ANALYSIS OF ERROR PROPAGATION IN PARTICLE FILTERS WITH APPROXIMATION[1]

BY BORIS N. ORESHKIN AND MARK J. COATES

McGill University

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.

1. Introduction. Particle filters have proven to be an effective approach for addressing difficult tracking problems [8]. 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 dynamics and/or observation models that are substantially nonlinear and non-Gaussian. A particle filter maintains a set of "particles" that are candidate state values of the system (e.g., 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.

The analysis of approximation error propagation and stability of nonlinear Markov filters has been an active research area for several decades [11, 18]. 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 [2–5, 7, 13, 15].

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 [3]; our investigation of parametric approximation is founded on error bounds for the greedy likelihood maximization algorithm, which was developed in [16] and analyzed in [21, 24].

1.1. *Leader node particle filter.* Throughout the paper, we will motivate the analysis by considering the concrete example of the "leader node" particle filter [17], an algorithm that has been proposed for distributed tracking in sensor networks. One of the major concerns in distributed sensor network tracking is balancing the tradeoff between tracking performance and network lifetime. The *leader node particle filter*, proposed in [17, 25] and refined and analyzed in [9, 23], achieves significant sensing and communication energy savings. The leader node, which performs the particle filtering, changes over time to follow the target and activates only a subset of nodes at any time instant. Thus only the active sensor nodes have to relay their measurements to a nearby location.

The setting corresponding to this filtering paradigm is depicted in Figure 1. A leader node (depicted by the large circles) is responsible for performing local tracking of the target (trajectory depicted by squares) based on the data acquired
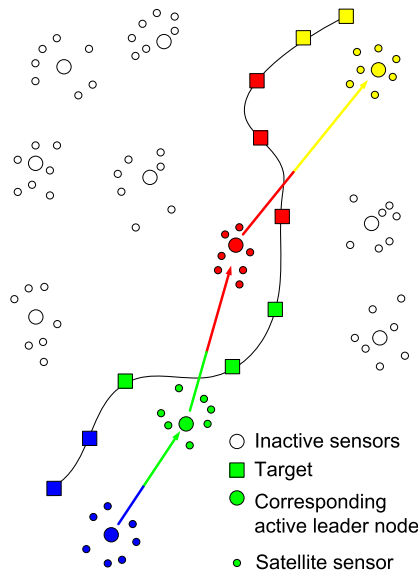

FIG. 1. *The leader node distributed particle filtering setting.*

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 neighborhood, incorporating them into its particle filter. Sensor management strategies are used to determine when to change leader node [23]. 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.

Mathematically, the leader node particle filter can be described as follows. Suppose that $\mathcal{L} = \{1, 2, \ldots, L\}$ is the set of possible leader nodes, 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 $|\mathcal{S}_\ell|$. Denote by ℓ_t the label of the leader node at time t. We adopt the following state-space model to describe the target evolution and measurement process:

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

$$Y_t^j = g_t^j(X_t, \zeta_t^j) \qquad \forall j \in \mathcal{S}_{\ell_t}. \tag{2}$$

Here $X_t \in \mathbb{R}^{d_x}$ is the target state vector at time t, $Y_t^j \in \mathbb{R}^{d_y^j}$ is the jth sensor measurement, ϱ_t and ζ_t^j 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^j is a nonlinear measurement map $g_t^j : \mathbb{R}^{d_x} \to \mathbb{R}^{d_y^j}$. The target model is the same at every leader node, but the observation process may be different.

1.2. *Feynman–Kac models.* Throughout the rest of this paper we adopt the methodology developed in [3] to analyze the behavior 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 briefly review the Feynman–Kac representation (for a much more detailed description and discussion, please refer to [3]).

