty)$ that characterize the properties of

the observation process in (2). This leads to the following definition of the unnormalized *prediction* Feynman-Kac model, for $h_t \in \mathcal{B}_b(E_t)$ and $t \in \mathbb{N}$.

$$\begin{aligned}\gamma_t(h_t) &= \mathbb{E}_{\eta_0}\left(h_t(X_t)\prod_{i=0}^{t-1} G_i(X_i)\right) \\ &= \int_{E_{[0:t]}} h_t(X_t)\prod_{i=0}^{t-1} G_i(X_i)\mathbb{P}_{\eta_0,t}(\mathrm{d}(x_0,\ldots,x_t))\end{aligned}$$

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

$$\eta_t(h_t) = \frac{\gamma_t(h_t)}{\gamma_t(1)}$$

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

$$\Psi_t(\nu)(\mathrm{d}x_t) = \frac{1}{\nu(G_t)}G_t(x_t)\nu(\mathrm{d}x_t).$$

This transformation is used to construct the key operator $\Phi_t : \mathcal{P}(E_{t-1}) \to \mathcal{P}(E_t)$, which is used to update the predictive posterior distribution from time step $t-1$ to time step t:

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

Thus this operator combines the fitness assessment described by the likelihood function G_{t-1} and the diffusion step described by the Markov kernel M_t

$$\Phi_t(\eta_{t-1}) = \Psi_{t-1}(\eta_{t-1})M_t$$

The repeated application of this operator, $\Phi_t(\eta_{t-1})_{t\geq 1}$, results in the semigroups $\Phi_{i,t}$, $i \leq t$ associated with the normalized Feynman-Kac distribution flows η_t.

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

The semigroup $\Phi_{i,t}$ describes the evolution of the normalized prediction Feynman-Kac model from time i to time t:

$$\eta_t = \Phi_{i,t}(\eta_i)$$

It is related to $G_{i,t} : E_i \to (0,\infty)$, the composite potential functions on E_i, and $P_{i,t} : \mathcal{P}(E_i) \to \mathcal{P}(E_t)$, the Markov kernels from E_i to E_t. In particular, $G_{i,t}$ is defined as the expectation of the composite potential constructed based on the observations acquired during time-steps $i,\ldots,t-1$ with respect to the shifted chain $M_{i+1}\ldots M_t$:

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

and $P_{i,t}$ is defined by the Feynman-Kac formulae as follows:

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

The Boltzmann-Gibbs transformation

$$\Psi_{i,t}(\eta_i)(h_i) = \frac{\eta_i(G_{i,t}h_i)}{\eta_i(G_{i,t})}$$

and the semigroup $\Phi_{i,t}(\eta_i)$

$$\Phi_{i,t}(\eta_i) = \Psi_{i,t}(\eta_i)P_{i,t}$$

can then be represented via these two quantities.

B. Feynman-Kac formulae and the Bayesian framework

It is instructive to draw parallels between the Feynman-Kac description of the filtering process discussed above and the Bayesian formulation of sequential filtering. The integral operator, $M_t(x_{t-1}, \mathrm{d}x_t)$, describing the evolution of signal diffusion in (1) is most naturally related to the state transition density (assuming one exists):

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

The measurement equation (2), which is compactly described by the potential function $G_t(x_t)$ in the Feynman-Kac framework, is directly related to the likelihood function $p_t(y_t|x_t)$ in the Bayesian framework:

$$G_t(x_t) = p_t(y_t|x_t).$$

We can then see how the diffusion step within the Feynman-Kac framework is related to the prediction step within the Bayesian framework:

$$\Phi_t(\eta_{t-1}) = \int_{E_{t-1}} \Psi_{t-1}(\eta_{t-1})(\mathrm{d}x_{t-1}) M_t(x_{t-1}, \mathrm{d}x_t) \quad \text{Feynman-Kac}$$

$$p_t(x_t|y_{1:t-1}) = \int_{E_{t-1}} p(x_{t-1}|y_{1:t-1}) p_t(x_t|x_{t-1})\mathrm{d}x_{t-1} \qquad \text{Bayes}$$

Thus the operator Φ_t generates the normalized predictive posterior distribution, $\eta_t(\mathrm{d}x_t) = \Phi_t(\eta_{t-1})(\mathrm{d}x_t) = p(x_t|y_{1:t-1})\mathrm{d}x_t$, and the Boltzmann-Gibbs transformation $\Psi_t(\eta_t)(\mathrm{d}x_t) = p(x_t|y_{1:t})\mathrm{d}x_t$ generates the normalized posterior distribution using the update step analogous to that of the Bayes model:

$$\Psi_{t-1}(\eta_{t-1}) = \frac{G_{t-1}(x_{t-1})\eta_{t-1}(\mathrm{d}x_{t-1})}{\int_{E_{t-1}} G_t(x_{t-1})\eta_{t-1}(\mathrm{d}x_{t-1})} \qquad \text{Feynman-Kac}$$

$$p_{t-1}(x_{t-1}|y_{1:t-1}) = \frac{p_{t-1}(y_{t-1}|x_{t-1}) p_{t-1}(x_{t-1}|y_{1:t-2})}{\int_{E_{t-1}} p_{t-1}(y_{t-1}|x_{t-1}) p_{t-1}(x_{t-1}|y_{1:t-2})\mathrm{d}x_{t-1}} \quad \text{Bayes}$$

The normalization constant $\eta_{t-1}(G_{t-1}) = p(y_{t-1}|y_{1:t-2})$ can be interpreted as equivalent to Bayes' evidence at time $t-1$, $p(y_{t-1}|y_{1:t-2})$. We conclude that the Feynman-Kac formulae are directly related to the predict-update Bayesian recursion. The difference between the two formulations lies in the fact that the Feynman-Kac formulae describe the evolution of distributions, while the Bayesian framework describes the evolution of the corresponding densities (assuming that these densities exist).

C. N-particle approximations

Following [4], we can define a particle filter by developing an N-particle approximation to the Feynman-Kac model. This approximation consists of N path particles:

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

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

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

where δ is the Dirac delta function.

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

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

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

D. Regularity Conditions

The analysis we present relies on certain assumptions about the regularity and mixing properties of the Markov kernels and likelihood potential functions. We adopt the same assumptions used for deriving time-uniform bounds in [4].

We define the following condition on the Markov functions:

- $(M)_{\mathrm{u}}^{(m)}$ There exists an integer $m \geq 1$ and strictly positive number $\epsilon_{\mathrm{u}}(M) \in (0,1)$ such that for any $i \geq 0$ and $x_i, y_i \in E_i$ we have

$$M_{i,i+m}(x_i, \cdot) = M_{i+1} M_{i+2} \ldots M_{i+m}(x_i, \cdot) \geq \epsilon_{\mathrm{u}}(M) M_{i,i+m}(y_i, \cdot).$$

The following regularity condition is defined for the likelihood potentials:

- $(G)_{\mathrm{u}}$ There exists a strictly positive number $\epsilon_{\mathrm{u}}(G) \in (0,1]$ such that for any t and $x_t, y_t \in E_t$

$$G_t(x_t) \geq \epsilon_{\mathrm{u}}(G) G_t(y_t)$$

The Dobrushin contraction coefficient ($\beta(P_{i,t}) \in [0,1]$) plays a key role in our analysis. This is defined as follows:

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

If $(G)_{\mathrm{u}}$ and $(M)_{\mathrm{u}}^{(m)}$ hold, then according to Proposition 4.3.3. [4] we have the following estimate for the Dobrushin contraction coefficient:

$$\beta(P_{i,t}) \leq \left(1 - \epsilon_{\mathrm{u}}^2(M) \epsilon_{\mathrm{u}}^{(m-1)}(G)\right)^{\lfloor (t-i)/m \rfloor}.$$

III. Foundational Results

Our analysis of error propagation relies on results that characterize the impact of sampling operations and the manner in which distribution discrepancies are propagated in the Feynman-Kac framework. In this section, we state and prove some of these foundational results, which build upon relationships presented in [4].

We first introduce some necessary notation. Denote by $(\mu_k)_{k\geq 1}$ a sequence of probability measures on $(E, \mathcal{E})$, and by $(h_k)_{k\geq 1}$ a sequence of $\mathcal{E}$-measurable functions. Denote by $m(X) = \frac{1}{N}\sum_{k=1}^{N} \delta_{X^k}$ the N-empirical measure associated with a collection of independent random variables $(X^k)_{k\geq 1}$ with respective distributions $(\mu_k)_{k\geq 1}$. Further, define $m(X)(h) \triangleq \frac{1}{N}\sum\limits_{k=1}^{N} h_k(X_k)$ and $\sigma^2(h) \triangleq \frac{1}{N}\sum\limits_{k=1}^{N} \mathrm{osc}^2(h_k)$.

Let the sampling operator $S^N : \mathcal{P}(E) \to \mathcal{P}(E^N)$ be defined as:

$$S^N(\eta)(h) = \frac{1}{N}\sum_{k=1}^{N} h(\xi^k)\ . \tag{4}$$

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

A. Bounds on errors induced by sampling

The following result bounds the weak-sense L_p error induced by the sampling operator for functions with finite oscillations. We recall that for a measure $P \in \mathcal{P}(E)$, $P(h) \triangleq \int_E h(x) P(dx)$.

Lemma 1. *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.

Proof: Since $\mathbb{E}\{[P - S^N(P)](h)\} = P(h) - P(h) = 0$, we have from the Chernov-Hoeffding inequality:

$$\mathbb{P}\{|[P - S^N(P)](h)| \geq \epsilon\} \leq 2e^{-\frac{2N\epsilon^2}{\sigma^2(h)}}$$

We note that

$$\mathbb{P}\{|[P - S^N(P)](h)|^p \geq \epsilon\} = \mathbb{P}\{|[P - S^N(P)](h)| \geq \epsilon^{1/p}\}$$

and we have from the Chernov-Hoeffding inequality:

$$\mathbb{P}\{|[P - S^N(P)](h)| \geq \epsilon^{1/p}\} \leq 2e^{-2N\epsilon^{2/p}/\sigma^2(h)}$$

Next we recall the following property:

$$\mathbb{E}\{|[P - S^N(P)](h)|\} = \int_0^\infty \mathbb{P}\{|[P - S^N(P)](h)| \geq \epsilon\}\mathrm{d}\epsilon$$

And finally we obtain:

$$\begin{aligned}
\mathbb{E}\{|[P - S^N(P)](h)|^p\}^{\frac{1}{p}} &= \left[\int_0^\infty \mathbb{P}\{|[P - S^N(P)](h)| \geq \epsilon^{1/p}\}\mathrm{d}\epsilon\right]^{\frac{1}{p}} \\
&\leq \left[2\int_0^\infty e^{-2N\epsilon^{2/p}/\sigma^2(h)}\mathrm{d}\epsilon\right]^{\frac{1}{p}} \\
&= \left[\sigma^p(h)p(2N)^{-\frac{p}{2}}\Gamma\left[\frac{p}{2}\right]\right]^{\frac{1}{p}}
\end{aligned}$$

Applying Lemma 7.3.3 of [4] allows us to set $c(1) = 1$ instead of $c(1) = 2^{-1/2}\Gamma[1/2] = \sqrt{\pi/2}$, and this completes the proof. ■

An almost identical result applies for the N-empirical measure $m(X)$. Lemma 2 tightens Lemma 7.3.3 from [4] and extends it to include non-integer p (for a comparison of the two bounds see Appendix A).

Lemma 2. *For a sequence of $\mathcal{E}$-measurable functions $(h_k)_{k\geq 1}$ with finite oscillations and satisfying $\mu_k(h_k) = 0$ for all $k \geq 1$,*

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

for any $p \geq 1$, where $c(p)$ is defined as in Lemma 1.

The proof follows the same argument as that of Lemma 1, noting that the condition $\mathbb{E}(m(X)(h)) = 0$, which is necessary for application of the Chernoff-Hoeffding bound, follows from the condition $\mu_k(h_k) = 0$.

The following theorem provides a bound on the moment-generating function of the empirical measure $m(X)$. The result employs Lemma 1 to tighten Theorem 7.3.1 of [4] (for a comparison of the two bounds see Appendix B).

Theorem 1. *For any sequence of $\mathcal{E}$–measurable functions $(h_k)_{k\geq 1}$ such that $\mu_k(h_k) = 0$ for all $k \geq 1$ and $\sigma(h) < \infty$, we have for any ε*

$$\mathbb{E}\left\{e^{\varepsilon\sqrt{N}|m(X)(h)|}\right\} \leq 1 + \varepsilon\sigma(h)\left(1 - \sqrt{\frac{\pi}{2}} + \sqrt{\frac{\pi}{2}}e^{\frac{\varepsilon^2}{8}\sigma^2(h)}\left[1 + \mathrm{Erf}\left[\frac{\varepsilon\sigma(h)}{\sqrt{8}}\right]\right]\right)$$

Proof: We first utilize the power series representation of the exponential:

$$\begin{aligned}
\mathbb{E}\left\{e^{\varepsilon|m(X)(h)|}\right\} &= \sum_{n\geq 0}\frac{\varepsilon^n}{n!}\mathbb{E}\left\{|m(X)(h)|^n\right\} \\
&= \varepsilon^0\mathbb{E}\left\{|m(X)(h)|^0\right\} + \varepsilon\mathbb{E}\left\{|m(X)(h)|\right\} + \sum_{n\geq 2}\frac{\varepsilon^n}{n!}\mathbb{E}\left\{|m(X)(h)|^n\right\}
\end{aligned}$$

Utilizing Lemma 1 we have:

$$\begin{aligned}
&\mathbb{E}\left\{e^{\varepsilon|m(X)(h)|}\right\} \\
&\leq 1+\frac{\varepsilon\sigma(h)}{\sqrt{N}}+\sum_{n\geq 2}\frac{\varepsilon^n}{n!}\sigma^n(h)n(2N)^{-n/2}\Gamma\left[n/2\right] \\
&= 1+\frac{\varepsilon\sigma(h)}{\sqrt{N}}+\sum_{n\geq 2}\left[\frac{\varepsilon\sigma(h)}{(2N)^{1/2}}\right]^n\frac{\Gamma\left[n/2\right]}{(n-1)!} \\
&= 1+\frac{\varepsilon\sigma(h)}{\sqrt{N}}-\frac{\varepsilon\sigma(h)\sqrt{\pi}}{\sqrt{2N}}+\frac{\varepsilon\sigma(h)\sqrt{\pi}}{\sqrt{2N}}e^{\frac{\varepsilon^2\sigma^2(h)}{8N}}\left[1+\mathrm{Erf}\left[\frac{\varepsilon\sigma(h)}{\sqrt{8N}}\right]\right]
\end{aligned}$$

Choosing $\varepsilon=\varepsilon\sqrt{N}$ and rearranging terms completes the proof. ■

The following corollary containing a more tractable variation of the previous theorem can be useful for deriving the exponential inequalities for the particle approximations of Feynman-Kac models.

Corollary 1. *For any sequence of $\mathcal{E}$–measurable functions $(h_k)_{k\geq 1}$ such that $\mu_k(h_k)=0$ for all $k\geq 1$ we have for any ε*

$$\sigma(h)<\infty\Longrightarrow\mathbb{E}\left\{e^{\varepsilon\sqrt{N}|m(X)(h)|}\right\}\leq\left(1+\sqrt{2\pi}\varepsilon\sigma(h)\right)e^{\frac{\varepsilon^2}{8}\sigma^2(h)}$$

Proof: The proof is straightforward since $\sup_x\mathrm{Erf}(x)=1$, $1-\sqrt{\pi/2}<0$ and $e^{\frac{\varepsilon^2}{8}\sigma^2(h)}\geq 1$. ■

We note that the simplified estimate of the moment-generating function in Corollary 1 is much tighter than the bound in Theorem 7.3.1 of [4] for asymptotically large deviations ε while the more complex bound in Theorem 1 is uniformly tighter over the range of ε.

B. Error propagation

Proposition 4.3.7. in [4] underpins our analysis of the stability of the semigroups $\Phi_{i,t}$.

Proposition 1 (Del Moral [4], Proposition 4.3.7)**.** *For any $0\leq i\leq t$, $\mu_i\in\mathcal{P}(E_i)$, and $h_t\in\mathcal{B}_b(E_t)$ with $\mathrm{osc}(h_t)\leq 1$, respectively $||h_t||\leq 1$, there exists a function h_i in $\mathcal{B}_b(E_i)$ with $\mathrm{osc}(h_i)\leq 1$, respectively $||h_i||\leq 1$, such that for any $\eta_i\in\mathcal{P}(E_i)$ we have*

$$|[\Phi_{i,t}(\eta_i)-\Phi_{i,t}(\mu_i)](h_t)|\leq\beta(P_{i,t})\frac{||G_{i,t}||_{\mathrm{osc}}}{\eta_i(G_{i,t})}|(\eta_i-\mu_i)(h_i)| \tag{5}$$

and respectively

$$|[\Phi_{i,t}(\eta_i)-\Phi_{i,t}(\mu_i)](h_t)|\leq\beta(P_{i,t})\frac{2||G_{i,t}||}{\eta_i(G_{i,t})}|(\eta_i-\mu_i)(h_i)| \tag{6}$$

Using the following logic for some positive function φ we can obtain a result intermediate between (5) and (6):

$$\begin{aligned}
||\varphi||_{\mathrm{osc}} &= ||\varphi||+\mathrm{osc}(\varphi)=||\varphi||+\sup_{x,y\in E}|\varphi(x)-\varphi(y)| \\
&\leq||\varphi||\left(1+\frac{1}{||\varphi||}\sup_{x,y\in E}\left|1-\frac{\varphi(y)}{\varphi(x)}\right||\varphi(x)|\right) \\
&\leq||\varphi||\left(1+\frac{||\varphi||}{||\varphi||}\sup_{x,y\in E}\left|1-\frac{\varphi(y)}{\varphi(x)}\right|\right) \\
&\leq||\varphi||\left(2-\frac{\inf_{y\in E}\varphi(y)}{\sup_{x\in E}\varphi(x)}\right)
\end{aligned}$$

Summarizing, we can formulate the following proposition.

Proposition 2. *For any $0 \leq i \leq t$, $\mu_i \in \mathcal{P}(E_i)$, and $h_t \in \mathcal{B}_b(E_t)$ with $\mathrm{osc}(h_t) \leq 1$ there exists a function h_i in $\mathcal{B}_b(E_i)$ with $\mathrm{osc}(h_i) \leq 1$ such that for any $\eta_i \in \mathcal{P}(E_i)$ we have*

$$|[\Phi_{i,t}(\eta_i) - \Phi_{i,t}(\mu_i)](h_t)| \leq \beta(P_{i,t}) \frac{||G_{i,t}||}{\eta_i(G_{i,t})} \left[2 - \frac{\inf_{y_i \in E_i} G_{i,t}(y_i)}{||G_{i,t}||}\right] |(\eta_i - \mu_i)(h_i)|$$

These results describe the propagation of one-step approximation error through the non-linear operator $\Phi_{i,t}$. They reveal the link between the initial error at time i and the propagated error at time t through the properties of the potential functions $G_{i,t}$ and the Dobrushin contraction coefficient $\beta(P_{i,t})$.

According to Proposition 4.3.3. [4], the oscillations of the potential functions can be bounded as follows under assumptions $(G)_{\mathrm{u}}$ and $(M)_{\mathrm{u}}^{(m)}$:

$$\frac{\inf_{x_i \in E_i} G_{i,t}(x_i)}{||G_{i,t}||} \geq \epsilon_{\mathrm{u}}(M) \epsilon_{\mathrm{u}}^m(G)$$
$$\frac{||G_{i,t}||}{\eta_i(G_{i,t})} \leq \epsilon_{\mathrm{u}}^{-1}(M) \epsilon_{\mathrm{u}}^{-m}(G)$$

Thus Proposition 2 implies that under assumptions $(G)_{\mathrm{u}}$ and $(M)_{\mathrm{u}}^{(m)}$ the error propagation in the sequential Feynman-Kac filter can be characterized as follows:

$$|[\Phi_{i,t}(\eta_i) - \Phi_{i,t}(\mu_i)](h_t)| \leq \left(1 - \epsilon_{\mathrm{u}}^2(M) \epsilon_{\mathrm{u}}^{(m-1)}(G)\right)^{\lfloor (t-i)/m \rfloor} \times \frac{2 - \epsilon_{\mathrm{u}}(M) \epsilon_{\mathrm{u}}^m(G)}{\epsilon_{\mathrm{u}}(M) \epsilon_{\mathrm{u}}^m(G)} |(\eta_i - \mu_i)(h_i)| \tag{7}$$

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 $\eta_i^N - \Phi_i(\eta_{i-1}^N), i = 0, \ldots, t$ [4]:

$$\eta_t^N - \eta_t = \sum_{i=0}^{t} \left[\Phi_{i,t}(\eta_i^N) - \Phi_{i,t}(\Phi_i(\eta_{i-1}^N))\right]. \tag{8}$$

IV. Particle Filters with Intermittent Subsampling

This section presents an analysis of the error propagation in particle filters that perform intermittent subsampling approximation steps. We focus on the case where the number of particles N is constant and the subsampling approximation step always uses N_b particles. It is relatively straightforward to generalize our results to other settings, where the number of particles N and the size of the subsample approximation may vary over time. Our main results are a time-uniform bound on the weak-sense L_p-error and an associated exponential inequality.

A. Algorithm Description

Denote by δ_t a binary variable which indicates whether a subsampling approximation is performed at time-step t. In our analysis, we will assume that this variable is the outcome of a decision function based on the set of particles $\{\xi_{t-1}^k\}_{k=1}^N$ and observations y_t at the previous time-step. Similar results could be obtained should the decision function be of a more general nature (for example, based on either the entire history of the particle filter, $(\xi'^k)_{k=1}^N$, and/or the entire history of measurements $y_{1:t}$). We define $\delta_0(\cdot, \cdot) = 0$, and we assume there exist probabilities:

$$\mathbb{E}\{\delta_t\} = \mathbb{P}\{\delta_t = 1\} = q_t$$

The expectation is with respect to the Monte-Carlo sampling, measurement noise and the possible target trajectories. The value of q_t characterizes the frequency of subsampling approximation at time-step t.

The subsample approximation particle filter can then be expressed as:

$$\eta_t^N = S^N \circ S^{N_{\mathrm{b}}}(\Phi_t(\eta_{t-1}^N)) \quad \text{if } \delta_t = 1,$$
$$\eta_t^N = S^N(\Phi_t(\eta_{t-1}^N)) \quad \text{if } \delta_t = 0 \tag{9}$$

Example: leader node particle filter with subsampling

Suppose that $\mathcal{L} = \{1, 2, \ldots, L\}$ is the set of leader node labels and every leader node with label $\ell \in \mathcal{L}$ has a set of satellite nodes $\mathcal{S}_\ell$ that take measurements and transmit them to the leader node. The number of such satellite nodes in the vicinity of the leader node ℓ is $|\mathcal{S}_\ell|$. The state-space model describing the target evolution and measurement process at every leader node is then a simple modification of the general state-space model described in Section II:

$$X_t = f_t(X_{t-1}, \varrho_t) \tag{10}$$

$$Y_t^j = g_t^j(X_t^j, \zeta_t^j) \quad \forall j \in \mathcal{S}_{\ell_t} \tag{11}$$

Thus the target model is the same at every leader node but the observation process may be different.

For the leader node particle filter, the decision function $\delta_t(\{\xi_{t-1}^j\}_{j=1}^N, Y_t^{\mathcal{S}_{\ell_t}})$, is used to decide whether or not to change a leader node at time t. Here $Y_t^{\mathcal{S}_{\ell_t}}$ denotes the set of all measurements recorded at time t by the set of satellite sensor nodes $\mathcal{S}_{\ell_t}$ of leader node ℓ_t. The leader node algorithm with subsampling can be described as follows:

$$\Phi_{\ell_t}(\eta_{t-1}^N) \Rightarrow \eta_t^{N_\mathrm{b}} \longrightarrow \eta_t^{N_\mathrm{b}} \Rightarrow \eta_t^N \quad \text{if } \delta_t = 1, \tag{12}$$

$$\Phi_{\ell_t}(\eta_{t-1}^N) \Rightarrow \eta_t^N \quad \text{if } \delta_t = 0 \tag{13}$$

Here the implication sign $\Rightarrow$ represents a sampling operation and the right arrow $\longrightarrow$ denotes the communication process.

If $\delta_t = 1$, the current leader node ℓ_t determines the next leader node ℓ_{t+1} (through a sensor management algorithm, see e.g. [27]), and calculates the N_b-particle approximation to the current predictive posterior distribution, η_t^N, by sub-sampling the output of the standard N-particle propagation step. Finally, the leader node ℓ_t transmits $\eta_t^{N_\mathrm{b}}$ to the next leader node ℓ_{t+1}, which recovers the N-particle approximation by up-sampling. If $\delta_t = 0$, the current leader node performs standard particle propagation.

In the leader node setting, the uniform condition $(G)_\mathrm{u}$ can be formulated as follows:

- $(G)_\mathrm{u}^\ell$ There exists a strictly positive number $\epsilon_\mathrm{u}(G) \in (0, 1]$ such that for any ℓ, t and $x_t, y_t \in E_t$

$$G_{\ell_t}(x_t) \geq \epsilon_\mathrm{u}^{K_\mathrm{u}}(G) G_{\ell_t}(y_t)$$

Indeed, $(G)_\mathrm{u}^\ell$ holds if we take $\epsilon_\mathrm{u}(G) = \inf_{t\geq 0} \min_{\ell_t \in \mathcal{L}} \epsilon_{\ell_t}(G)$ and $K_\mathrm{u} = \max_{\ell \in \mathcal{L}} |\mathcal{S}_\ell|$. Thus Proposition 2 implies that in the leader node setting and under assumptions $(G)_\mathrm{u}^\ell$ and $(M)_\mathrm{u}^{(m)}$ the error propagation in the sequential Feynman-Kac filter can be characterized as follows:

$$\begin{aligned} |[\Phi_{\ell_i,\ell_t}(\eta_i) - \Phi_{\ell_i,\ell_t}(\mu_i)](h_t)| \leq & \left(1 - \epsilon_\mathrm{u}^2(M)\epsilon_\mathrm{u}^{(m-1)K_\mathrm{u}}(G)\right)^{\lfloor (t-i)/m \rfloor} \\ & \times \frac{2 - \epsilon_\mathrm{u}(M)\epsilon_\mathrm{u}^{mK_\mathrm{u}}(G)}{\epsilon_\mathrm{u}(M)\epsilon_\mathrm{u}^{mK_\mathrm{u}}(G)} |(\eta_i - \mu_i)(h_i)| \end{aligned} \tag{14}$$

B. Time-uniform error bounds and exponential inequalities

We now analyze the global approximation error for particle filtering with intermittent subsampling. We first present a theorem that specifies a time-uniform bound on the weak-sense L_p error.

Theorem 2. *Suppose assumptions $(G)_u$ and $(M)_u^{(m)}$ hold. Suppose further that $\mathbb{P}\{\delta_i = 1\} \leq q_u$ for any $i \geq 0$ and $0 \leq q_u \leq 2/3$. Then for a positive integer χ such that $N = \chi N_b$, $t \geq 0$, $p \geq 1$ and $h_t \in \mathrm{Osc}_1(E_t)$ we have the time uniform estimate*

$$\sup_{t\geq 0} \mathbb{E}\left\{|[\eta_t^N - \eta_t](h_t)|^p\right\}^{1/p} \leq \frac{\epsilon_{u,m} c^{1/p}(p)}{\sqrt{N}} \left(q_u^{1/p}\sqrt{\chi} + (1 - q_u)^{1/p}\right)$$

where the constant $\epsilon_{u,m}$ is[1]*:*

$$\epsilon_{u,m} = \frac{m(2 - \epsilon_u(M)\epsilon_u^m(G))}{\epsilon_u^3(M)\epsilon_u^{(2m-1)}(G)}. \tag{15}$$

[1]In the leader node setting this is given by the following expression $\epsilon_{\mathrm{u},m} = m(2 - \epsilon_\mathrm{u}(M)\epsilon_\mathrm{u}^{mK_\mathrm{u}}(G))/\epsilon_\mathrm{u}^3(M)\epsilon_\mathrm{u}^{(2m-1)K_\mathrm{u}}(G)$.

Proof: We begin by applying Minkowski's inequality to (8)

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

and then applying Proposition 2:

$$\begin{aligned}&\sum_{i=0}^{t}\mathbb{E}\left\{\left|\left[\Phi_{\ell_i,\ell_t}(\eta_i^N)-\Phi_{\ell_i,\ell_t}(\Phi_{\ell_i}(\eta_{i-1}^N))\right](h_i)\right|^p\right\}^{\frac{1}{p}}\\&\leq\sum_{i=0}^{t}\mathbb{E}\left\{\left|\beta(P_{\ell_i,\ell_t})\frac{||G_{\ell_i,\ell_t}||}{\eta_i(G_{\ell_i,\ell_t})}\left[2-\frac{\inf_{y_i\in E_i}G_{\ell_i,\ell_t}(y_i)}{||G_{\ell_i,\ell_t}||}\right]\right|^p\left|\left[\eta_i^N-\Phi_{\ell_i}(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{1/p}.\end{aligned}$$

Furthermore, applying (7) we have:

$$\begin{aligned}&\sum_{i=0}^{t}\mathbb{E}\left\{\left|\left[\Phi_{\ell_i,\ell_t}(\eta_i^N)-\Phi_{\ell_i,\ell_t}(\Phi_{\ell_i}(\eta_{i-1}^N))\right](h_i)\right|^p\right\}^{\frac{1}{p}}\\&\leq\frac{2-\epsilon_{\mathrm{u}}(M)\epsilon_{\mathrm{u}}^m(G)}{\epsilon_{\mathrm{u}}(M)\epsilon_{\mathrm{u}}^m(G)}\\&\qquad\times\sum_{i=0}^{t}\left(1-\epsilon_{\mathrm{u}}^2(M)\epsilon_{\mathrm{u}}^{(m-1)}(G)\right)^{\lfloor(t-i)/m\rfloor}\mathbb{E}\left\{\left|\left[\eta_i^N-\Phi_{\ell_i}(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{1/p}.\end{aligned}$$

Next we analyze each individual expectation under the sum above. In particular, using the structure of the algorithm defined in (24) and the definition of sampling operator introduced in (4) we can rewrite the terms comprising the sum in the following explicit way:

$$\begin{aligned}&\mathbb{E}\left\{\left|\left[\eta_i^N-\Phi_{\ell_i}(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{\frac{1}{p}}\\&=\mathbb{E}\left\{\left|\left[\delta_i S^N\circ S^{N_{\mathrm{b}}}(\Phi_{\ell_i}(\eta_{i-1}^N))+(1-\delta_i)S^N(\Phi_{\ell_i}(\eta_{i-1}^N))-\Phi_{\ell_i}(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{\frac{1}{p}}\end{aligned}\tag{16}$$

Grouping the terms and using Minkowski's inequality again we conclude the following:

$$\begin{aligned}\mathbb{E}\left\{\left|\left[\eta_i^N-\Phi_{\ell_i}(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{\frac{1}{p}}&\leq\mathbb{E}\left\{\left|\delta_i\left[S^N\circ S^{N_{\mathrm{b}}}(\Phi_{\ell_i}(\eta_{i-1}^N))-\Phi_{\ell_i}(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{\frac{1}{p}}\\&\quad+\mathbb{E}\left\{\left|(1-\delta_i)\left[S^N(\Phi_{\ell_i}(\eta_{i-1}^N))-\Phi_{\ell_i}(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{\frac{1}{p}}.\end{aligned}$$

Adding and subtracting $\delta_i S^{N_{\mathrm{b}}}(\Phi_{\ell_i}(\eta_{i-1}^N))$ in the first term on the right and applying Minkowski's inequality again we have:

$$\begin{aligned}&\mathbb{E}\left\{\left|\left[\eta_i^N-\Phi_{\ell_i}(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{\frac{1}{p}}\\&\leq\mathbb{E}\left\{\left|\delta_i\left[S^N\circ S^{N_{\mathrm{b}}}(\Phi_{\ell_i}(\eta_{i-1}^N))-S^{N_{\mathrm{b}}}(\Phi_{\ell_i}(\eta_{i-1}^N))\right](h_i)\right|^p\right\}^{\frac{1}{p}}\\&\quad+\mathbb{E}\left\{\left|\delta_i\left[S^{N_{\mathrm{b}}}(\Phi_{\ell_i}(\eta_{i-1}^N))-\Phi_{\ell_i}(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{\frac{1}{p}}\\&\quad+\mathbb{E}\left\{\left|(1-\delta_i)\left[S^N(\Phi_{\ell_i}(\eta_{i-1}^N))-\Phi_{\ell_i}(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{\frac{1}{p}}\end{aligned}\tag{17}$$

We see that each error term under the sum splits into three individual terms, describing the approximation paths the leader node algorithm can follow at time i. If $N=\chi N_{\mathrm{b}}$ then the N-particle approximation after subsampling can be recovered from the N_{b}-particle approximation without error by replicating the N_{b}-particle approximation χ times. Thus the first term in (17) is zero.

The analysis of the remaining two terms is similar. We first concentrate on the second term. Recall that $\delta_i=\delta_i(\{\xi_{i-1}^j\}_{j=1}^N,Y_i^{\mathcal{S}_{\ell_i}})$. Thus given the σ-algebra $\mathcal{F}_{i-1}$ and the realization of the current measurement, $Y_i^{\mathcal{S}_{\ell_i}}=y_i^{\mathcal{S}_{\ell_i}}$, the output of the decision rule is independent of the sampling error, $[S^N(\Phi_{\ell_i}(\eta_{i-1}^N))-\Phi_{\ell_i}(\eta_{i-1}^N)](h_i)$. We exploit

this Markovian nature of the decision rule and apply Lemma 1 to the conditional expectation rendering the following bound:

$$\begin{aligned}
&\mathbb{E}\left\{\left|\delta_i\left[S^{N_\mathrm{b}}(\Phi_{\ell_i}(\eta_{i-1}^N))-\Phi_{\ell_i}(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{1/p} \\
&=\mathbb{E}\left\{\delta_i\mathbb{E}\left\{\left|\left[S^{N_\mathrm{b}}(\Phi_{\ell_i}(\eta_{i-1}^N))-\Phi_{\ell_i}(\eta_{i-1}^N)\right](h_i)\right|^p\middle|\mathcal{F}_{i-1},Y_i^{\mathcal{S}_{\ell_i}}=y_i^{\mathcal{S}_{\ell_i}}\right\}\right\}^{1/p} \\
&\leq\frac{c^{1/p}(p)}{\sqrt{N_\mathrm{b}}}q_i^{1/p}
\end{aligned} \tag{18}$$

Combining the analysis results for all three terms we obtain:

$$\mathbb{E}\left\{\left|\left[\eta_i^N-\Phi_{\ell_i}(\eta_{i-1}^N)\right](h_t)\right|^p\right\}^{1/p}\leq c^{1/p}(p)\left(q_i^{1/p}\frac{1}{\sqrt{N_\mathrm{b}}}+(1-q_i)^{1/p}\frac{1}{\sqrt{N}}\right)$$

We note that the expression in brackets has the form $\varphi(q_i)=q_i^{1/p}(\alpha+\beta)+(1-q_i)^{1/p}\alpha$ for some $\beta>\alpha\geq 0$. For $p\geq 1$, $\varphi(q_i)$ has maximum at $q_i=q_{\max}$ with

$$q_{\max}=\frac{1}{1+\left[\frac{\alpha+\beta}{\alpha}\right]^{p/(1-p)}}.$$

We have that $\varphi(q_i)$ is non-decreasing on $q_i\in[0,q_{\max}]$ and non-increasing on $q_i\in(q_{\max},1]$. Noting that $[(\alpha+\beta)/\alpha]^{p/(1-p)}$ is increasing in p we obtain:

$$q_{\max}\geq\frac{1}{1+\left[\frac{\alpha}{\alpha+\beta}\right]}\geq\inf_{\beta:\beta>\alpha}\frac{1}{1+\left[\frac{\alpha}{\alpha+\beta}\right]}=2/3.$$

Thus if $q_\mathrm{u}\leq 2/3\leq q_{\max}$ then for any $i\geq 0$ we have the time-uniform estimate:

$$\mathbb{E}\left\{\left|\left[\eta_i^N-\Phi_{\ell_i}(\eta_{i-1}^N)\right](h_t)\right|^p\right\}^{1/p}\leq c^{1/p}(p)\left(q_\mathrm{u}^{1/p}\frac{1}{\sqrt{N_\mathrm{b}}}+(1-q_\mathrm{u})^{1/p}\frac{1}{\sqrt{N}}\right)$$

Finally, noting [4] that:

$$\sum_{i=0}^{t}\left(1-\epsilon_\mathrm{u}^2(M)\epsilon_\mathrm{u}^{(m-1)}(G)\right)^{\lfloor(t-i)/m\rfloor}\leq\frac{m}{\epsilon_\mathrm{u}^2(M)\epsilon_\mathrm{u}^{(m-1)}(G)} \tag{19}$$

we complete the proof of theorem. ■

The result can be generalized to cases where N is not an integer multiple of N_b, at the expense of a slight loosening of the bound.

Corollary 2. *Suppose the assumptions of Theorem 2 apply, except we allow any integer $N_b<N$. Then for any $t\geq 0$, $p\geq 1$ and $h_t\in\mathrm{Osc}_1(E_t)$ we have the time uniform estimate*

$$\sup_{t\geq 0}\mathbb{E}\left\{|[\eta_t^N-\eta_t](h_t)|^p\right\}^{1/p}\leq\epsilon_{u,m}c^{1/p}(p)\left(q_u^{1/p}\left[\frac{1}{\sqrt{N}}+\frac{1}{\sqrt{N_b}}\right]+(1-q_u)^{1/p}\frac{1}{\sqrt{N}}\right)$$

where the constant $\epsilon_{u,m}$ is defined as in (15).

The corollary follows by allowing for sampling error to arise in the first term in (17):

$$\mathbb{E}\left\{\left|\delta_i\left[S^N\circ S^{N_\mathrm{b}}(\Phi_{\ell_i}(\eta_{i-1}^N))-S^{N_\mathrm{b}}(\Phi_{\ell_i}(\eta_{i-1}^N))\right](h_i)\right|^p\right\}^{\frac{1}{p}}\leq\frac{c^{1/p}(p)}{\sqrt{N}}q_i^{1/p}.$$

and incorporating this error bound throughout the rest of the proof of Theorem 2.