To describe the probabilistic model corresponding to the state-space framework above, we need to introduce additional notation. Let $(E_t, \mathcal{E}_t)$, $t \in \mathbb{N}$, be a sequence of measurable spaces such that $X_t \in E_t$. 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

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

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

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

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

and for a sequence of functions $(h_i)_{1 \leq i \leq N} \in \mathcal{B}_b(E)^N$ we define $\sigma^2(h)$ as

$$\sigma^2(h) \triangleq \frac{1}{N} \sum_{k=1}^{N} \mathrm{osc}^2(h_i).$$

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 $(E_{t-1}, \mathcal{E}_{t-1})$ to $(E_t, \mathcal{E}_t)$

$$(\mu_{t-1}M_t)(A_t) = \int \mu_{t-1}(dx_{t-1})M_i(x_{t-1}, A_t).$$

The target state vector in (1) thus evolves according to a nonhomogeneous discrete-time Markov chain X_t with transitions M_t from E_{t-1} into E_t. These transitions and the initial distribution η_0 define the canonical probability space

$$\left(\Omega = \prod_{t \geq 0} E_t, (\mathcal{F}_t)_{t \in \mathbb{N}}, (X_t)_{t \in \mathbb{N}}, \mathbb{P}_{\eta_0} \right),$$

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 $\mathcal{F}_\infty = \sigma(\bigcup_{i \geq 0} \mathcal{F}_i)$. To characterize the properties of the observation process in (2) we introduce bounded and nonnegative potential functions $G_t^j : E_t \to [0, \infty), \forall j \in \mathcal{S}_{\ell_t}$. Assuming that in every leader node neighborhood $\mathcal{S}_{\ell_t}$ observation noises, ζ_t^j, $j \in \mathcal{S}_{\ell_t}$, in (2) are independent[2] the composite potential function at leader node ℓ_t can be written via the product of the individual potential functions of satellite nodes, G_t^j

$$G_t^{\mathcal{S}_{\ell_t}} = \prod_{j \in \mathcal{S}_{\ell_t}} G_t^j.$$

Then the propagation of the leader node Feynman–Kac model is described by a pair of prediction-update operators. Since the prediction operator is only concerned

---

[2]The assumption of independence among the sensor observations is not critical for the error analysis performed in the paper, but is adopted because it allows for a more concrete discussion and concise presentation of results.

with the dynamics of the target, it coincides with the Markov transition M_t for all possible leader nodes ℓ_t at time t. On the other hand, the Boltzmann–Gibbs transformation on $\mathcal{P}(E_t)$ is leader node dependent and is defined for any $\eta \in \mathcal{P}(E_t)$

$$\Psi_t^{\ell_t}(\eta)(\mathrm{d}x_t) = \frac{G_t^{\mathcal{S}_{\ell_t}}(x_t)\eta(\mathrm{d}x_t)}{\eta(G_t^{\mathcal{S}_{\ell_t}})}.$$

Using the diffusion M_{t+1} and Boltzmann–Gibbs transformation $\Psi_t^{\ell_t}$ we can identify an operator $\Phi_{t+1}^{\ell_t} : \mathcal{P}(E_t) \to \mathcal{P}(E_{t+1})$ which describes, for a given leader node ℓ_t, the evolution of the normalized prediction flow from time t to time $t+1$

$$\Phi_{t+1}^{\ell_t}(\eta) = \Psi_t^{\ell_t}(\eta)M_{t+1}.$$

To describe the evolution of the leader nodes we define a sensor management rule $\Upsilon_t^{\ell_t} : \mathcal{P}(E_{t+1}) \times \mathfrak{I}_t \to \mathcal{L} \times \{0, 1\}$. This mapping defines the next leader node, $\ell_{t+1} \in \mathcal{L}$, and the decision, $\Delta_{t+1} \in \{0, 1\}$, on whether or not the leader node has to be exchanged: $\Delta_{t+1} = 1$ if we decide to transfer the processing and measurement process to the leader node other than the current one. Sensor management rules usually operate according to the informativeness of the sensors and the predicted position of the target.[3] Thus the decision is made based on the utility of measurements provided by different leader nodes given the current filtering distribution $\Phi_{t+1}^{\ell_t}(\eta) \in \mathcal{P}(E_{t+1})$ and the information about leader nodes, $\mathcal{I}_t \in \mathfrak{I}_t$. Information $\mathcal{I}_t$ may include coordinates of nodes in the sensor network, measurement models for every node, costs of performing a hand-off from the current leader node to other leader nodes etc. The operation of the nonlinear mapping $\Upsilon_t^{\ell_t}$ is described by the equation

$$(\ell_{t+1}, \Delta_{t+1}) = \Upsilon_t^{\ell_t}(\Phi_{t+1}^{\ell_t}(\eta), \mathcal{I}_t).$$