Corollary 3. *Under the same assumptions as Theorem 2, we have for $p\in\mathbb{N}$ and $h_t\in\mathrm{Osc}_1(E_t)$ the time uniform estimate*

$$\sup_{t\geq 0}\mathbb{E}\left\{|[\eta_t^N-\eta_t](h_t)|^p\right\}^{1/p}\leq\frac{\epsilon_{u,m}c^{1/p}(p)}{\sqrt{N}}\left(q_u\chi^{p/2}+(1-q_u)\right)^{1/p} \tag{20}$$

Proof: Starting with (16), we perform a different error decomposition:

$$\begin{aligned}
&\mathbb{E}\left\{\left|\left[\eta_i^N - \Phi_{\ell_i}(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{\frac{1}{p}} \\
&= \mathbb{E}\left\{\left|\delta_i\left[S^N \circ S^{N_\mathrm{b}}(\Phi_{\ell_i}(\eta_{i-1}^N)) - \Phi_{\ell_i}(\eta_{i-1}^N)\right](h_i)\right.\right. \\
&+ \left.\left.(1-\delta_i)\left[S^N(\Phi_{\ell_i}(\eta_{i-1}^N)) - \Phi_{\ell_i}(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{\frac{1}{p}} \\
&\leq \mathbb{E}\left\{\delta_i^p\left|\left[S^N \circ S^{N_\mathrm{b}}(\Phi_{\ell_i}(\eta_{i-1}^N)) - \Phi_{\ell_i}(\eta_{i-1}^N)\right](h_i)\right|^p\right. \\
&+ \sum_{k=1}^{p-1}\binom{p}{k}\delta_i^k(1-\delta_i)^{p-k}\left|\left[S^N \circ S^{N_\mathrm{b}}(\Phi_{\ell_i}(\eta_{i-1}^N)) - \Phi_{\ell_i}(\eta_{i-1}^N)\right](h_i)\right|^k \\
&\times \left|\left[S^N(\Phi_{\ell_i}(\eta_{i-1}^N)) - \Phi_{\ell_i}(\eta_{i-1}^N)\right](h_i)\right|^{p-k} \\
&+ \left.(1-\delta_i)^p\left|\left[S^N(\Phi_{\ell_i}(\eta_{i-1}^N)) - \Phi_{\ell_i}(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{1/p}.
\end{aligned} \tag{21}$$

We observe that $\delta_i(1-\delta_i) = 0$ and that if $N = \chi N_\mathrm{b}$ for integer χ, we can reconstruct an N-sample representation from the N_b sample with no additional error. Thus:

$$\begin{aligned}
\mathbb{E}\left\{\left|\left[\eta_i^N - \Phi_i(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{\frac{1}{p}} \leq \mathbb{E}\left\{\delta_i\left|\left[S^{N_\mathrm{b}}(\Phi_i(\eta_{i-1}^N)) - \Phi_i(\eta_{i-1}^N)\right](h_i)\right|^p\right. \\
\left.+ (1-\delta_i)\left|\left[S^N(\Phi_i(\eta_{i-1}^N)) - \Phi_i(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{1/p}.
\end{aligned}$$

Applying the same conditioning as in (18) and utilizing Lemma 1

$$\begin{aligned}
\mathbb{E}\left\{\left|\left[\eta_i^N - \Phi_i(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{\frac{1}{p}} &\leq \left(\frac{q_i c(p)}{N_\mathrm{b}^{p/2}} + \frac{(1-q_i)c(p)}{N^{p/2}}\right)^{1/p} \\
&= \frac{c(p)^{1/p}}{\sqrt{N}}\left(q_i\chi^{p/2} + (1-q_i)\right)^{1/p}
\end{aligned} \tag{22}$$

We note that $\chi \geq 1$ and $q_i\chi + (1-q_i) \leq q_\mathrm{u}\chi + (1-q_\mathrm{u})$ under the assumption $q_i \leq q_\mathrm{u}$. The final step in the proof involves applying (19) as in the proof of Theorem 2. ∎

The intuitive implication of Theorem 2 and Corollary 3 is that rare approximation events have limited effect on the average error performance of the subsample approximation particle filter. The L_2 error bound for the standard particle filter is the same as (20) of Corollary 3 taken with $p = 2$, except for the term $(q_\mathrm{u}\chi + (1-q_\mathrm{u}))^{1/2}$. This expression thus quantifies the performance deterioration, in terms of L_2 error bounds, due to the subsample approximation step. If the compression factor, χ, is $\chi = 10$, and subsample approximations occur with probability 0.1, then the deterioration of the root mean-square performance captured, in terms of bounds, by the factor $(0.1 \times 10 + (1-0.1))^{1/2}$ is around 40%. The communication overhead, on the other hand, represented by the total number of particles transmitted during leader node hand-off, is reduced by a factor of 10. The compressed particle cloud exchanges are most efficient in scenarios where the targets being tracked have slow dynamics and the density of leader nodes is relatively low (both implying rare hand-off events), but the tracking accuracy requirements and leader-to-leader communication costs are high.

Example: leader node particle filter: If the compression factor, χ, is $\chi = 10$, and subsample approximations occur with probability 0.1, then the deterioration of the root mean-square performance deterioration captured, in terms of bounds, by the factor $(0.1 \times 10 + (1-0.1))^{1/2}$, is around 40%. The communication overhead, on the other hand, represented by the total number of particles transmitted during leader node hand-off, is reduced by a factor of 10. The compressed particle cloud exchanges are most efficient in scenarios where the targets being tracked have slow dynamics and the density of leader nodes is relatively low (both implying rare hand-off events), but the tracking accuracy requirements and leader-to-leader communication costs are high.

Theorem 3 below provides the exponential estimate for the probability of large deviations of the approximate Feynman-Kac flows associated with the subsample approximation particle filter. Before proceeding to Theorem 3 we state a technical lemma.

Lemma 3. *Let X and Y be real random variables taking values in $\mathcal{X} \subseteq \mathbb{R}$ and $\mathcal{Y} \subseteq \mathbb{R}$ and let the joint distribution of these variables be $P_{X,Y}(d(x,y))$. Then for any $\varepsilon \in \mathbb{R}$ we have:*

$$\mathbb{P}\{X+Y \geq \varepsilon\} \leq \mathbb{P}\{X \geq \varepsilon/2\} + \mathbb{P}\{Y \geq \varepsilon/2\}$$

Proof: Let us define subsets $\mathcal{X}_{x\geq y} \subseteq \mathcal{X}$, $\mathcal{X}_{x\geq y} = \{x \in \mathcal{X} : x \geq y, y \in \mathcal{Y}\}$ and $\mathcal{X}_{x<y} = \mathcal{X}^c_{x\geq y}$, $\mathcal{X}_{x<y} = \{x \in \mathcal{X} : x < y, y \in \mathcal{Y}\}$. Denote by $\mathbf{1}_{\text{cond}}$ the indicator function, taking value 1 where the condition cond holds and 0 elsewhere. Now write the explicit expression for $\mathbb{P}\{X+Y \geq \varepsilon\}$:

$$\begin{aligned}\mathbb{P}\{X+Y \geq \varepsilon\} &= \int_{\mathcal{Y}}\int_{\mathcal{X}} \mathbf{1}_{x+y\geq\varepsilon} P_{X,Y}(\mathrm{d}(x,y)) \\ &= \int_{\mathcal{Y}}\int_{\mathcal{X}_{x\geq y}} \mathbf{1}_{x+y\geq\varepsilon} P_{X,Y}(\mathrm{d}(x,y)) + \int_{\mathcal{Y}}\int_{\mathcal{X}_{x<y}} \mathbf{1}_{x+y\geq\varepsilon} P_{X,Y}(\mathrm{d}(x,y))\end{aligned}$$

Since $\mathbf{1}_{x+y\geq\varepsilon} \leq \mathbf{1}_{2x\geq\varepsilon}$ on $\mathcal{X}_{x\geq y}$ and $\mathbf{1}_{x+y\geq\varepsilon} \leq \mathbf{1}_{2y\geq\varepsilon}$ on $\mathcal{X}_{x<y}$ we have

$$\begin{aligned}\mathbb{P}\{X+Y \geq \varepsilon\} &\leq \int_{\mathcal{Y}}\int_{\mathcal{X}_{x\geq y}} \mathbf{1}_{x\geq\varepsilon/2} P_{X,Y}(\mathrm{d}(x,y)) + \int_{\mathcal{Y}}\int_{\mathcal{X}_{x<y}} \mathbf{1}_{y\geq\varepsilon/2} P_{X,Y}(\mathrm{d}(x,y)) \\ &\leq \int_{\mathcal{Y}}\int_{\mathcal{X}} \mathbf{1}_{x\geq\varepsilon/2} P_{X,Y}(\mathrm{d}(x,y)) + \int_{\mathcal{Y}}\int_{\mathcal{X}} \mathbf{1}_{y\geq\varepsilon/2} P_{X,Y}(\mathrm{d}(x,y)),\end{aligned}$$

and the claim of lemma follows. ■

Theorem 3. *Suppose assumptions $(G)_u$ and $(M)_u^{(m)}$ hold. Suppose further that $\mathbb{P}\{\delta_i = 1\} \leq q_u$ for $i \geq 0$ and $0 \leq q_u \leq 1$. Then for any $N_b < N$, $t \geq 0$ and $h_t \in \mathrm{Osc}_1(E_t)$ we have*

$$\begin{aligned}\sup_{t\geq 0}\ \mathbb{P}\left\{|[\eta_t^N - \eta_t](h_t)| \geq \epsilon\right\} \leq &\left(1 + 4\sqrt{2\pi}\frac{\varepsilon\sqrt{N}}{\epsilon_{u,m}}\right) e^{-\frac{N\varepsilon^2}{2\epsilon^2_{u,m}}} \\ &+ q_u\left(1 + 4\sqrt{2\pi}\frac{\varepsilon\sqrt{N_b}}{\epsilon_{u,m}}\right) e^{-\frac{N_b\varepsilon^2}{2\epsilon^2_{u,m}}}\end{aligned}$$

Proof: Using the triangle inequality in (8) we have

$$\left|[\eta_t^N - \eta_t](h_t)\right| \leq \sum_{i=0}^{t} \left|\left[\Phi_{i,t}(\eta_i^N) - \Phi_{i,t}(\Phi_i(\eta_{i-1}^N))\right](h_i)\right|$$

Following the methodology presented in Theorem 2 and denoting $\omega_i = \left(1 - \epsilon_{\mathrm{u}}^2(M)\epsilon_{\mathrm{u}}^{(m-1)}(G)\right)^{\lfloor (t-i)/m \rfloor}$ and $a = \frac{2-\epsilon_{\mathrm{u}}(M)\epsilon_{\mathrm{u}}^m(G)}{\epsilon_{\mathrm{u}}(M)\epsilon_{\mathrm{u}}^m(G)}$ we have:

$$\left|[\eta_t^N - \eta_t](h_t)\right| \leq a\sum_{i=0}^{t} \omega_i \left|\left[\eta_i^N - \Phi_i(\eta_{i-1}^N)\right](h_i)\right|.$$

Using the structure of the algorithm defined in (24) and the definition of sampling operator introduced in (4) we obtain the following (similarly to Theorem 2):

$$\begin{aligned}\left|[\eta_t^N - \eta_t](h_t)\right| \leq\ & a\sum_{i=0}^{t} \omega_i\delta_i \left|\left[S^N \circ S^{N_\mathrm{b}}(\Phi_i(\eta_{i-1}^N)) - S^{N_\mathrm{b}}(\Phi_i(\eta_{i-1}^N))\right](h_i)\right| \\ &+ a\sum_{i=0}^{t} \omega_i(1-\delta_i) \left|\left[S^N(\Phi_i(\eta_{i-1}^N)) - \Phi_i(\eta_{i-1}^N)\right](h_i)\right| \\ &+ a\sum_{i=0}^{t} \omega_i\delta_i \left|\left[S^{N_\mathrm{b}}(\Phi_i(\eta_{i-1}^N)) - \Phi_i(\eta_{i-1}^N)\right](h_i)\right| \\ &= Z_1 + Z_2,\end{aligned}$$

where

$$Z_1 = a\sum_{i=0}^{t}\omega_i\delta_i\left|\left[S^N\circ S^{N_\mathrm{b}}(\Phi_i(\eta_{i-1}^N)) - S^{N_\mathrm{b}}(\Phi_i(\eta_{i-1}^N))\right](h_i)\right| + a\sum_{i=0}^{t}\omega_i(1-\delta_i)\left|\left[S^N(\Phi_i(\eta_{i-1}^N)) - \Phi_i(\eta_{i-1}^N)\right](h_i)\right|$$

$$Z_2 = a\sum_{i=0}^{t}\omega_i\delta_i\left|\left[S^{N_\mathrm{b}}(\Phi_i(\eta_{i-1}^N)) - \Phi_i(\eta_{i-1}^N)\right](h_i)\right|$$

Noting that

$$\sup_{t\geq 0}\ \mathbb{P}\left\{|[\eta_t^N-\eta_t](h_t)|\geq\epsilon\right\}\leq\sup_{t\geq 0}\ \mathbb{P}\left\{Z_1+Z_2\geq\epsilon\right\}$$

and applying Lemma 3 we have:

$$\sup_{t\geq 0}\ \mathbb{P}\left\{|[\eta_t^N-\eta_t](h_t)|\geq\epsilon\right\}\leq\sup_{t\geq 0}\ \mathbb{P}\left\{Z_1\geq\epsilon/2\right\}+\sup_{t\geq 0}\ \mathbb{P}\left\{Z_2\geq\epsilon/2\right\}.$$

Now applying Markov inequality we conclude:

$$\begin{aligned}\sup_{t\geq 0}\ \mathbb{P}\left\{|[\eta_t^N-\eta_t](h_t)|\geq\epsilon\right\}&\leq\sup_{t\geq 0}\ \mathbb{P}\left\{e^{\tau_1 Z_1}\geq e^{\tau_1\epsilon/2}\right\}+\sup_{t\geq 0}\ \mathbb{P}\left\{e^{\tau_2 Z_2}\geq e^{\tau_2\epsilon/2}\right\}\\&\leq\sup_{t\geq 0}\ e^{-\tau_1\epsilon/2}\mathbb{E}\left\{e^{\tau_1 Z_1}\right\}+\sup_{t\geq 0}e^{-\tau_2\epsilon/2}\mathbb{E}\left\{e^{\tau_2 Z_2}\right\}\end{aligned}$$

Next we apply the exponential series expansion

$$\mathbb{E}\left\{e^{\tau_1 Z_1}\right\}=\sum_{n\geq 0}\frac{\tau_1^n}{n!}\mathbb{E}\{Z_1^n\} \tag{23}$$

and use the fact that according to the following conditioning argument and Lemma 1 we have

$$\begin{aligned}\mathbb{E}\{Z_1^n\}^{1/n}&=\left(\mathbb{E}\{Z_1^n|\delta_i=1\}\mathbb{P}\{\delta_i=1\}+\mathbb{E}\{Z_1^n|\delta_i=0\}\mathbb{P}\{\delta_i=0\}\right)^{1/n}\\&\leq a\sum_{i=0}^{t}\omega_i\mathbb{E}\left\{\left(\delta_i\left|\left[S^N\circ S^{N_\mathrm{b}}(\Phi_i(\eta_{i-1}^N)) - S^{N_\mathrm{b}}(\Phi_i(\eta_{i-1}^N))\right](h_i)\right|\right.\right.\\&\left.\left.+\ (1-\delta_i)\left|\left[S^N(\Phi_i(\eta_{i-1}^N)) - \Phi_i(\eta_{i-1}^N)\right](h_i)\right|\right)^n\right\}^{1/n}\\&=a\sum_{i=0}^{t}\omega_i\left(\mathbb{P}\{\delta_i=1\}\mathbb{E}\left\{\left|\left[S^N\circ S^{N_\mathrm{b}}(\Phi_i(\eta_{i-1}^N)) - S^{N_\mathrm{b}}(\Phi_i(\eta_{i-1}^N))\right](h_i)\right|^n|\delta_i=1\right\}\right.\\&\left.+\ \mathbb{P}\{\delta_i=0\}\mathbb{E}\left\{\left|\left[S^N(\Phi_i(\eta_{i-1}^N)) - \Phi_i(\eta_{i-1}^N)\right](h_i)\right|^n|\delta_i=0\right\}\right)^{1/n}\\&\leq a\sum_{i=0}^{t}\omega_i(q_i c(n)N^{-n/2}+(1-q_i)c(n)N^{-n/2})^{1/n}\\&=\frac{c^{1/n}(n)}{\sqrt{N}}a\sum_{i=0}^{t}\omega_i.\end{aligned}$$

Noting that $a\sum_{i=0}^{t}\omega_i\leq\epsilon_{\mathrm{u},m}$ we have:

$$\mathbb{E}\{Z_1^n\}\leq\epsilon_{\mathrm{u},m}^n c(n)N^{-n/2}$$

Substituting this into (23) and employing the same simplifications as in the proofs of Theorem 1 and Corollary 1 we obtain:

$$\begin{aligned} e^{-\varepsilon\tau_1/2}\mathbb{E}\left\{e^{\tau_1 Z_1}\right\} &\leq \sum_{n\geq 0}\left(\frac{\tau_1\epsilon_{\mathrm{u},m}}{\sqrt{N}}\right)^n \frac{c(n)}{n!}e^{-\varepsilon\tau_1/2} \\ &\leq \left[1+\frac{\tau_1\epsilon_{\mathrm{u},m}}{\sqrt{N}}+\sum_{n\geq 2}\left(\frac{\tau_1\epsilon_{\mathrm{u},m}}{\sqrt{2N}}\right)^n \frac{\Gamma(n/2)}{(n-1)!}\right]e^{-\varepsilon\tau_1/2} \\ &= \left(1+\sqrt{2\pi}\frac{\tau_1\epsilon_{\mathrm{u},m}}{\sqrt{N}}\right)e^{\frac{\tau_1^2\epsilon_{\mathrm{u},m}^2}{8N}-\varepsilon\tau_1/2} \end{aligned}$$

Choosing $\tau_1 = \frac{2\varepsilon N}{\epsilon_{\mathrm{u},m}^2}$ we have

$$e^{-\varepsilon\tau_1/2}\mathbb{E}\left\{e^{\tau_1 Z_1}\right\} \leq \left(1+4\sqrt{2\pi}\frac{\varepsilon\sqrt{N}}{\epsilon_{\mathrm{u},m}}\right)e^{-\frac{N\varepsilon^2}{2\epsilon_{\mathrm{u},m}^2}}$$

Similar analysis yields

$$e^{-\varepsilon\tau_2/2}\mathbb{E}\left\{e^{\tau_2 Z_2}\right\} \leq q_{\mathrm{u}}\left(1+\sqrt{2\pi}\frac{\tau_2\epsilon_{\mathrm{u},m}}{\sqrt{N_{\mathrm{b}}}}\right)e^{\frac{\tau_2^2\epsilon_{\mathrm{u},m}^2}{8N_{\mathrm{b}}}-\varepsilon\tau_2/2},$$

which after choosing $\tau_2 = \frac{2N_{\mathrm{b}}\varepsilon}{\epsilon_{\mathrm{u},m}^2}$ results in:

$$e^{-\varepsilon\tau_2/2}\mathbb{E}\left\{e^{\tau_2 Z_2}\right\} \leq q_{\mathrm{u}}\left(1+4\sqrt{2\pi}\frac{\varepsilon\sqrt{N_{\mathrm{b}}}}{\epsilon_{\mathrm{u},m}}\right)e^{-\frac{N_{\mathrm{b}}\varepsilon^2}{2\epsilon_{\mathrm{u},m}^2}}.$$

This completes the proof. ■

V. Particle Filtering with Intermittent Parametric Approximations

In this section we analyze the error behaviour of a particle filter that incorporates intermittent parametric mixture estimation of the filtering density. The probability density estimation problem consists of estimating an unknown probability density given the i.i.d. sample $\{\xi_i\}_{i=1}^N$ from this density. As before, let $(E,\mathcal{E})$ be a measurable space. Denote λ a σ-finite measure on $\mathcal{E}$. Throughout this section it is assumed that the underlying distribution has a density if its Radon-Nikodym derivative with respect to λ exists.

We assume that with the sequence of the approximate filtering distributions, $\Phi_i(\eta_{i-1}^N)(\mathrm{d}x_i)$, there exists an associated and well-behaved sequence of approximate filtering densities $\frac{1}{\mathrm{d}x_i}\Phi_i(\eta_{i-1}^N)(\mathrm{d}x_i)$ so that the mixture density estimation problem is well-defined. The main result of the section, constituted in Theorem 6, is a time-uniform, weak-sense L_p error bound characterizing the expected behaviour of the parametric approximation leader node particle filter.

A. Parametric Approximation Particle Filter Algorithm

For this algorithm, the binary variable δ_t now indicates whether a parametric approximation is performed at time-step t. Again we assume that it is the outcome of a decision function based on the set of particles $\{\xi_{t-1}^k\}_{k=1}^N$ and observations Y_{t-1} at the previous time-step. Denote by $\mathbb{W}_{N_{\mathrm{p}}} : \mathcal{P}(E) \to \mathcal{P}(E^{N_{\mathrm{p}}})$ an operator that represents a parametric mixture approximation procedure that involves N_{p} mixture components (we will provide a concrete example below). The parametric approximation particle filter can then be expressed as:

$$\begin{aligned} \eta_t^N = S^N \circ \mathbb{W}_{N_{\mathrm{p}}}(\Phi_t(\eta_{t-1}^N)) \quad \text{if } \delta_t = 1, \\ \eta_t^N = S^N(\Phi_t(\eta_{t-1}^N)) \quad \text{if } \delta_t = 0 \end{aligned} \tag{24}$$

Example: parametric approximation leader node particle filter

When it employs parametric approximation, the leader node particle filter can be represented as follows.

$$\begin{aligned} \Phi_{\ell_t}(\eta_{t-1}^N) \Rightarrow \eta_t^N \Rrightarrow \widehat{\eta}_t^{N_{\mathrm{p}}} \longrightarrow \widehat{\eta}_t^{N_{\mathrm{p}}} \Rightarrow \eta_t^N \quad \text{if } \delta_t = 1, \\ \Phi_{\ell_t}(\eta_{t-1}^N) \Rightarrow \eta_t^N \quad \text{if } \delta_t = 0 \end{aligned}$$

Here the $\Rrightarrow$ represents the local distribution parametric approximation process and N_{p} is the number of the components in the mixture. As before, $\Rightarrow$ represents an N-particle sampling operation and $\longrightarrow$ represents communication between leader nodes. Thus the second particle filter we define relies upon a parametric approximation of the distribution $\Phi_i(\eta_{i-1}^N)$ based on the current particle set (a sample from this distribution).

B. Parametric Approximation

Within the GML framework proposed by Li and Barron [19] the discrepancy between the target density f and its estimate is measured by the Kullback-Leibler (KL) divergence. For any two measures ν and μ on E KL-divergence can be defined as follows:

$$D(\nu||\mu) = \int_E \log \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \mathrm{d}\nu \tag{25}$$

We will also abuse notation by writing KL-divergence for two arbitrary densities f and g in a similar fashion:

$$D(f||g) = \int_E \log \frac{f(x)}{g(x)} f(x) \mathrm{d}\lambda(x) \tag{26}$$

Consider the following class of bounded parametric densities:

$$\mathcal{H}_i = \left\{ \phi_{\theta_i}(x) : \theta_i \in \Theta_i, a_i \leq \inf_{\theta_i, x_i} \phi_{\theta_i}(x_i), \sup_{\theta_i, x_i} \phi_{\theta_i}(x_i) \leq b_i \right\}$$

where $0 < a_i < b_i < \infty$ and $\Theta_i \subset \mathbb{R}^{d_i}$ defines the parameter space, and $\inf$ and $\sup$ are taken over Θ_i and E_i. In the setting where the intermittent approximation is accomplished using parametric approximation, we are looking for a sequence of mixture density estimators of the filtering densities. We thus define the class of bounded parametric densities, $\phi_{\theta_i}(x)$, indexing it by time-step i to emphasize that the parameterization can be time-varying. The approximation is restricted to a class of discrete N_{p}-component convex combinations of the form:

$$\mathcal{C}_{N_{\mathrm{p}},i} = \mathrm{conv}_{N_{\mathrm{p}}}(\mathcal{H}_i) = \left\{ g : g(x) = \sum_{j=1}^{N_{\mathrm{p}}} \alpha_{i,j} \phi_{\theta_j}(x), \phi_{\theta_j} \in \mathcal{H}_i, \sum_{j=1}^{N_{\mathrm{p}}} \alpha_{i,j} = 1,\ \alpha_{i,j} \geq 0 \right\}$$

As N_{p} grows without bound, $\mathcal{C}_{N_{\mathrm{p}},i}$ converges to the class of continuous convex combinations:

$$\mathcal{C}_i = \mathrm{conv}(\mathcal{H}_i) = \left\{ g : g(x) = \int_\Theta \phi_\theta(x) \mathbb{P}_i(\mathrm{d}\theta), \phi_\theta \in \mathcal{H}_i \right\}$$

The general framework for the greedy approximation of arbitrary cost functions is discussed in [28]. The particular instance of this more general framework is the Greedy Maximum Likelihood (GML) for mixture approximation (see [19]). The corresponding computational routine, a sequential greedy maximum likelihood, associated with this procedure is summarized in the form of Algorithm 1.

Algorithm 1: GML

Given $g_1 \in \mathcal{H}$
for $k = 2$ *to* N_p **do**
    Find $\phi_{\theta_k} \in \mathcal{H}$ and $0 \leq \alpha_k \leq 1$ to maximize the function:
    $(\theta_k^*, \alpha_k^*) = \arg\max_{\alpha_k, \theta_k} \sum_{j=1}^{N} \log((1-\alpha_k) g_{k-1}(x_j) + \alpha_k \phi_{\theta_k}(x_j))$
    Let $g_k = (1-\alpha_k^*) g_{k-1} + \alpha_k^* \phi_{\theta_k^*}$
endfor

To link the greedy maximum likelihood maximization in Algorithm 1 to the minimization of KL-divergence we recall that if ν is a known distribution and μ is the KL-based fit to this distribution, the KL-divergence minimization

problem can be written as follows (assuming that the corresponding densities exist):

$$\begin{aligned}\min_{\mu\in\mathcal{P}(E)} D(\nu||\mu) &= \min_{\mu\in\mathcal{P}(E)} \int_E \log\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\mathrm{d}\nu \\ &= \min_{\mu\in\mathcal{P}(E)} \left[\int_E \log\frac{\mathrm{d}\nu}{\mathrm{d}\lambda(x)}\mathrm{d}\nu - \int_E \log\frac{\mathrm{d}\mu}{\mathrm{d}\lambda(x)}\mathrm{d}\nu\right] \\ &= \max_{\mu\in\mathcal{P}(E)} \int_E \log\frac{\mathrm{d}\mu}{\mathrm{d}\lambda(x)}\mathrm{d}\nu\end{aligned}$$

In practice ν itself is unknown, but a sample from this distribution may be available. The approximation of the true expectation with respect to ν above by the expectation with respect to its empirical counterpart, $S^N(\nu)$, leads to the maximum likelihood density estimator:

$$\begin{aligned}\min_{\mu\in\mathcal{P}(E)} D(\nu||\mu) &= \max_{\mu\in\mathcal{P}(E)} \mathbb{E}_\nu \log\frac{\mathrm{d}\mu}{\mathrm{d}\lambda(x)} \\ &\approx \max_{\mu\in\mathcal{P}(E)} \mathbb{E}_{S^N(\nu)} \log\frac{\mathrm{d}\mu}{\mathrm{d}\lambda(x)} \\ &= \max_{\mu\in\mathcal{P}(E)} \sum_{i=1}^{N} \log\frac{\mathrm{d}\mu}{\mathrm{d}\lambda(x)}(x_i)\end{aligned}$$

Thus the error committed by resorting to the suboptimal GML algorithm consists of two contributions.

First, there is the error associated with the limitations of the approximation class $\mathcal{C}$: even the best possible $\mu \in \mathcal{C}$ will have non-zero $D(\nu||\mu)$ if $\nu \notin \mathcal{C}$. We will call this error the approximation error. Second, there is the combined error associated with the greedy optimization and the approximation of the true expectation by its empirical counterpart — we will call this the estimation error. In the following we analyze these errors for the one-step approximation and then link the results of the analysis to the overall error of the parametric approximation particle filter.

C. Local Error Analysis

The attractive features of Algorithm 1 are threefold. First, the algorithm simplifies the ML density estimation procedure. Instead of facing the N_p-mixture estimation problem we only have to solve N_p 2-mixture estimation problems [19]. Second, there are several bounds on approximation and sampling errors of Algorithm 1 in terms of KL-divergence (see [19] and [23]). In this section we extend the existing results and perform the L_p error analysis. Third, it was shown [19] that the performance of the greedy algorithm converges to the performance of the optimal mixture estimation algorithm as N and N_p become large. Thus if these conditions hold, the results obtained for the parametric approximation particle filter that uses GML are also applicable if other density estimators are employed.

Here we state the relevant results from [19] that will be of use in further analysis. The following notation is introduced to facilitate presentation. Assuming that f is a target density and $g \in \mathcal{C}$ we denote $D(f||\mathcal{C}) = \inf_{g\in\mathcal{C}} D(f||g)$, the least possible divergence between a target density, f, and a member g from the class of continuous convex combinations $\mathcal{C}$. Furthermore, assuming that the target density f is known, the analytical estimator $g^{N_\mathrm{p}} \in \mathcal{C}_{N_\mathrm{p}}$ can be obtained by solving the following greedy recursion for $i = 2 \dots N_\mathrm{p}$ (see Algorithm 1):

$$(\theta_k^*, \alpha_k^*) = \arg\max_{\alpha_k,\theta_k} \int_{x\in E} \log((1-\alpha_k)g_{k-1}(x) + \alpha_k\phi_{\theta_k}(x))f(x)\mathrm{d}x.$$

Alternatively, $\widehat{g}^{N_\mathrm{p}} \in \mathcal{C}_{N_\mathrm{p}}$ is an empirical N_p-mixture estimator constructed using Algorithm 1 based on a sample from the target density, f. The following theorem (see [19]) reveals an important general property of the GML algorithm.

Theorem 4 (Li and Barron [19], Theorem 2)**.** *For every $g_\mathcal{C}(x) \in \mathcal{C}$*

$$D(f||g^{N_p}) \le D(f||g_\mathcal{C}) + \frac{\gamma c_{f,\mathcal{C}}^2}{N_p}.$$

Here,

$$c_{f,\mathcal{C}}^2 = \int \frac{\int_\Theta \phi_\theta^2(x)\mathbb{P}(\mathrm{d}\theta)}{(\int_\Theta \phi_\theta(x)\mathbb{P}(\mathrm{d}\theta))^2} f(x)\mathrm{d}x,$$

and $\gamma = 4[\log(3\sqrt{e}) + \sup_{\theta_1,\theta_2\in\Theta, x\in E} \log(\phi_{\theta_1}(x)/\phi_{\theta_2}(x))]$

One of the consequences [19] of Theorem 4 is the following relationship between an arbitrary $g_{\mathcal{C}}(x) \in \mathcal{C}$ and the empirical GML algorithm output $\widehat{g}^{N_\mathrm{p}} \in \mathcal{C}_{N_\mathrm{p}}$:

$$\frac{1}{N}\sum_{i=1}^N \log \widehat{g}^{N_\mathrm{p}}(x_i) \geq \frac{1}{N}\sum_{i=1}^N \log g_{\mathcal{C}}(x_i) - \frac{\gamma c_{f,\mathcal{C}}^2}{N_\mathrm{p}}. \tag{27}$$

Clearly, it also follows directly from Theorem 4 that $D(f||g^{N_\mathrm{p}}) \leq D(f||\mathcal{C}) + \frac{\gamma c_{f,\mathcal{C}}^2}{N_\mathrm{p}}$. Thus Theorem 4 establishes a strong formal argument that shows that the greedy density estimate converges to the best possible estimate as N_p grows without bound. Similar results for the empirical estimator $\widehat{g}^{N_\mathrm{p}}$ appear in [23] and [19].

Our next goal is to connect the existing results on the performance of the GML in terms of the KL-divergence to its performance in terms of L_p error metric. Our next result reveals the L_p error bound characterizing the average performance of the GML algorithm. One of the components of the bound is the packing number $\mathcal{D}(\varepsilon, \mathcal{H}, d_N)$, which is the the maximum number of ε-separated points in $\mathcal{H}$ (the class of parametric density functions) under the empirical semimetric d_N. The empirical semimetric is defined for $h_1, h_2 \in \mathcal{H}$ as

$$d_N^2(h_1,h_2) = \frac{1}{N}\sum_{k=1}^N (h_1(x_k) - h_2(x_k))^2.$$

Theorem 5. *Suppose $\widehat{g}^{N_p} \in \mathcal{C}_{N_p}$ is constructed using Algorithm 1 and $\widehat{\mathcal{G}}^{N_p} \in \mathcal{P}(E)$ is the distribution associated with $\widehat{g}^{N_p}$. Suppose further that there exists density f associated with the target distribution $F \in \mathcal{P}(E)$. Then for any $h \in \mathcal{B}_b(E)$ with $||h||_{\mathrm{osc}} \leq 1$, $p \geq 1$, and $N, N_p \in \mathbb{N}$ we have:*

$$\mathbb{E}\left\{|[\widehat{\mathcal{G}}^{N_p} - F](h)|^p\right\}^{1/p} \leq \sqrt{2}\left[\frac{8}{a\sqrt{N}}\left(2c^{2/p}(p/2)\right.\right.$$
$$\left.\left.+(p/4)!C\mathbb{E}\int_0^b \sqrt{\log\left(1+\mathcal{D}(\varepsilon,\mathcal{H},d_N)\right)}d\varepsilon\right) + \frac{\gamma c_{f,\mathcal{C}}^2}{N_p} + D(f||\mathcal{C})\right]^{1/2}$$

where C is a universal constant[2].

Proof: Using Pinsker's inequality, $\int |f-g| \leq \sqrt{2D(f||g)}$, [7] we have

$$\begin{aligned}
\mathbb{E}\left\{|[\widehat{\mathcal{G}}^{N_\mathrm{p}} - F](h)|^p\right\}^{1/p} &= \mathbb{E}\left\{\left(\int_E [\widehat{g}^{N_\mathrm{p}}(x) - f(x)]h(x)\mathrm{d}x\right)^p\right\}^{1/p} \\
&\leq ||h||\mathbb{E}\left\{\left(\int_E |\widehat{g}^{N_\mathrm{p}}(x) - f(x)|\mathrm{d}x\right)^p\right\}^{1/p} \\
&\leq \mathbb{E}\left\{\left(\sqrt{2D(f||\widehat{g}^{N_\mathrm{p}})}\right)^p\right\}^{1/p} \\
&= \sqrt{2}\left[\mathbb{E}\left\{D(f||\widehat{g}^{N_\mathrm{p}})^{p/2}\right\}^{2/p}\right]^{1/2}
\end{aligned}$$

Now, suppose $p \geq 2$. The following decomposition can be used to analyze the previous expression:

$$D(f||\widehat{g}^{N_\mathrm{p}}) = D(f||\widehat{g}^{N_\mathrm{p}}) - D(f||\mathcal{C}) + D(f||\mathcal{C})$$

[2]See [25] for details.

Denoting $g^* = \arg\min_{g\in\mathcal{C}} D(f||g)$ we have the following modification of the decomposition proposed by Rakhlin et al. in [23]:

$$\begin{aligned} D(f||\widehat{g}^{N_\mathrm{p}}) - D(f||\mathcal{C}) &= -\int \log \widehat{g}^{N_\mathrm{p}}(x) F(\mathrm{d}x) + \int \log g^*(x) F(\mathrm{d}x) \\ &= -\int \log \widehat{g}^{N_\mathrm{p}}(x) F(\mathrm{d}x) + \frac{1}{N}\sum_{i=1}^{N} \log \widehat{g}^{N_\mathrm{p}}(x_i) \\ &+ \frac{1}{N}\sum_{i=1}^{N} \log g^*(x_i) - \frac{1}{N}\sum_{i=1}^{N} \log \widehat{g}^{N_\mathrm{p}}(x_i) \\ &+ \int \log g^*(x) F(\mathrm{d}x) - \frac{1}{N}\sum_{i=1}^{N} \log g^*(x_i) \end{aligned}$$

Applying (27) to the middle term we see:

$$D(f||\widehat{g}^{N_\mathrm{p}}) - D(f||\mathcal{C}_i) \leq |[F - S^N(F)](\log \widehat{g}^{N_\mathrm{p}})| + |[F - S^N(F)](\log g^*)| + \frac{\gamma c_{f,\mathcal{C}}^2}{N_\mathrm{p}}$$

By the definition of $D(f||\mathcal{C})$ it follows that $D(f||\widehat{g}^{N_\mathrm{p}}) - D(f||\mathcal{C}) \geq 0$ and thus we conclude:

$$\left|D(f||\widehat{g}^{N_\mathrm{p}}) - D(f||\mathcal{C})\right| \leq 2\sup_{g\in\mathcal{C}} \left|[F - S^N(F)]\,(\log g)\right| + \frac{\gamma c_{f,\mathcal{C}}^2}{N_\mathrm{p}}$$

This allows splitting the effect of approximation and sampling errors by applying Minkowski's inequality (since $p \geq 2$):

$$\begin{aligned} \mathbb{E}\left\{D(f||\widehat{g}^{N_\mathrm{p}})^{p/2}\right\}^{2/p} &= \mathbb{E}\left\{\left|D(f||\widehat{g}^{N_\mathrm{p}}) - D(f||\mathcal{C}) + D(f||\mathcal{C})\right|^{p/2}\right\}^{2/p} \\ &\leq 2\mathbb{E}\left\{\left[\sup_{g\in\mathcal{C}} \left|[F - S^N(F)]\,(\log g)\right|\right]^{p/2}\right\}^{2/p} + \frac{\gamma c_{f,\mathcal{C}}^2}{N_\mathrm{p}} + D(f||\mathcal{C}). \end{aligned}$$

The next step of the proof makes use of a symmetrization argument. We recall the definition of the Rademacher sequence (ε_k) as a sequence of independent random variables taking values in $\{-1, +1\}$ with $\mathbb{P}\{\varepsilon_k = 1\} = \mathbb{P}\{\varepsilon_k = -1\} = 1/2$. Denote by S_ε^N the generator of the signed Rademacher measure (with x_k being the samples from μ):

$$S_\varepsilon^N(\mu) = \frac{1}{N}\sum_{k=1}^{N} \varepsilon_k \delta_{x_k}.$$

Using the symmetrization lemma (see e.g. Lemma 2.3.1 in [25] or Lemma 6.3 in [18]) we deduce:

$$\mathbb{E}\left\{\left[\sup_{g\in\mathcal{C}} \left|[F - S^N(F)]\,(\log g)\right|\right]^{p/2}\right\}^{2/p} \leq 2\mathbb{E}\left\{\left[\sup_{g\in\mathcal{C}} \left|S_\varepsilon^N(F)\,(\log g)\right|\right]^{p/2}\right\}^{2/p}$$

Denoting $\kappa(x) = g(x) - 1$ and using the fact [22] that $\varphi(\kappa(x)) = a\log(\kappa(x) + 1)$ is a contraction[3], we apply the comparison inequality (Theorem 4.12 in [18]), observing that $[\cdot]^{p/2}$ is convex and increasing for $p \geq 2$ and κ is a

[3]The function $\varphi : \mathbb{R} \to \mathbb{R}$ is a contraction if we have $|\varphi(x) - \varphi(y)| \leq |x - y|, \forall x, y \in E$.

bounded function:

$$\mathbb{E}\left\{\left[\sup_{g\in\mathcal{C}}\left|S_\varepsilon^N(F)\left(\log g\right)\right|\right]^{p/2}\right\}^{2/p} = \mathbb{E}\left\{\left[\frac{1}{2}\frac{2}{a}\sup_{g\in\mathcal{C}}\left|S_\varepsilon^N(F)\left(a\log(\kappa+1)\right)\right|\right]^{p/2}\right\}^{2/p}$$
$$\leq \frac{2}{a}\mathbb{E}\left\{\left[\sup_{g\in\mathcal{C}}\left|S_\varepsilon^N(F)\left(g-1\right)\right|\right]^{p/2}\right\}^{2/p}$$
$$\leq \frac{2}{a}\mathbb{E}\left\{\left[\sup_{g\in\mathcal{C}}\left|S_\varepsilon^N(F)(g)\right|\right]^{p/2}\right\}^{2/p} + \frac{2}{a}\mathbb{E}\left\{\left|S_\varepsilon^N(F)(1)\right|^{p/2}\right\}^{2/p}$$

Using Lemma 1 we have:

$$\mathbb{E}\left\{\left|S_\varepsilon^N(F)(1)\right|^{p/2}\right\}^{2/p} = \mathbb{E}\left\{\left|\frac{1}{N}\sum_{i=1}^N\varepsilon_i\right|^{p/2}\right\}^{2/p} \leq \frac{2c^{2/p}(p/2)}{\sqrt{N}}$$