The sequence of mappings $\Upsilon_{i-1}^{\ell_{i-1}}, \dots, \Upsilon_{t-2}^{\ell_{t-2}}$ defines the sequence of leader nodes $\ell_{i,t} = \ell_i, \dots, \ell_{t-1}$ that can be used to define the semigroups $\Phi_{i,t}^{\ell_{i,t}}, i \leq t$, associated with the normalized Feynman–Kac distribution flows

$$\Phi_{i,t}^{\ell_{i,t}} = \Phi_t^{\ell_{t-1}} \circ \Phi_{t-1}^{\ell_{t-2}} \circ \cdots \circ \Phi_{i+1}^{\ell_i}.$$

The semigroup $\Phi_{i,t}^{\ell_{i,t}}$ describes the evolution of the normalized prediction Feynman–Kac model from time i to time t through the sequence of leader nodes $\ell_{i,t}$

$$\eta_t^{\ell_{0,t}} = \Phi_{i,t}^{\ell_{i,t}}(\eta_i^{\ell_{0,i}}).$$

Using the analysis tools developed in [3] $\Phi_{i,t}^{\ell_{i,t}}$ can further be related to potential functions on E_i, $G_{i,t}^{\ell_{i,t}} : E_i \to (0, \infty)$, and $P_{i,t}^{\ell_{i,t}} : \mathcal{P}(E_i) \to \mathcal{P}(E_t)$, the Markov kernels from E_i to E_t. In particular, using the expectation with respect to the shifted

---

[3]We discuss an example of such an algorithm, based on [23], in Section 5.

chain,

$$\mathbb{E}_{i,x_i}\{h_{i,t}(X_{i+1}, \ldots, X_t)\}$$
$$= \int h_{i,t}(x_{i+1}, \ldots, x_t) M_{i+1}(x_i, dx_{i+1}) \cdots M_t(x_{t-1}, dx_t),$$

and defining $G_{i,t}^{\ell_{i,t}}$ as

$$G_{i,t}^{\ell_{i,t}}(x_i) = \mathbb{E}_{i,x_i}\left\{ \prod_{i \le j < t} G_j^{\mathcal{S}_{\ell_j}}(X_j) \right\},$$

we can introduce the multi-step Boltzmann–Gibbs transformation on E_i for any $\eta \in \mathcal{P}(E_i)$ and $h_i \in \mathcal{B}_b(E_i)$, $\Psi_{i,t}^{\ell_{i,t}}(\eta)(h_i) = \eta(G_{i,t}^{\ell_{i,t}} h_i)/\eta(G_{i,t}^{\ell_{i,t}})$. Defining $P_{i,t}^{\ell_{i,t}}$ by the Feynman–Kac formulae,

$$P_{i,t}^{\ell_{i,t}}(h_t) \propto \mathbb{E}_{i,x_i}\left\{ h_t(X_t) \prod_{i \le j < t} G_j^{\mathcal{S}_{\ell_j}}(X_j) \right\},$$

we can represent the semigroup $\Phi_{i,t}^{\ell_{i,t}}$ as follows:

$$\Phi_{i,t}^{\ell_{i,t}}(\eta) = \Psi_{i,t}^{\ell_{i,t}}(\eta) P_{i,t}^{\ell_{i,t}}.$$

1.3. *Dobrushin contraction and regularity conditions.* The Dobrushin contraction coefficient $(\beta_{i,t}(P) \in [0, 1])$ plays a key role in our analysis. For a fixed leader node sequence, $\ell_{i,t}$, this can be defined as follows:

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

Here the total variation metric $\| \cdot \|_{\text{tv}}$ is defined for any $\mu, \eta \in \mathcal{P}(E)$ as $\|\mu(\cdot) - \eta(\cdot)\|_{\text{tv}} = \sup\{|\mu(A) - \eta(A)| : A \in \mathcal{F}\}$. We can also define the (worst case) Dobrushin contraction coefficient, which is independent of the leader node sequence, $\beta_{i,t}(P) = \sup_{\ell_{i,t}} \beta^{\ell_{i,t}}(P_{i,t}^{\ell_{i,t}})$.