On the other hand, we have for any $g\in\mathcal{C}$ a corresponding $\phi_\theta\in\mathcal{H}$:

$$\begin{aligned}
|S_\varepsilon^N(F)(g)| &= \left|\frac{1}{N}\sum_{i=1}^N\varepsilon_i g(x_i)\right| \\
&= \left|\frac{1}{N}\sum_{i=1}^N\varepsilon_i\int_{\theta\in\Theta}\phi_\theta(x_i)\mathbb{P}(\mathrm{d}\theta)\right| \\
&= \left|\int_{\theta\in\Theta}\frac{1}{N}\sum_{i=1}^N\varepsilon_i\phi_\theta(x_i)\mathbb{P}(\mathrm{d}\theta)\right| \\
&\leq \int_{\theta\in\Theta}\left|\frac{1}{N}\sum_{i=1}^N\varepsilon_i\phi_\theta(x_i)\right|\mathbb{P}(\mathrm{d}\theta) \\
&\leq \sup_{\theta\in\Theta}\left|\frac{1}{N}\sum_{i=1}^N\varepsilon_i\phi_\theta(x_i)\right|
\end{aligned}$$

Thus we have

$$\mathbb{E}\left\{\left[\sup_{g\in\mathcal{C}}|S_\varepsilon^N(F)(g)|\right]^{p/2}\right\}^{2/p} \leq \mathbb{E}\left\{\left[\sup_{g\in\mathcal{H}}|S_\varepsilon^N(F)(g)|\right]^{p/2}\right\}^{2/p}$$

The Orlicz norm [4], [25] $\pi_{\psi_p}(Y)$ of a random variable Y is defined, for a nondecreasing convex function $\psi_p(x)=e^{x^p}-1$, as

$$\pi_{\psi_p}(Y) = \inf\{C>0 : \mathbb{E}\{\psi_p(|Y|/C)\}\leq 1\}.$$

By Hoeffding's inequality the Rademacher process $S_\varepsilon^N(F)(g)$ is sub-Gaussian for the semimetric d_N [25]. Using the fact that $\mathbb{E}\{X^p\}^{1/p}\leq(p/2)!\pi_{\psi_2}(X)$ (see e.g. Lemma 7.3.5 in [4] or [25], p. 105, Problem 4) we deduce:

$$\mathbb{E}\mathbb{E}_\varepsilon\left\{\left[\sup_{g\in\mathcal{H}}|S_\varepsilon^N(F)(g)|\right]^{p/2}\right\}^{2/p} \leq (p/4)!\mathbb{E}\pi_{\psi_2}(\sup_{g\in\mathcal{H}}|S_\varepsilon^N(F)(g)|)$$

In addition, since $S_\varepsilon^N(F)(g)$ is sub-Gaussian, we have for some universal constant C (see Proof of Corollary 2.2.8 in [25]):

$$\mathbb{E}\pi_{\psi_2}(\sup_{g\in\mathcal{H}}|S_\varepsilon^N(F)(g)|) \leq \frac{C}{\sqrt{N}}\mathbb{E}\int_0^b\sqrt{\log\left(1+\mathcal{D}(\varepsilon,\mathcal{H},d_N)\right)}\mathrm{d}\varepsilon$$

Combining the above we have:

$$\mathbb{E}\left\{|[\widehat{\mathcal{G}}^{N_{\mathrm{p}}} - F](h)|^p\right\}^{1/p} \leq \sqrt{2}\left[\frac{8}{a\sqrt{N}}\left(2c^{2/p}(p/2)\right.\right.$$
$$\left.\left.+(p/4)!C\mathbb{E}\int_0^b \sqrt{\log\left(1+\mathcal{D}(\varepsilon,\mathcal{H},d_N)\right)}\mathrm{d}\varepsilon\right) + \frac{\gamma c_{f,\mathcal{C}}^2}{N_{\mathrm{p}}} + D(f||\mathcal{C})\right]^{1/2}$$

Finally, suppose $1 \leq p < 2$. In this case using Jensen's inequality we have:

$$\mathbb{E}\left\{D(f||\widehat{g}^{N_{\mathrm{p}}})^{p/2}\right\}^{2/p} \leq \mathbb{E}\left\{D(f||\widehat{g}^{N_{\mathrm{p}}})\right\}$$

Thus the above analysis applies if we choose $p = 2$ and the proof is now complete. ■

Corollary 4. *Suppose that the assumptions of Theorem 5 hold. Suppose in addition that $f \in \mathcal{C}$ then we have for any $p \geq 1$:*

$$\mathbb{E}\left\{|[\widehat{\mathcal{G}}^{N_p} - F](h)|^p\right\}^{1/p} \leq \sqrt{2}\left[\frac{8}{a\sqrt{N}}\left(2c^{2/p}(p/2)\right.\right.$$
$$\left.\left.+(p/4)!C\mathbb{E}\int_0^b \sqrt{\log\left(1+\mathcal{D}(\varepsilon,\mathcal{H},d_N)\right)}d\varepsilon\right) + 4\log(3\sqrt{e}(b/a))\frac{(b/a)^2}{N_p}\right]^{1/2}$$

Proof: The proof follows from the fact that under the additional assumption we have $D(f||\mathcal{C}) = 0$. Furthermore, we note that under this assumption $c_{f,\mathcal{C}}^2 \leq (b/a)^2$ and $\gamma = 4\log(3\sqrt{e}(b/a))$ ■

D. Time-uniform error bounds

In this section we present a result specifying time-uniform error bounds for particle filters performing intermittent parametric approximation. The result links the properties of Markov transitions $M_i(x_{i-1}, \mathrm{d}x_i)$ and error bounds for parametric GML approximation (Theorem 5) with the propagation of approximation errors through Feynman-Kac operators. It is based on the following observations. For an absolutely continuous Markov kernel with density $p_i(x_i|x_{i-1})$, we can write [4]:

$$M_i(x_{i-1}, \mathrm{d}x_i) = \Pr\{X_i \in \mathrm{d}x_i | X_{i-1} = x_{i-1}\} = p_i(x_i|x_{i-1})\mathrm{d}x_i = p_{\vartheta_i}(x_i)\mathrm{d}x_i,$$

where we explicitly assume that the structure of the kernel M_i can be captured by a set of parameters $\vartheta_i \in \Theta_i \subset \mathbb{R}^{d_i}$ (these parameters may include the state-value x_{i-1}). We can further define a class $\mathcal{M}_i$ of such densities:

$$\mathcal{M}_i = \left\{p_{\vartheta_i}(x_i) : \vartheta_i \in \Theta_i \subset \mathbb{R}^{d_i}\right\}.$$

Thus if M_i is such that $p_{\vartheta_i}(x_i) \in \mathcal{M}_i$ and $\mathcal{M}_i \subseteq \mathcal{H}_i$ then the assumption $(M)_{\mathrm{u}}^{(m)}$ is satisfied with $m = 1$ and $\epsilon_{\mathrm{u}}(M) = a/b$, yielding for any $x_{i-1}, y_{i-1} \in E_{i-1}$:

$$M_i(x_{i-1}, \cdot) \geq \frac{a}{b} M_i(y_{i-1}, \cdot)$$

Furthermore, using the definitions of the one-step Boltzman-Gibbs transformation and the associated Feynman-Kac operator we see that the distribution at time i is related to the distribution at time $i-1$ as follows:

$$\begin{aligned}\eta_i &= \Phi_i(\eta_{i-1}) = \Psi_{i-1}(\eta_{i-1})M_i \\ &= \int_{E_{i-1}} M_i(x_{i-1}, \mathrm{d}x_i)\Psi_{i-1}(\eta_{i-1})(\mathrm{d}x_{i-1}) \\ &= \int_{E_{i-1}} M_i(x_{i-1}, \mathrm{d}x_i)\frac{G_{i-1}(x_{i-1})\eta_{i-1}(\mathrm{d}x_{i-1})}{\eta_{i-1}(G_{i-1})}.\end{aligned}$$

Thus for an absolutely continuous Markov kernel with $p_{\vartheta_i}(x_i) \in \mathcal{M}_i$ we can rewrite the previous equation with a suitable change of measure:

$$\frac{\eta_i(\mathrm{d}x_i)}{\mathrm{d}x_i} = \int_{\Theta_i} p_{\vartheta_i}(x_i)\mathbb{P}(\mathrm{d}\vartheta_i).$$

This implies that for an N-particle approximation η_{i-1}^N we have that $\frac{\eta_i(\mathrm{d}x_i)}{\mathrm{d}x_i} \in \mathrm{conv}_N(\mathcal{M}_i)$ and, as N grows without bound, we have $\frac{\eta_i(\mathrm{d}x_i)}{\mathrm{d}x_i} \in \mathrm{conv}(\mathcal{M}_i)$. Thus the performance of GML approximation algorithm is determined by the properties of Markov transition kernel $M_i(x_{i-1}, \mathrm{d}x_i)$ and the class of approximating densities $\mathcal{H}_i$. In particular, for an absolutely continuous Markov kernel with $p_{\vartheta_i}(x_i) \in \mathcal{M}_i$ and a sufficiently rich class $\mathcal{H}_i$, such that $\mathcal{M}_i \subseteq \mathcal{H}_i$ we have asymptotically unbiased approximation:

$$D\left(\frac{\eta_i(\mathrm{d}x_i)}{\mathrm{d}x_i}\middle\|\mathcal{C}\right) = 0.$$

The preceding discussion can be summarized in the form of a concise assumption:

- $(\mathcal{H})_{\mathrm{u}}$: The Markov kernels associated with the target dynamics are absolutely continuous and can be expressed in the form $M_i(x_{i-1}, \mathrm{d}x_i) = p_{\vartheta_i}(x_i)\mathrm{d}x_i$. The class of densities associated with M_i is defined as $\mathcal{M}_i = \left\{p_{\vartheta_i}(x_i) : \vartheta_i \in \Theta_i \subset \mathbb{R}^{d_i}\right\}$. For each $\mathcal{M}_i$ there exists an approximation class $\mathcal{H}_i$ and strictly positive numbers $a_{\mathrm{u}} = \inf_{i\geq 0} a_i$, $b_{\mathrm{u}} = \sup_{i\geq 0} b_i$ satisfying $0 < a_{\mathrm{u}} < b_{\mathrm{u}} < \infty$ such that for any $i \geq 0$ we have

$$\mathcal{M}_i \subseteq \mathcal{H}_i \qquad \text{and hence} \qquad M_i(x_{i-1}, \cdot) \geq \frac{a_{\mathrm{u}}}{b_{\mathrm{u}}} M_i(y_{i-1}, \cdot)$$

The following result describes the analog of Theorem 2 for the case of a parametric approximation particle filter using the GML algorithm.

Theorem 6. *Suppose assumptions $(G)_u$ and $(\mathcal{H})_u$ hold. Suppose further that $\mathbb{P}\{\delta_i = 1\} \leq q_u$ for any $i \geq 0$ and $0 \leq q_u \leq 1$. Then for any $N_p, N \geq 1$, $t \geq 0$, $p \geq 1$ and $h_t \in \mathrm{Osc}_1(E_t)$ we have the time uniform bound*

$$\sup_{t\geq 0}\mathbb{E}\left\{|[\eta_t^N - \eta_t](h_t)|^p\right\}^{1/p} \leq \epsilon_u \left[\frac{c^{1/p}(p)}{\sqrt{N}} + q_u^{1/p}\left[\frac{16}{a\sqrt{N}}\left(2c^{2/p}(p/2)\right.\right.\right.$$
$$\left.\left.\left. + C(p/4)!\sup_{i\geq 0}\mathbb{E}\int_0^{b_i}\sqrt{\log\left(1+\mathcal{D}(\varepsilon, \mathcal{H}_i, d_N)\right)}d\varepsilon\right) + 8\log(3\sqrt{e}(b/a))\frac{(b/a)^2}{N_p}\right]^{1/2}\right]$$

where the constant ϵ_u is given by:

$$\epsilon_u = \frac{2 - (a_u/b_u)\epsilon_u(G)}{(a_u/b_u)^3\epsilon_u(G)}.$$

Proof: Using the same argument as in Theorem 2 we have

$$\mathbb{E}\left\{|[\eta_t^N - \eta_t](f_t)|^p\right\}^{1/p}$$
$$\leq \frac{2 - \epsilon_{\mathrm{u}}(M)\epsilon_{\mathrm{u}}(G)}{\epsilon_{\mathrm{u}}(M)\epsilon_{\mathrm{u}}(G)}\sum_{i=0}^{t}\left(1-\epsilon_{\mathrm{u}}^2(M)\right)^{(t-i)}\mathbb{E}\left\{\left|\left[\eta_i^N - \Phi_i(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{1/p}.$$

Based on the Minkowski inequality we have the decomposition for each individual expectation under the sum above:

$$\mathbb{E}\left\{\left|\left[\eta_i^N - \Phi_i(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{\frac{1}{p}}$$
$$\leq \mathbb{E}\left\{\left|\delta_i\left[S^N(\widehat{\mathcal{G}}^{N_{\mathrm{p}}}) - \widehat{\mathcal{G}}^{N_{\mathrm{p}}}\right](h_i) + (1-\delta_i)\left[S^N(\Phi_i(\eta_{i-1}^N)) - \Phi_i(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{\frac{1}{p}}$$
$$+ \mathbb{E}\left\{\left|\delta_i\left[\widehat{\mathcal{G}}^{N_{\mathrm{p}}} - \Phi_i(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{\frac{1}{p}}$$

Using the same conditioning argument as in Theorem 2 and applying Corollary 4 based on the assumption $(\mathcal{H})_{\mathrm{u}}$ to the second term we have:

$$\begin{aligned}
&\mathbb{E}\left\{\left|\delta_i\left[\widehat{\mathcal{G}}^{N_{\mathrm{p}}}-\Phi_i(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{1/p}\\
&=\mathbb{E}\left\{\mathbb{E}\left\{\delta_i\left|\left[\widehat{\mathcal{G}}^{N_{\mathrm{p}}}-\Phi_i(\eta_{i-1}^N)\right](h_i)\right|^p\middle|\mathcal{F}_{i-1},Y_i^{\mathcal{S}_{\ell_i}}=y_i^{\mathcal{S}_{\ell_i}}\right\}\right\}^{1/p}\\
&\leq q_i^{1/p}\sqrt{2}\left[\frac{8}{a_i\sqrt{N}}\left(2c^{2/p}(p/2)+(p/4)!C\mathbb{E}\int_0^{b_i}\sqrt{\log\left(1+\mathcal{D}(\varepsilon,\mathcal{H}_i,d_N)\right)}\mathrm{d}\varepsilon\right)\right.\\
&\quad\left.+\,4\log(3\sqrt{e}(b_i/a_i))\frac{(b_i/a_i)^2}{N_{\mathrm{p}}}\right]^{1/2}
\end{aligned}$$

Applying conditioning and Lemma 1 and the same conditioning argument as in Theorem 3 to the remaining term we have:

$$\begin{aligned}
&\mathbb{E}\left\{\left|\left[\eta_i^N-\Phi_i(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{1/p}\leq\frac{c^{1/p}(p)}{\sqrt{N}}\\
&+q_i^{1/p}\left(\sqrt{2}\left[\frac{8}{a_i\sqrt{N}}\left(2c^{2/p}(p/2)+(p/4)!C\mathbb{E}\int_0^{b_i}\sqrt{\log\left(1+\mathcal{D}(\varepsilon,\mathcal{H}_i,d_N)\right)}\mathrm{d}\varepsilon\right)\right.\right.\\
&\left.\left.+\,4\log(3\sqrt{e}(b_i/a_i))\frac{(b_i/a_i)^2}{N_{\mathrm{p}}}\right]^{1/2}\right)
\end{aligned}$$

We conclude that since $q_i\leq q_{\mathrm{u}}$ then for any $i\geq 0$ we have the time-uniform estimate:

$$\begin{aligned}
&\mathbb{E}\left\{\left|\left[\eta_i^N-\Phi_i(\eta_{i-1}^N)\right](h_i)\right|^p\right\}^{1/p}\leq\frac{c^{1/p}(p)}{\sqrt{N}}\\
&+q_{\mathrm{u}}^{1/p}\sqrt{2}\left[\frac{8}{a_{\mathrm{u}}\sqrt{N}}\left(2c^{2/p}(p/2)+(p/4)!C\sup_{i\geq 0}\mathbb{E}\int_0^{b_i}\sqrt{\log\left(1+\mathcal{D}(\varepsilon,\mathcal{H}_i,d_N)\right)}\mathrm{d}\varepsilon\right)\right.\\
&\left.+\,4\log(3\sqrt{e}(b_{\mathrm{u}}/a_{\mathrm{u}}))\frac{(b_{\mathrm{u}}/a_{\mathrm{u}})^2}{N_{\mathrm{p}}}\right]^{1/2}
\end{aligned}$$

This along with a variation of (19) for $m=1$ and the fact that according to assumption $(\mathcal{H})_{\mathrm{u}}$, $\epsilon_{\mathrm{u}}(M)\geq a_{\mathrm{u}}/b_{\mathrm{u}}$, completes the proof of theorem. ■

The above theorem provides an error bound for the parametric approximation particle filter (using the GML algorithm to perform approximation) that is similar in structure to that specified for the subsampling approximation particle filter. The error bound consists of two distinct contributions, one corresponding to the normal operation of the filter and the other capturing the impact of the parametric approximation operation. The theorem establishes a sufficiency requirement on the sequence of approximating classes $\mathcal{H}_i$ leading to unbiased approximation of distribution flows. The requirement is that the Markov transition kernel must have an associated bounded density and this density must be a member of the class $\mathcal{H}_i$. This condition is reminiscent of the modeling assumptions that underpin Gaussian sum particle filtering (see e.g. [12]), where the premise is that the filtering density can asymptotically be represented as an infinite sum of Gaussians.

VI. Numerical Experiments

In this section we present the results of numerical experiments exploring the performance of the leader node particle filter. The experiments provide an example of how the subsampling and parametric approximation particle filters can be applied in a practical tracking problem. They provide an opportunity to compare the performance of the two algorithms and to examine whether practical behaviour is similar to that predicted by the theoretical analysis.

We adopt the following information acquisition and target movement models. The state of the target is two-dimensional with dynamics [11]

$$X_t=X_{t-1}+r_0([\cos\varphi_t;\sin\varphi_t])+u_t.$$

Here r_0 is a constant (we set $r_0 = 0.02$) and φ_t, u_t are independent and uniformly distributed $u_t \sim U[0,1]$, $\varphi_t \sim U[-\pi, \pi]$. $K_l = 20$ leader nodes and $K_s = 200$ satellite nodes are distributed uniformly in the unit square. A satellite sensor node j with coordinates $s_j = [s_{1,j}, s_{2,j}]$ can transmit its measurement to any active leader node within the connectivity radius r_c. The connectivity radius is set to $r_c = \sqrt{2\log(K_s)/K_s}$ (note that if every node can be a leader node, $K_l = K_s$, the resultant network topology is a random geometric graph). We assume that any active leader node can transmit an approximation of its posterior representation to any other potential leader node.

The measurement equation of every satellite sensor is the binary detector [2] capable of detecting a target within radius r_d with probability p_d and false alarm rate p_f:

$$\mathbb{P}\{Y_t^j = 1 | X_t\} = \begin{cases} p_d & \text{if} \quad X_t \in \mathcal{X}_d^j \\ p_f & \text{if} \quad X_t \notin \mathcal{X}_d^j \end{cases},$$

where the detection region $\mathcal{X}_d^j$ of satellite sensor j is defined as $\mathcal{X}_d^j = \{x : \|x - s_j\|_2 \leq r_d\}$. To perform sensor selection step we use the mutual information (MI) criterion [20]:

$$\ell_{t+1} = \arg \max_{\ell_{t+1} \in \mathcal{L}} I(X_{t+1}, Y_{t+1}^{\mathcal{S}_{\ell_{t+1}}} | y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) \tag{28}$$

Here $y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}$ denotes the entire history of measurements, and the random variable $Y_{t+1}^{\mathcal{S}_{\ell_{t+1}}}$ denotes the (potential) set of measurements at time $t+1$ by the set of satellite sensor nodes ($\mathcal{S}_{\ell_{t+1}}$) of a candidate leader node ℓ_{t+1}. The calculation of the mutual information in the multiple sensor framework is generally a computationally demanding exercise. In the binary sensor framework, the calculations can be simplified using an efficient approximation (see Appendix D for details).

Williams et al. pointed out in [27] that the application of the one-step mutual information criterion for sensor selection can result in undesirable leader node bouncing (frequent, unnecessary hand-off). To prevent this, Williams et al. proposed a computationally demanding finite-time horizon dynamic program [27]. In our simulations we use a simpler randomized algorithm to control the leader node exchange rate. In this algorithm the current leader node flips a biased coin with the probability of the flip outcome being 1 equal to λ. If the outcome is 1 then the current leader node calculates the mutual information criterion. It then determines if the current particle representation should be transferred to a new leader node that is more likely to make informative measurements. If the outcome is 0, then no calculations are performed. With this approach, the computational load for each leader node is significantly reduced and the communication overhead can be regulated by the choice of λ. However, the value of λ should be tailored depending on the application (as the mobility of the target increases, leader node hand-offs must be considered more frequently). In our experiments we fix $\lambda = 1/5$.

We consider two leader node particle filtering algorithms, with one employing non-parametric approximation (subsampling) and the other using parametric approximation. To create a subsample for transmission in the non-parametric framework we use the general residual resampling scheme [8]. The parametric leader node particle filter is implemented using the GML algorithm with N_{p} components. Each component consists of a two-dimensional Gaussian density with diagonal covariance matrix. The mean vector and covariance matrix are estimated using the particle representation available at the current leader node. To implement the GML algorithm we used the standard MATLAB nonlinear optimization routine `fmincon` (see Appendix C for details of the implementation).

In the following we report the simulation results obtained using the set-up discussed above. All results are achieved using 5000 Monte Carlo trials, and in each trial a new trajectory of the target is generated.

We first demonstrate that the discussed sensor selection procedure (leader node exchange rule) has good information fusion properties. Fig. 2 depicts the performance in terms of Root Mean Squared Error (RMSE) between the true position of the target and its estimate using different information diffusion schemes. The first scheme denoted by $\bigtriangledown$ corresponds to the situation when the leader node is selected at the initialization and is fixed throughout the tracking exercise. The second and third schemes denoted by $+$ and $\square$ respectively correspond to non-parametric leader node algorithms using $N_{\mathrm{b}} = 10$ and $N_{\mathrm{b}} = 300$ particles for communications respectively. The fourth scheme denoted by $\circ$ corresponds to the centralized scenario when all the measurements available from every sensor at every time step t are used to track the target. Note that the baseline particle filter uses $N = 300$ particles[4] in all scenarios (so the $N_{\mathrm{b}} = 300$ case corresponds to no subsampling). We can see from Fig. 2 that

[4]This value was selected after experimentation with multiple values of N because it provides sufficient accuracy without inducing unnecessary computational overhead. The primary purpose of the simulations is to examine the impact of the approximation steps.

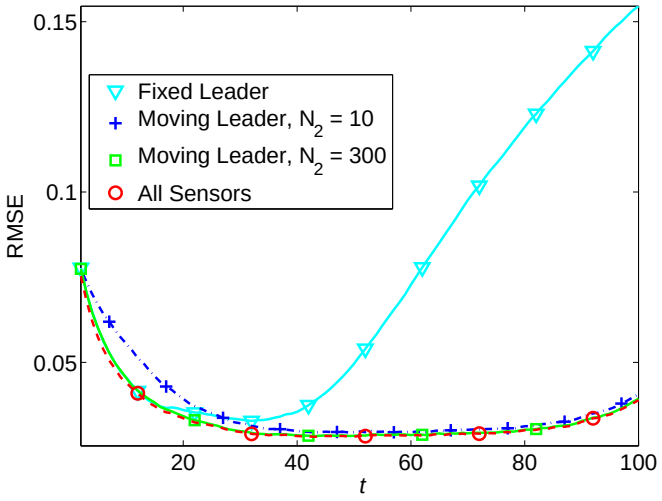


Fig. 2. Performance (RMSE) of different fusion schemes versus time: $\bigtriangledown$ denotes the scheme with fixed leader node selected at initialization; $+$ denotes the scheme with leader node selected using approximate Mutual Information (MI) criterion and non-parametric (subsampling) approximation with $N_{\mathrm{b}} = 10$; $\square$ denotes the scheme with leader node selected using approximate MI criterion but no subsampling approximation ($N_{\mathrm{b}} = 300$); and $\circ$ denotes the centralized scheme using the entire set of measurements from all sensors at every step.

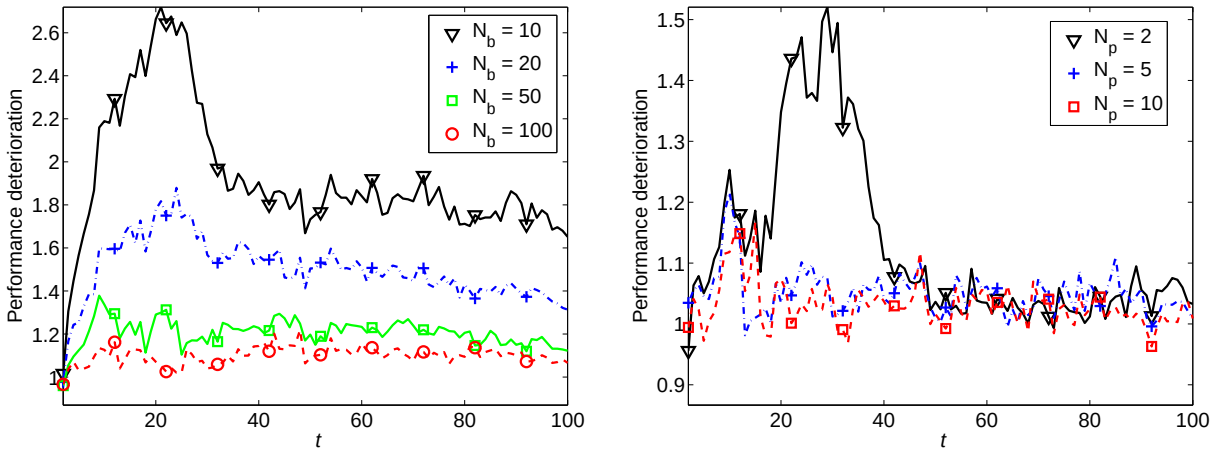
the centralized scheme has the best performance in terms of RMSE. However, it is only marginally better than the leader node scenario without compression ($N = N_{\mathrm{b}} = 300$). This highlights the effectiveness of the leader node particle filtering method and confirms that leader node selection based on the approximate mutual information is a valid approach. Compared to the centralized scheme, the communication and power consumption costs are significantly decreased since only a small subset of nodes is activated at any particular time step.

The leader node particle filter that uses a very small number of transmitted particles ($N_{\mathrm{b}} = 10$) performs comparably well. This suggests that there are practical scenarios where a particle filter can incorporate aggressive approximation to reduce communication overhead without incurring a significant penalty in tracking accuracy. The fixed leader node approach performs poorly, because the activated sensors only provide useful information when the target is nearby. As the target moves further away, the particle cloud approximating the filtering distribution becomes very diffuse, and tracking accuracy is 4 times worse than that of any of the other schemes.

In the next set of results, we explore the approximation error, i.e. the error induced by both sampling and the additional parametric/subsampling approximations. The RMSE combines both approximation error and estimation error resulting from the inaccuracy and/or ambiguity of the measurement information. We can estimate a *Root Mean Squared Approximation Error* (RMSAE) by calculating the error between a candidate particle filter and an "ideal" reference particle filter. As our reference filter, we employ a particle filter that uses $N = 3000$ particles, with no approximation during hand-off. For each of the 5000 Monte Carlo trials, we apply this reference filter to generation location estimates. The approximation error for our test filters is measured relative to these estimates rather than the true locations.

Figure 3(a) depicts how the tracking performance is affected as the number of particles in the subsampling step (N_{b}) changes; Figure 3(b) provides similar results for the parametric approximation method as the number of components in the mixture model (N_p) is varied. The performance is measured in terms of the RMSAE increase relative to a leader node particle filter that performs no approximation, i.e. uses $N_{\mathrm{b}} = N = 300$ particles during hand-off.

Fig. 3 indicates that the performance of the leader node particle filter has interesting dynamic structure. In particular, in the time period $t \in [1, 50]$ we can see an articulated transient behaviour (see Fig. 3(a), $N_{\mathrm{b}} = 10$ in particular). The transient in these curves arises because the particle representation of the target location density is initially highly dispersed and multi-modal, making it relatively difficult to approximate using either a subsampling or parametric method with a small number of particles/mixture components. However, as time progresses ($t \in [51, 100]$) the particle representation of the target becomes more localized and closer to uni-modal, so approximation performance improves significantly. Qualitatively, the performance deteriorates gracefully with respect to the extent of the compression during hand-off (reduction in number of particles or components), as theoretically predicted in



(a) Non-parametric leader node particle filter

(b) Parametric leader node particle filter

Fig. 3. Deterioration of performance as a function of (a) varying number of transmitted particles for the subsampling approximation leader node particle filter; and (b) varying number of transmitted mixture components for the parametric approximation leader node particle filter. The performance deterioration is measured as the ratio of the Root Mean Squared Approximation Error (RMSAE) of the candidate particle filtering algorithm with intermittent approximation (subsampling or parametric) to that of a leader node particle filter that performs no approximation ($N_{\mathrm{b}} = 300$). See Section VI for a definition of the RMSAE.

the previous sections.

For the final performance analysis, we define a *compression factor* as the ratio of the number of particles used during regular particle filter computations to the number of *values* transmitted during the hand-off. For the subsample approximation case, this is simply N/N_{b}. In our case of a Gaussian mixture, variance information is transmitted in addition to the locations of the Gaussians and the mixture weights, so the factor is $2N/5N_{\mathrm{p}}$. Figure 4 presents a box-plot depicting performance deterioration (ratio of approximation error of the leader node with $N_b < N$ and the leader node with $N_b = N$) versus the compression factor. Both the median and the maximal deviations of the performance deterioration scale smoothly with changing compression factor. Parametric approximation clearly outperforms subsampling.

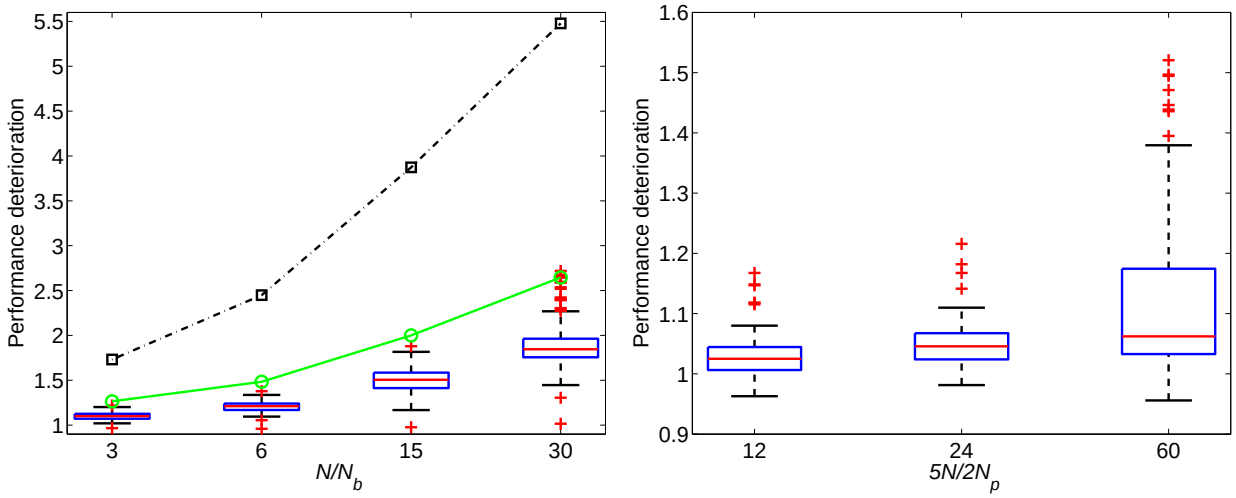
For the subsampling case, Theorem 2 and Corollary 3 provide an analytical bound on the expected approximation error. The curve based on these results is depicted in Figure 4(a) and provides a meaningful characterization of the expected performance deterioration. Indeed, the theoretical prediction based on the factor $(q_{\mathrm{u}}\chi + (1 - q_{\mathrm{u}}))^{1/2}$ from Corollary 3 closely coincides with the maximal performance deterioration observed for each compression factor. For comparison purposes, we include a similar characterization derived based on a simple worst-case assumption that the subsample approximation particle filter performs only as well as a particle filter that uses N_{b} particles at all times. The characterization based on the bounds developed in this chapter clearly provides a better indication of the performance deterioration.

VII. Related Work

The analysis of approximation error propagation and stability of non-linear Markov filters has been an active research area for several decades. In [13] Kunita studied the asymptotic behaviour of the error and stability of the filter that has an ergodic signal transition semigroup with respect to the initial distribution. Ocone and Pardoux [21] addressed the stability of linear filters with respect to a non-Gaussian initial condition and examined the stability of non-linear filters in the case where the signal diffusion is convergent. The important conclusion drawn by Ocone and Pardoux based on results in [13], [21] is that if the signal diffusion is stable with respect to its initial condition then the optimal filter inherits this property and it is also stable with respect to the initial condition.

Although interesting, the results in [13], [21] address the optimal filtering scenario, and more relevant to our study is the analysis of approximately optimal filters (especially particle filters). Important results concerning the stability of particle filters have been developed over the past decade [3]–[6], [8], [15]–[17], [26].

The Feynman-Kac semigroup approach to the stability analysis of particle filters has been described and developed by Del Moral, Miclo and Guionnet in [4]–[6]. The authors study the stability properties of general non-linear



(a) Non-parametric leader node particle filter. □ denotes the naive performance deterioration characterization, $\sqrt{N/N_{\mathrm{b}}}$. ∘ denotes the proposed characterization captured by the factor $(q_{\mathrm{u}}\chi + (1-q_{\mathrm{u}}))^{1/2}$ in Corollary 3.

(b) Parametric leader node particle filter

Fig. 4. Box-plots showing the relationship between deterioration of approximation performance and compression factor. The performance deterioration is measured as the ratio of the Root Mean Squared Approximation Error (RMSAE) of the candidate particle filtering algorithm with intermittent approximation (subsampling or parametric) to that of a leader node particle filter that performs no approximation ($N_{\mathrm{b}} = 300$). The compression factor, defined in Section VI, is the ratio of N to the number of values transmitted during leader node exchange (N_b or $2.5N_p$). The boxes show lower quartile, median and upper quartile of the 5000 Monte Carlo trials. Whiskers depict 1.5 times the interquartile range and capture most of the extreme values, and the $+$ values denote outliers extending beyond the whiskers.

Feynman-Kac semigroups under a variety of assumptions. The Dobrushin contraction coefficient of the underlying Markov chain plays a central role in the analysis. In [6], Del Moral and Miclo formulate the conditions for the exponential asymptotic stability of the Feynman-Kac semigroup and bound the Lyapunov constant and Dobrushin coefficient. One of the applications of these results is a time-uniform upper bound on the error of interacting particle systems. In [4], Del Moral provides an extensive analysis of the properties of Feynman-Kac semigroups. His analysis forms the basis for our study in this paper, particularly in the case of the subsampling approximation particle filter.

Stability analysis for particle filters is frequently built on relatively strong assumptions about the mixing and ergodicity properties of the underlying Markov transitions of the signal (target state). Our analysis in this paper is no exception, with the regularity conditions $(M)_{\mathrm{u}}^{(m)}$ and $(\mathcal{H})_{\mathrm{u}}$ underpinning our results. There have been some efforts to relax these types of assumptions. In [16], [17], Le Gland and Oudjane study the stability and convergence rates for particle filters using the Hilbert projective metric. In [16], they relax the signal mixing assumptions by employing a specific, "robust" particle filter architecture with truncated likelihood functions. In [17], the mixing assumption is applied not to the Markov kernel governing signal diffusion, but to the non-negative kernel that governs the evolution of the particle filter. This kernel combines the effects of the Markov transitions and the likelihood potentials, so mixing behaviour can arise from either of these two components.

The papers cited thus far addressed the analysis of particle filters with fixed population size (number of particles). In the subsampling approximation particle filter analyzed in this paper, the number of particles varies over time. Crisan et al. examine the stability of branching and interacting particle systems in [3]; in these systems the population size also varies, because at each time step a particle generates a random number of offspring. The population size forms a positive integer-valued martngale with respect to the filtration and the properties of the resulting particle filter depend on the initial number of particles. The variation in the number of particles is clearly very different from that of the subsampling approximation particle filter, so the results are not directly applicable.

Thus far we have discussed previous work that has addressed particle filter stability when the error arises due to the sampling approximation. The sampling error is dependent on the resampling schemes, and Douc et al. have provided theoretical results that allow various resampling schemes to be compared [8]. Other work has considered additional sources of error. Vaswani et al. analyzed the performance of a particle filter in the case of signal model

mismatch (when the true underlying Markov transitions differ from the model used to update the filter) [26]. They showed, using the same assumptions as in [17], that the particle filter is stable if the mismatch persists for only a finite interval of time.

Le Gland et al. propose and analyze the kernel-based *regularized particle filter* in [15], [17], and this work is the most closely related to our study of the parametric approximation particle filter. From an algorithmic standpoint, there are also similarities with the Gaussian sum particle filter [12], but the theoretical analysis of this filter is less developed. The regularized particle filters described in [15], [17] incorporate a step in which the N-sample pointwise density approximation is replaced by a continuous density approximation, using a kernel-based density estimation approach. During resampling, N particles are generated by sampling from this continuous density. The practical benefit of this approach is the increase in the diversity of the particle system, eliminating the potential for degeneracy and improving the stability of the algorithm. Le Gland et al. provide uniform convergence results for the regularized particle filters. Although there is some similarity to the parametric approximation particle filter we analyze, the purpose of the approximation is very different. It is not performed intermittently to reduce computation or communication cost, but rather is performed every time step with a complex model (N components).

There has been some work addressing the analysis of the leader node particle filter [11]. Although simulation (and to some extent, experimental) results indicate that instability effects are rarely observed in the leader node particle filtering, prior to our work, the theoretical bounds on estimation error for leader node particle filtering using intermittent parametric approximation grow exponentially over time [11]. The analysis in [11] used maximum log-error to model the approximation error propagation. Because the analysis does not take into account the structure of the dynamic system (and indeed imposes no assumptions on its mixing properties and regularity), the resulting bounds diverge.

VIII. Concluding Remarks

We have presented the analysis of particle filters that perform intermittent approximation. Our main results have the form of upper bounds on the expected L_p error of particle filters that occasionally employ either subsampling or parametric approximations of the filtering distribution. Such approximation steps become necessary when particle filters are deployed on resource-constrained platforms, where the resource can be energy, memory or computational power. The important conclusion of our analysis is that these approximation steps do not induce instability, and moreover, the frequency of the approximation steps significantly affects the extent of performance degradation. If the approximation steps are rare, then the compression can be very high (very few subsamples or very few mixture components) and the error remains reasonable. Numerical experiments indicate that the bound for the subsample approximation particle filter provides a meaningful characterization of practical performance.

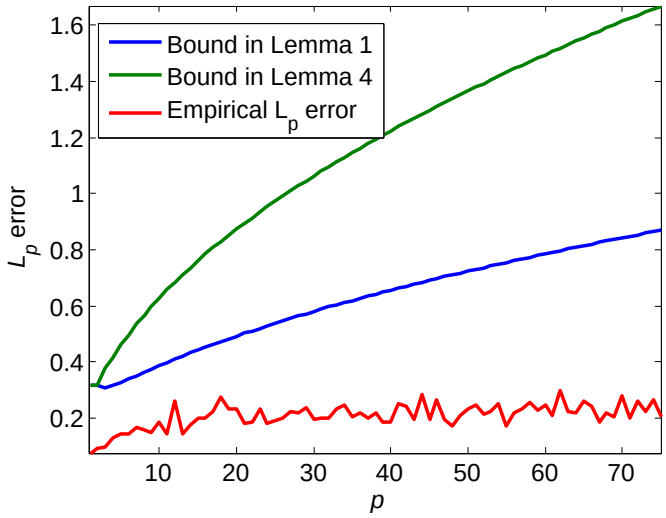


Fig. 5. Comparison of bounds in Lemma 4 and Lemma 1 for the L_p error of the N-sample mean estimator of the uniform random variable distributed over the interval $[0, 10]$. Sample size, $N = 1000$.

APPENDIX A
THE COMPARISON OF LOCAL APPROXIMATION ERROR BOUNDS

The following local error analysis result from [4] is presented here without a proof.

Lemma 4 (Del Moral [4], Lemma 7.3.3)**.** *For any $p \geq 1$ and sequence of $\mathcal{E}$-measurable functions $(h_i)_{i\geq 1}$ with finite oscillations such that $\mu_i(h_i) = 0$ for all $i \geq 1$ we have*

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

where the following definitions are used

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

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

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

It is relatively straightforward to see why the sequence of constants $c(p)$ provides tighter bounds in Lemma 1 than the sequence $d(p)$ in Lemma 4. For example, for the even $p = 2n$ the ratio of the two sequences is

$$\begin{aligned}\frac{d(2n)}{c(2n)} &= \frac{(2n)!2^{-n}}{n!(2n)\Gamma(n)2^{-n}} \\ &= \frac{(2n-1)!}{n(n-1)!\Gamma(n)} = \frac{\Gamma(2n)}{n\Gamma(n)\Gamma(n)} \\ &= \frac{1}{n\mathrm{B}(n,n)}. \end{aligned} \tag{29}$$

Here B is the beta function. $\mathrm{B}(n, n)$ is a quickly decaying function. In particular, for large n Stirling's approximation gives a simple expression for beta function, $\mathrm{B}(n, n) \sim \sqrt{2\pi}n^{-1/2}2^{-2n+1/2}$, yielding the large n Stirling's approximation for (29):

$$\frac{d(2n)}{c(2n)} \sim \frac{1}{\sqrt{2\pi n}}2^{2n-1/2}.$$

This shows that $c(p)$ grows much slower with p than $d(p)$. This improved nature of constants $c(p)$ results in better estimates of moment generating function in Theorem 1 and Corollary 1, and better exponential inequality in Theorem 3. The comparison of the bound in Lemma 4 with the bound in Lemma 1 is provided in Fig. 5 for the L_p error of the N-sample mean estimator in the case of uniform random variable distributed over the interval $[0, s]$. We verify that in this case the function that we are analyzing is:

$$\begin{aligned}\mathbb{E}\{|[P - S^N(P)](h)|^p\}^{\frac{1}{p}} &= \mathbb{E}\left\{\left|\frac{1}{N}\sum_{i=1}^{N} h(x_i)\right|^p\right\}^{\frac{1}{p}} \\ &= \mathbb{E}\left\{\left|\frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)\right|^p\right\}^{\frac{1}{p}}\end{aligned}$$

We conclude that our test function has the following form in this setting: $h(x_i) = x_i - \mu$ and oscillations of this function can be estimated straightforwardly:

$$\begin{aligned}\sigma^2(h) &= \frac{1}{N}\sum_{i=1}^{N} \operatorname{osc}^2(h_i) \\ &= \frac{1}{N}\sum_{i=1}^{N} \sup\{|h_i(x_i) - h_i(y_i)|; x_i, y_i \in E_i\}^2 \\ &= \frac{1}{N}\sum_{i=1}^{N} \sup\{|x_i - \mu - y_i + \mu|; x_i, y_i \in [0, s]\}^2 \\ &= \frac{1}{N}\sum_{i=1}^{N} \sup\{|x_i - y_i|; x_i, y_i \in [0, s]\}^2 \\ &= \frac{1}{N}\sum_{i=1}^{N} s^2 = s^2\end{aligned}$$

Hence we have in this case $\sigma(h) = s$ and the bound from Lemma 1 takes the form: $\mathbb{E}\{|[P - S^N(P)](h)|^p\}^{\frac{1}{p}} \leq c(p)^{1/p}\frac{s}{\sqrt{N}}$. On the other hand, applying Lemma 4 gives: $\mathbb{E}\{|[P - S^N(P)](h)|^p\}^{\frac{1}{p}} \leq d(p)^{1/p}\frac{s}{\sqrt{N}}$. Fig. 5 depicts the two bounds plotted for different values of p and compares it with the actual errors observed during simulations. We used the following settings to obtain this plot: $N = 1000$, $s = 10$.

APPENDIX B
THE ESTIMATES OF THE MOMENT GENERATING FUNCTION

In this appendix we show how the impact of improved constants in Lemma 1 can be used to improve the estimate of the moment generating function in Theorem 7.3.1 [4]. We now state the Theorem 7.3.1.

Theorem 7 (Del Moral [4], Theorem 7.3.1)**.** *For any sequence of $\mathcal{E}$–measurable functions $(h_i)_{i\geq 1}$ such that $\mu_i(h_i) = 0$ for all $i \geq 1$ we have for any ε*

$$\sigma(h) < \infty \Longrightarrow \mathbb{E}\left\{e^{\varepsilon\sqrt{N}|m(X)(h)|}\right\} \leq (1 + \varepsilon\sigma(h))\, e^{\frac{\varepsilon^2}{2}\sigma^2(h)}$$

We note that the simplified estimate of MGF in Corollary 1 is much tighter than the bound in Theorem 7 for asymptotically large deviations ε while the more complex bound in Theorem 1 outperforms the one in the Theorem 7 uniformly over the range of ε. The comparison of the bounds obtained in Theorem 7 and Theorem 1 with the empirical estimate is provided in Fig. 6. The test setup is the same as in Appendix A The parameters of the simulation can be summarized as follows: scale parameter, $s = 10$, number of iid samples, $N = 100$, averaging is performed over $M = 10000$ trials, ε ranges from 0 to 1. Similar results are obtained in other settings.

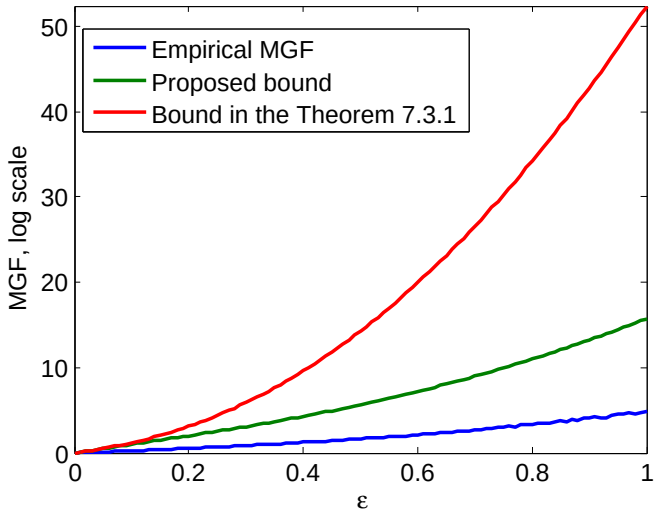


Fig. 6. Comparison of bounds in Theorem 7 and Theorem 1 for the moment generating function of N-sample mean estimator of the uniform random variable distributed over the interval $[0, 10]$. Sample size, $N = 100$.

APPENDIX C
GML IMPLEMENTATION DETAILS (OBJECTIVE FUNCTION AND ITS DERIVATIVES)

In this section we present the derivatives of the objective function of the GML algorithm. As was mentioned earlier, the injection of this information into the numerical optimization routine results in a significant (two times) acceleration of the GML speed. Assuming that $\{\xi^{(j)}\}_{j=1}^{N}$ is the current particle set, ϕ_{θ_i} comes from the class of two-dimensional Gaussian densities with diagonal covariance matrix

$$\phi_{\theta_i}(\xi^{(j)}) = \frac{1}{2\pi\sigma_{2,i}\sigma_{1,i}} e^{-\frac{\left(\xi_1^{(j)}-\mu_{1,i}\right)^2}{2\sigma_{1,i}^2} - \frac{\left(\xi_2^{(j)}-\mu_{2,i}\right)^2}{2\sigma_{2,i}^2}}$$

and ith-step GML objective is written as follows (according to Algorithm 1)

$$\mathcal{J}_i = -\sum_{j=1}^{N} \log[\alpha_i \phi_{\theta_i}(\xi^{(j)}) + (1-\alpha_i) g_{i-1}(\xi^{(j)})]$$

we can calculate the following set of the first- and second-order derivatives necessary to construct gradient and Hessian for the non-linear optimization routine at iteration i. The acceleration is achieved by evaluating the expensive exponential terms

$$V_{i,j} = e^{\frac{\left(\xi_1^{(j)}-\mu_{1,i}\right)^2}{2\sigma_{1,i}^2} + \frac{\left(\xi_2^{(j)}-\mu_{2,i}\right)^2}{2\sigma_{2,i}^2}}, \quad 1 \le j \le N$$

only once per GML iteration and vectorizing the code with respect to the terms of the type $(\xi_1^{(j)} - \mu_{1,i})$ and $\left(\left(\xi_1^{(j)} - \mu_{1,i}\right)^2 - \sigma_{1,i}^2\right)$.

$$\frac{\partial \mathcal{J}_i}{\partial \mu_{1,i}} = \sum_{j=1}^{N} \frac{-\alpha_i \left(\xi_1^{(j)} - \mu_{1,i}\right)}{\sigma_{1,i}^2 \left(\alpha_i + (1-\alpha_i) 2\pi\sigma_{1,i}\sigma_{2,i} V_{i,j} g_{i-1}(\xi^{(j)})\right)}$$

$$\frac{\partial \mathcal{J}_i}{\partial \mu_{2,i}} = \sum_{j=1}^{N} \frac{-\alpha_i \left(\xi_2^{(j)} - \mu_{2,i}\right)}{\sigma_{2,i}^2 \left(\alpha_i + (1-\alpha_i) 2\pi\sigma_{1,i}\sigma_{2,i} V_{i,j} g_{i-1}(\xi^{(j)})\right)}$$

$$\frac{\partial \mathcal{J}_i}{\partial \sigma_{1,i}} = \sum_{j=1}^{N} \frac{-\alpha_i \left(\left(\xi_1^{(j)} - \mu_{1,i} \right)^2 - \sigma_{1,i}^2 \right)}{\sigma_{1,i}^3 \left(\alpha_i + (1-\alpha_i) 2\pi \sigma_{1,i} \sigma_{2,i} V_{i,j} g_{i-1}(\xi^{(j)}) \right)}$$

$$\frac{\partial \mathcal{J}_i}{\partial \sigma_{2,i}} = \sum_{j=1}^{N} \frac{-\alpha_i \left(\left(\xi_2^{(j)} - \mu_{2,i} \right)^2 - \sigma_{2,i}^2 \right)}{\sigma_{2,i}^3 \left(\alpha_i + (1-\alpha_i) 2\pi \sigma_{1,i} \sigma_{2,i} V_{i,j} g_{i-1}(\xi^{(j)}) \right)}$$

$$\frac{\partial^2 \mathcal{J}_i}{\partial \mu_{1,i}^2} = \sum_{j=1}^{N} \frac{\alpha_i \left(\alpha_i \sigma_{1,i} + 2V_{i,j} \pi g_{i-1}(\xi^{(j)})(1-\alpha_i) \left(\left(\xi_1^{(j)} - \mu_{1,i} \right)^2 - \sigma_{1,i}^2 \right) \sigma_{2,i} \right)}{\sigma_{1,i}^3 \left(\alpha_i + (1-\alpha_i) 2\pi \sigma_{1,i} \sigma_{2,i} V_{i,j} g_{i-1}(\xi^{(j)}) \right)^2}$$

$$\frac{\partial^2 \mathcal{J}_i}{\partial \mu_{1,i} \partial \mu_{2,i}} = \sum_{j=1}^{N} \frac{2V_{i,j} \pi g_{i-1}(\xi^{(j)})(1-\alpha_i) \alpha_i \left(\xi_1^{(j)} - \mu_{1,i} \right) \left(\xi_2^{(j)} - \mu_{2,i} \right)}{\sigma_{1,i} \sigma_{2,i} \left(\alpha_i + (1-\alpha_i) 2\pi \sigma_{1,i} \sigma_{2,i} V_{i,j} g_{i-1}(\xi^{(j)}) \right)^2}$$

$$\frac{\partial^2 \mathcal{J}_i}{\partial \mu_{1,i} \partial \sigma_{1,i}} = \sum_{j=1}^{N} \frac{2\alpha_i \left(\xi_1^{(j)} - \mu_{1,i} \right) \left(\alpha_i + \frac{V_{i,j} \pi g_{i-1}(\xi^{(j)})(1-\alpha_i) \left(\left(\xi_1^{(j)} - \mu_{1,i} \right)^2 - 3\sigma_{1,i}^2 \right) \sigma_{2,i}}{\sigma_{1,i}} \right)}{\sigma_{1,i}^3 \left(\alpha_i + (1-\alpha_i) 2\pi \sigma_{1,i} \sigma_{2,i} V_{i,j} g_{i-1}(\xi^{(j)}) \right)^2}$$

$$\frac{\partial^2 \mathcal{J}_i}{\partial \mu_{1,i} \partial \sigma_{2,i}} = \sum_{j=1}^{N} \frac{2V_{i,j} \pi g_{i-1}(\xi^{(j)})(1-\alpha_i) \alpha_i \left(\xi_1^{(j)} - \mu_{1,i} \right) \left(\left(\xi_2^{(j)} - \mu_{2,i} \right)^2 - \sigma_{2,i}^2 \right)}{\sigma_{1,i} \sigma_{2,i}^2 \left(\alpha_i + (1-\alpha_i) 2\pi \sigma_{1,i} \sigma_{2,i} V_{i,j} g_{i-1}(\xi^{(j)}) \right)^2}$$

$$\frac{\partial^2 \mathcal{J}_i}{\partial \mu_{2,i}^2} = \sum_{j=1}^{N} \frac{\alpha_i \left(\alpha_i \sigma_{2,i} + 2V_{i,j} \pi g_{i-1}(\xi^{(j)})(1-\alpha_i) \sigma_{1,i} \left(\left(\xi_2^{(j)} - \mu_{2,i} \right)^2 - \sigma_{2,i}^2 \right) \right)}{\sigma_{2,i}^3 \left(\alpha_i + (1-\alpha_i) 2\pi \sigma_{1,i} \sigma_{2,i} V_{i,j} g_{i-1}(\xi^{(j)}) \right)^2}$$

$$\frac{\partial^2 \mathcal{J}_i}{\partial \mu_{2,i} \partial \sigma_{1,i}} = \sum_{j=1}^{N} \frac{2V_{i,j} \pi g_{i-1}(\xi^{(j)})(1-\alpha_i) \alpha_i \left(\xi_2^{(j)} - \mu_{2,i} \right) \left(\left(\xi_1^{(j)} - \mu_{1,i} \right)^2 - \sigma_{1,i}^2 \right)}{\sigma_{1,i}^2 \sigma_{2,i} \left(\alpha_i + (1-\alpha_i) 2\pi \sigma_{1,i} \sigma_{2,i} V_{i,j} g_{i-1}(\xi^{(j)}) \right)^2}$$

$$\frac{\partial^2 \mathcal{J}_i}{\partial \mu_{2,i} \partial \sigma_{2,i}} = \sum_{j=1}^{N} \frac{2\alpha_i \left(\xi_2^{(j)} - \mu_{2,i} \right) \left(\alpha_i + \frac{V_{i,j} \pi g_{i-1}(\xi^{(j)})(1-\alpha_i) \sigma_{1,i} \left(\left(\xi_2^{(j)} - \mu_{2,i} \right)^2 - 3\sigma_{2,i}^2 \right)}{\sigma_{2,i}} \right)}{\sigma_{2,i}^3 \left(\alpha_i + (1-\alpha_i) 2\pi \sigma_{1,i} \sigma_{2,i} V_{i,j} g_{i-1}(\xi^{(j)}) \right)^2}$$

$$\frac{\partial^2 \mathcal{J}_i}{\partial \sigma_{1,i}^2} = \sum_{j=1}^{N} \frac{2V_{i,j} \pi g_{i-1}(\xi^{(j)})(1-\alpha_i) \alpha_i \left(\left(\xi_1^{(j)} - \mu_{1,i} \right)^2 - \sigma_{1,i}^2 \right) \left(\left(\xi_2^{(j)} - \mu_{2,i} \right)^2 - \sigma_{2,i}^2 \right)}{\sigma_{1,i}^2 \sigma_{2,i}^2 \left(\alpha_i + (1-\alpha_i) 2\pi \sigma_{1,i} \sigma_{2,i} V_{i,j} g_{i-1}(\xi^{(j)}) \right)^2}$$

$$\frac{\partial^2 \mathcal{J}_i}{\partial \sigma_{1,i} \partial \sigma_{2,i}} = \sum_{j=1}^{N} \frac{2\alpha_i \left(\xi_2^{(j)} - \mu_{2,i} \right) \left(\alpha_i \sigma_{2,i} + V_{i,j} \pi g_{i-1}(\xi^{(j)})(1-\alpha_i) \sigma_{1,i} \left(\left(\xi_2^{(j)} - \mu_{2,i} \right)^2 - 3\sigma_{2,i}^2 \right) \right)}{\sigma_{2,i}^4 \left(\alpha_i + (1-\alpha_i) 2\pi \sigma_{1,i} \sigma_{2,i} V_{i,j} g_{i-1}(\xi^{(j)}) \right)^2}$$

$$\frac{\partial^2 \mathcal{J}_i}{\partial \sigma_{2,i}^2} = \sum_{j=1}^{N} \frac{2V_{i,j}\pi g_{i-1}(\xi^{(j)})(1-\alpha_i)\alpha_i \left(\xi_1^{(j)} - \mu_{1,i}\right)\left(\left(\xi_2^{(j)} - \mu_{2,i}\right)^2 - \sigma_{2,i}^2\right)}{\sigma_{1,i}\sigma_{2,i}^2 \left(\alpha_i + (1-\alpha_i)2\pi\sigma_{1,i}\sigma_{2,i}V_{i,j}g_{i-1}(\xi^{(j)})\right)^2}$$

Gradient and Hessian calculated using the above formulae can be inserted into any standard non-linear optimization routine to boost its performance. Note that as we mentioned above, for more efficient operation, the exponential terms should be evaluated only once for every iteration of the non-linear optimization routine — during the evaluation of the objective function.

APPENDIX D
APPROXIMATE CALCULATION OF THE LEADER-NODE SELECTION CRITERION

In this section we show how to calculate the information based leader-node selection criterion based on the definition (28). We first note that from the relationship between the mutual information and conditional entropy, $I(X,Y|Z=z) = H(Y|Z=z) - H(Y|X,Z=z)$ we have

$$I(X_{t+1}, Y_{t+1}^{\mathcal{S}_{\ell_{t+1}}} | y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) = H(Y_{t+1}^{\mathcal{S}_{\ell_{t+1}}} | y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) - H(Y_{t+1}^{\mathcal{S}_{\ell_{t+1}}} | X_{t+1}, y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) \tag{30}$$

Second, recalling our assumption of the conditional independence of the measurements we can see

$$H(Y_{t+1}^{\mathcal{S}_{\ell_{t+1}}} | X_{t+1}, y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) = \sum_{j \in \mathcal{S}_{\ell_{t+1}}} H(Y_t^j | X_{t+1}, y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) \tag{31}$$

Using the definition of the conditional entropy (where $y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}$ is the sequence of measurements that has already been realized [27])

$$\begin{aligned} H(Y_{t+1}^{\mathcal{S}_{\ell_{t+1}}} | X_{t+1}, y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) &= -\sum_{j \in \mathcal{S}_{\ell_{t+1}}} \int \log p(y_{t+1}^j | x_{t+1}, y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) p(x_{t+1}, y_{t+1}^j | y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) \mathrm{d}x_{t+1} \mathrm{d}y_{t+1}^j \\ &= -\sum_{j \in \mathcal{S}_{\ell_{t+1}}} \int \log p(y_{t+1}^j | x_{t+1}, y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) p(y_{t+1}^j | x_{t+1}, y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) p(x_{t+1} | y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) \mathrm{d}x_{t+1} \mathrm{d}y_{t+1}^j \end{aligned}$$

Since the true predictive density $p(x_{t+1}|y_{1:t}^{\mathcal{S}_{\ell_{1:t}}})$ is unknown we have to use its Monte-Carlo approximation consisting of the set of diffused (predictive) particles $\{\xi_{t+1}^{(i)}\}_{i=1}^N$. This results in the following efficient approximation of the above integral:

$$H(Y_{t+1}^{\mathcal{S}_{\ell_{t+1}}} | X_{t+1}, y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) = -\sum_{j \in \mathcal{S}_{\ell_{t+1}}} \frac{1}{N} \sum_{i=1}^{N} \int \log p(y_{t+1}^j | \xi_{t+1}^{(i)}) p(y_{t+1}^j | \xi_{t+1}^{(i)}) \mathrm{d}y_{t+1}^j$$

According to our sensor model, the likelihood function can be represented as follows:

$$p(y_{t+1}^j | \xi_{t+1}^{(i)}) = p_d^{y_{t+1}^j \Delta_i^j} (1-p_d)^{(1-y_{t+1}^j)\Delta_i^j} p_f^{y_{t+1}^j (1-\Delta_i^j)} (1-p_f)^{(1-y_{t+1}^j)(1-\Delta_i^j)}, \tag{32}$$

where $\Delta_i^j = \mathbf{1}_{\xi_{t+1}^{(i)} \in \mathcal{X}_d^j}$. Straightforward calculation gives

$$\begin{aligned} H(Y_{t+1}^{\mathcal{S}_{\ell_{t+1}}} | X_{t+1}, y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) = &-\sum_{j \in \mathcal{S}_{\ell_{t+1}}} q^j (p_d \log p_d + (1-p_d)\log(1-p_d)) \\ &+ (1-q^j)(p_f \log p_f + (1-p_f)\log(1-p_f)). \end{aligned}$$

Here $q^j = \frac{1}{N}\sum_{i=1}^N \Delta_i^j$ is the average number of particles in the detection region of sensor j. Thus we have constructed an efficient Monte-Carlo approximation to the second summand in the expression for the mutual information between the predicted state X_{t+1} and the measurements arising in the neighborhood of the leader

node ℓ_{t+1}. The first summand, $H(Y_{t+1}^{\mathcal{S}_{\ell_{t+1}}}|y_{1:t}^{\mathcal{S}_{\ell_{1:t}}})$ is much more difficult to approximate directly using Monte-Carlo technique. One form of decomposing this term [27]

$$H(Y_{t+1}^{\mathcal{S}_{\ell_{t+1}}}|y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) = \sum_{j=1}^{|\mathcal{S}_{\ell_{t+1}}|} H(Y_{t+1}^{j}|Y_{t+1}^{1:j-1}, y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) \tag{33}$$

implies that for a general measurement model the evaluation complexity grows exponentially in the size of the neighborhood $|\mathcal{S}_{\ell_{t+1}}|$. This is because the sequence of measurements $Y_{t+1}^{1:j-1}$ is unknown and averaging over all possible cases is required. Our experiments revealed that approximating this term using Monte-Carlo sampling of possible measurements is also inefficient. However, for sensors with uninformative (noisy) measurements the following approximation can be used

$$H(Y_{t+1}^{\mathcal{S}_{\ell_{t+1}}}|y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) = \sum_{j=1}^{|\mathcal{S}_{\ell_{t+1}}|} H(Y_{t+1}^{j}|y_{1:t}^{\mathcal{S}_{\ell_{1:t}}})$$

The intuition behind this approximation can be explained as follows. We can represent each term in the decomposition in the following way:

$$\begin{aligned} H(Y_{t+1}^{j}|Y_{t+1}^{1:j-1}, y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) = -\int \log\left[\int p(y_{t+1}^{j}|x_{t+1}) p(x_{t+1}|y_{t+1}^{1:j-1}, y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) \mathrm{d}x_{t+1}\right] \\ \left(\int p(y_{t+1}^{1:j}|x_{t+1}) p(x_{t+1}|y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) \mathrm{d}x_{t+1}\right) \mathrm{d}y_{t+1}^{1:j} \end{aligned}$$

If measurements $Y_{t+1}^{1:j-1}$ are uninformative with respect to the current predictive density $p(x_{t+1}|y_{1:t}^{\mathcal{S}_{\ell_{1:t}}})$ their incorporation will not significantly affect the density and the following will hold $p(x_{t+1}|y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) \approx p(x_{t+1}|y_{t+1}^{1:j-1}, y_{1:t}^{\mathcal{S}_{\ell_{1:t}}})$. In our particular setting a measurement Y_{t+1}^{j} is uninformative (according to the likelihood model (32)) if all the particles are either simultaneously inside or outside the detection region $\mathcal{X}_d^j$ of sensor j. This implies that Δ_i^j is same for all i and, consequently, the updated weight of every particle does not depend on the realization of measurement Y_{t+1}^{j}. Such a measurement can be excluded from mutual information calculation without affecting the accuracy of calculation. We observed in our simulations that when the particle representation of predictive density becomes localized most of the sensors in the neighbourhoods of leader-nodes become uninformative and we introduced the approximation $p(x_{t+1}|y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) \approx p(x_{t+1}|y_{t+1}^{1:j-1}, y_{1:t}^{\mathcal{S}_{\ell_{1:t}}})$ into the calculation of the mutual information. This resulted in a significant simplification (dimensionality reduction) of calculations:

$$\begin{aligned} H(Y_{t+1}^{j}|Y_{t+1}^{1:j-1}, y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) &\approx -\int \log\left[\int p(y_{t+1}^{j}|x_{t+1}) p(x_{t+1}|y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) \mathrm{d}x_{t+1}\right] \\ &\quad \left(\int p(y_{t+1}^{1:j}|x_{t+1}) p(x_{t+1}|y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) \mathrm{d}x_{t+1}\right) \mathrm{d}y_{t+1}^{1:j} \\ &= -\int \log\left[\int p(y_{t+1}^{j}|x_{t+1}) p(x_{t+1}|y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) \mathrm{d}x_{t+1}\right] \\ &\quad \left(\int p(y_{t+1}^{j}|x_{t+1}) p(x_{t+1}|y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) \mathrm{d}x_{t+1}\right) \mathrm{d}y_{t+1}^{j}. \end{aligned}$$

Furthermore, using Monte-Carlo representation of predictive density we can approximate the inner integral:

$$\int p(y_{t+1}^{j}|x_{t+1}) p(x_{t+1}|y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) \mathrm{d}x_{t+1} \approx q^j p_d^{y_{t+1}^j} (1-p_d)^{1-y_{t+1}^j} + (1-q^j) p_f^{y_{t+1}^j} (1-p_f)^{1-y_{t+1}^j}$$

Finally, calculating the outer integral we obtain:

$$\begin{aligned} H(Y_{t+1}^{j}|Y_{t+1}^{1:j-1}, y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) \approx &-(q^j p_d + (1-q^j)p_f) \log(q^j p_d + (1-q^j)p_f) \\ &- (q^j(1-p_d) + (1-q^j)(1-p_f)) \log(q^j(1-p_d) + (1-q^j)(1-p_f)) \end{aligned}$$

Thus according to (30), (31) and (33) we have the following approximate expression for the calculation of mutual information:

$$\begin{aligned}
I(X_{t+1}, Y_{t+1}^{\mathcal{S}_{\ell_{t+1}}} | y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) &= \sum_{j=1}^{|\mathcal{S}_{\ell_{t+1}}|} H(Y_{t+1}^{j} | Y_{t+1}^{1:j-1}, y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) - H(Y_{t+1}^{j} | X_{t+1}, y_{1:t}^{\mathcal{S}_{\ell_{1:t}}}) \\
&\approx \sum_{j \in \mathcal{S}_{\ell_{t+1}}} -(q^j p_d + (1-q^j)p_f) \log(q^j p_d + (1-q^j)p_f) \\
&\quad - (q^j(1-p_d) + (1-q^j)(1-p_f)) \log(q^j(1-p_d) + (1-q^j)(1-p_f)) \\
&\quad + q^j (p_d \log p_d + (1-p_d)\log(1-p_d)) \\
&\quad + (1-q^j)(p_f \log p_f + (1-p_f)\log(1-p_f))
\end{aligned}$$

Note that this is an extremely fast approximation since its complexity is proportional to $N|\mathcal{S}_{\ell_{t+1}}|$ operations as opposed to the exponential complexity of exact calculation, which is proportional to $N2^{|\mathcal{S}_{\ell_{t+1}}|}$.

REFERENCES

[1] COATES, M. (2004). Distributed particle filtering for sensor networks. In *Proc. IEEE/ACM Int. Symp. Information Processing in Sensor Networks*. Berkeley, CA.

[2] COATES, M. J. AND ING, G. (2005). Sensor network particle filters: Motes as particles. In *Proc. IEEE Workshop on Statistical Signal Processing (SSP)*. Bordeaux, France.

[3] CRISAN, D., DEL MORAL, P., AND LYONS, T. (1999). Discrete filtering using branching and interacting particle systems. *Markov Processes and Related Fields* **5**, 3, 293–318.

[4] DEL MORAL, P. (2004). *Feynman–Kac formulae. Genealogical and interacting particle approximations*. Springer, New York.

[5] DEL MORAL, P. AND GUIONNET, A. (2001). On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann. de l'Inst. H. Poincaré* **37**, 2, 155–194.

[6] DEL MORAL, P. AND MICLO, L. (2002). On the stability of nonlinear Feynman-Kac semigroups. *Ann. de la faculté des sciences de Toulouse* **(6)11**, 2, 135–175.

[7] DEVROYE, L. AND LUGOSI, G. (2001). *Combinatorial Methods in Density Estimation*. Springer-Verlag, New York.

[8] DOUC, R. AND CAPPE, O. (2005). Comparison of resampling schemes for particle filtering. In *Proc. 4th International Symp. Image and Sig. Process. and Analysis*. Zagreb, Croatia, 64–69.

[9] DOUCET, A., DE FREITAS, N., AND GORDON, N., Eds. (2001). *Sequential Monte Carlo Methods in Practice*. Springer–Verlag, Berlin.

[10] FOX, D. (2001). KLD-sampling: Adaptive particle filters. In *In Advances in Neural Information Processing Systems 14*. MIT Press, 713–720.

[11] IHLER, A., FISHER, J., AND WILLSKY, A. (2005). Particle filtering under communication constraints. In *Proc. IEEE Workshop on Statistical Signal Processing*. Bordeaux, France.

[12] KOTECHA, J. AND DJURIC, P. (2003). Gaussian sum particle filtering. *IEEE Trans. Sig. Process.* **51**, 10 (Oct.), 2602–2612.

[13] KUNITA, H. (1971). Asymptotic behavior of the nonlinear filtering errors of Markov processes. *Journal of Multivariate Analysis* **1**, 4 (Dec.), 365–393.

[14] KWOK, C., FOX, D., AND MEILA, M. (2004). Real-time particle filters. *Proc. IEEE* **92**, 3 (Mar.), 469–484.

[15] LE GLAND, F., MUSSO, C., AND OUDJANE, N. (1998). An analysis of regularized interacting particle methods for nonlinear filtering. In *Proc. IEEE European Workshop Comp.-Intensive Methods in Control and Data Process.* 167–174.

[16] LE GLAND, F. AND OUDJANE, N. (2003). A robustification approach to stability and to uniform particle approximation of nonlinear filters : the example of pseudo-mixing signals. *Stochastic Processes and their Applications* **38**, 2 (Aug.), 279–316.

[17] LE GLAND, F. AND OUDJANE, N. (2004). Stability and uniform approximation of nonlinear filters using the Hilbert metric, and application to particle filters. *Annals of Applied Probability* **14**, 1 (Feb.), 144–187.

[18] LEDOUX, M. AND TALAGRAND, M. (1991). *Probability in Banach Spaces*. Springer.

[19] LI, J. Q. AND BARRON, A. R. (1999). Mixture density estimation. In *Advances in Neural Information Processing Systems 12*. MIT Press, 279–285.

[20] LIU, J., REICH, J., AND ZHAO, F. (2003). Collaborative in-network processing for target tracking. *EURASIP J. Appl. Signal Process.* **2003**, 1 (Jan.), 378–391.

[21] OCONE, D. AND PARDOUX, E. (1996). Asymptotic stability of the optimal filter with respect to its initial condition. *SIAM Journal on Control and Optimization* **34**, 1, 226–243.

[22] RAKHLIN, A. (2006). Applications of empirical processes in learning theory: Algorithmic stability and generalization bounds. Ph.D. thesis, Massachusetts Institute of Technology.

[23] RAKHLIN, A., PANCHENKO, D., AND MUKHERJEE, S. (2005). Risk bounds for mixture density estimation. *European Series App. Industrial Math. (ESAIM): Probab. and Statist. 9*, 220–229.

[24] SHENG, X. AND HU, Y.-H. (2005). Distributed particle filter with GMM approximation for multiple target localization and tracking in wireless sensor network. In *Proc. IEEE/ACM Int. Symp. Information Processing in Sensor Networks*. Los Angeles, CA.

[25] VAN DER VAART, A. AND WELLNER, J. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

[26] VASWANI, N. (2004). Change detection in partially observed nonlinear dynamic systems with unknown change parameters. In *Proc. Am. Control Conf.* Vol. **6**. Boston, MA, 5387–5393.

[27] WILLIAMS, J. L., FISHER, J. W., AND WILLSKY, A. S. (2007). Approximate dynamic programming for communication-constrained sensor network management. *IEEE Trans. Sig. Process.* **55**, 8, 4300–4311.

[28] ZHANG, T. (2003). Sequential greedy approximation for certain convex optimization problems. *IEEE Trans. Info. Theory* **49**, 3 (Mar.), 682–691.

[29] ZHAO, F., SHIN, J., AND REICH, J. (2002). Information-driven dynamic sensor collaboration. *IEEE Signal Processing Magazine* **19**, 2 (Mar.), 61–72.

TABLE I
GREEK AND OPERATOR NOTATION

Symbol	Definition
$\|\|\cdot\|\|$	supremum norm of function h $\|\|h\|\| = \sup_{x\in E} \|h(x)\|$
$\|\|\cdot\|\|_{\text{osc}}$	$\|\|h\|\|_{\text{osc}} = \|\|h\|\| + \text{osc}(h)$ (see Table II for definition of $\text{osc}(\cdot)$)
$\|\|\cdot\|\|_{\text{tv}}$	$\|\|P(\cdot) - Q(\cdot)\|\|_{\text{tv}} = \sup\{\|P(A) - Q(A)\| : A \in \mathcal{F}\}$
$\beta(P_{i,t})$	Dobrushin contraction coefficient $\beta(P_{i,t}) = \sup\{\|\|P_{i,t}(x_i,\cdot) - P_{i,t}(y_i,\cdot)\|\|_{\text{tv}}; x_i, y_i \in E_i\}$
χ	Positive integer indicating ratio of N to N_b
δ_t	binary variable indicating whether approximation occurs at time t
$\epsilon_{\text{u}}(M)$, $\epsilon_{\text{u}}(G)$, $\epsilon_{\text{u},m}$	Constants for regularity conditions
(ε_k)	Rademacher sequence of independent binary random variables
$\gamma_t(h_t)$	unnormalized prediction Feynman-Kac model
$\mu(h)$	For measure μ in $\mathcal{P}(E)$ and function h, $\mu(h) = \int_E h(x)\mu(dx)$
$\frac{1}{\text{d}\lambda(x_i)}\nu(\text{d}\lambda(x_i))$	Density associated with distribution $\nu \in \mathcal{P}(E_i)$ and σ-finite measure λ on measurable space $(E_i, \mathcal{E}_i)$
$\eta_t(h_t)$	normalized prediction Feynman-Kac model
$\pi_{\psi_p}(Y)$	Orlicz norm of a random variable Y for non-decreasing convex function $\psi_p(x) = e^{x^p} - 1$
Φ_t	Feynman-Kac update-operator $\eta_t = \Phi_t(\eta_{t-1})$
$\Phi_{i,t}$	semigroups associated with the normalized Feynman-Kac distribution flows, $\Phi_{i,t} = \Phi_t \circ \Phi_{t-1} \circ \ldots \circ \Phi_{i+1}$
Ψ_t	Boltzmann-Gibbs transformation $\Psi_t(\nu)(\text{d}x_t) = \frac{1}{\nu(G_t)} G_t(x_t)\nu(\text{d}x_t)$
$\sigma^2(h)$	Measure of oscillations for sequence of functions h_k $\sigma^2(h) \triangleq \frac{1}{N}\sum_{k=1}^{N} \text{osc}^2(h_k)$
Θ	parameter space
θ	parameters of mixture approximation
ξ_t^k	k-th particle at time t

TABLE II
NOTATION

Symbol	Definition
$\mathcal{B}_b(E)$	Banach space of bounded functions
$c(p)$	Constants ($c(p) = 1$ if $p = 1$; and $c(p) = 2^{-p/2} p\Gamma[p/2]$ if $p > 1$)
$\mathcal{C}_{N_\mathrm{p}}$	class of discrete N_p-component convex combinations of densities from $\mathcal{H}$
$D(f\|\|g)$	Kulback-Leibler divergence for densities f and g
$D(f\|\|\mathcal{C})$	$\inf_{g\in\mathcal{C}} D(f\|\|g)$
$\mathcal{D}(\varepsilon, \mathcal{H}, d_N)$	packing number - maximum number of ε-separated points in $\mathcal{H}$ using the empirical semimetric d_N
$d_N^2(h_1, h_2)$	empirical semi-metric for $h_1, h_2 \in \mathcal{H}$ $d_N^2(h_1, h_2) = \frac{1}{N}\sum_{i=1}^N (h_1(x_i) - h_2(x_i))^2$
$(E_t, \mathcal{E}_t)$, $t \in \mathbb{N}$	sequence of measurable spaces through which the target transitions
$\mathcal{G}^{N_\mathrm{p}}$	N_p component distribution associated with g^{N_p}
g^{N_p}	N_p component mixture density
$\widehat{g}^{N_\mathrm{p}}$	empirical N_p-mixture density estimate constructed from N samples from a target density
$G_t : E_t \to [0, \infty)$	bounded, non-negative potential functions characterizing Y_t
$(G)_\mathrm{u}$	time-uniform regularity condition for likelihood potentials
$\mathcal{H}$	class of bounded parametric densities
M_{t+1}	Markov chain transitions of X_t from E_t into E_{t+1}
$M_{i,t} = M_{i+1} \dots M_t$	composite integral operator from $(E_i, \mathcal{E}_i)$ to $(E_t, \mathcal{E}_t)$
$M(\cdot, \cdot)$	Markov transition kernel
$(M)_\mathrm{u}^{(m)}$	time-uniform regularity condition on the Markov functions
$m(X) = \frac{1}{N}\sum_{k=1}^N \delta_{X^k}$	the N-empirical measure
$m(X)(h)$	Application of empirical measure to a sequence of functions h_k: $m(X)(h) = \frac{1}{N}\sum_{k=1}^N h_k(X_k)$
N	number of particles in standard particle filter
N_b	number of particles in subsampled particle filter
N_p	number of components in mixture approximation
$\mathrm{osc}(h)$	$\sup(\|h(x) - h(y)\|;\ x, y \in E)$
$\mathrm{Osc}_1(E)$	convex set of $\mathcal{E}$-measurable test functions with finite oscillations ($\{h : \mathrm{osc}(h) \le 1\}$)
S^N	Sampling operator $S^N(\eta)(h) = \frac{1}{N}\sum_{k=1}^N h(\xi^k)$
$S_\varepsilon^N(\mu)$	generator of the signed Rademacher measure for distribution μ
X_t	$d_x \times 1$ target state vector at time t (Markov chain)
$X_t' = X_{[0:t]}$	the historical path process associated with X_t
Y_t	$d_y \times 1$ observation vector at time t
$Y_t = Y_{[1:t]}$	the history of observations from time 1 to t