The estimation of the Dobrushin contraction coefficient is possible if we adopt certain regularity assumptions regarding the components of the Feynman–Kac operator. In particular, we adopt the following condition on the Markov kernels:

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

$$M_{i,i+m}(x_i, \cdot) = M_{i+1}M_{i+2} \cdots M_{i+m}(x_i, \cdot) \ge \epsilon_u(M) M_{i,i+m}(y_i, \cdot).$$

The following regularity condition is defined for the potential functions:

- $(G)_u$ There exists a strictly positive number $\epsilon_u(G) \in (0, 1]$ such that for any ℓ_t, $t \ge 0$ and $x_t, y_t \in E_t$

$$G_t^{\mathcal{S}_{\ell_t}}(x_t) \ge \epsilon_u^{K_u}(G) G_t^{\mathcal{S}_{\ell_t}}(y_t),$$

$(G)_u$ holds if a milder condition, $G_t^j(x_t) \geq \epsilon_t(G_t^j)G_t^j(y_t)$, holds for all t, all potential functions G_t^j, $j \in \mathcal{S}_{\ell_t}$, and for all leader nodes $\ell_t \in \mathcal{L}$. In this case we can take $\epsilon_u(G) = \inf_{t \geq 0} \min_{\ell_t \in \mathcal{L}} \min_{j \in \mathcal{S}_{\ell_t}} \epsilon_t(G_t^j)$ and $K_u = \max_{\ell_t \in \mathcal{L}} |\mathcal{S}_{\ell_t}|$.

The following two propositions that summarize results presented in [3], Proposition 4.3.3, Corollary 4.3.3 and Proposition 4.3.7, will be used for analyzing approximation error propagation in the leader-node algorithm (being employed in the proofs of Theorems 2, 3 and 6).

PROPOSITION 1 (Proposition 4.3.3. and Corollary 4.3.3 [3]). *When $(G)_u$ and $(M)_u^{(m)}$ are satisfied we have for the Dobrushin contraction coefficient $\beta_{i,t}(P) = \sup_{\ell_{i,t}} \beta^{\ell_{i,t}}(P_{i,t}^{\ell_{i,t}})$*

$$(3) \qquad \beta_{i,t}(P) \leq \left(1 - \epsilon_u^2(M)\epsilon_u^{(m-1)K_u}(G)\right)^{\lfloor (t-i)/m \rfloor}$$

and the oscillations of the potential functions,

$$(4) \qquad \frac{\inf_{x_i \in E_i} G_{i,t}^{\ell_{i,t}}(x_i)}{\|G_{i,t}^{\ell_{i,t}}\|} \geq \epsilon_u(M)\epsilon_u^{mK_u}(G), \qquad \frac{\|G_{i,t}^{\ell_{i,t}}\|}{v(G_{i,t}^{\ell_{i,t}})} \leq \epsilon_u^{-1}(M)\epsilon_u^{-mK_u}(G).$$

PROPOSITION 2 (Proposition 4.3.7 [3]). *For any $0 \leq p \leq n$, $\mu_p \in \mathcal{P}(E_p)$, and $f_n \in \mathcal{B}_b(E_n)$ with $\mathrm{osc}(f_n) \leq 1$ there exists a function $f_{p,n}^{\mu_p}$ in $\mathcal{B}_b(E_p)$ with $\mathrm{osc}(f_{p,n}^{\mu_p}) \leq 1$ such that for any $\eta_p \in \mathcal{P}(E_p)$ we have*

$$(5) \qquad |\Phi_{p,n}(\eta_p) - \Phi_{p,n}(\mu_p)| \leq \beta(P_{p,n})\frac{\|G_{p,n}\|_{\mathrm{osc}}}{\eta_p(G_{p,n})}|(\eta_p - \mu_p)(f_{p,n}^{\mu_p})|.$$

Proposition 2 implies that for any $\mu, \nu \in \mathcal{P}(E_i)$ and $h_t \in \mathrm{Osc}_1(E_t)$ there exists $h_i \in \mathrm{Osc}_1(E_i)$ such that

$$(6) \qquad |[\Phi_{i,t}^{\ell_{i,t}}(\nu) - \Phi_{i,t}^{\ell_{i,t}}(\mu)](h_t)| \leq \beta^{\ell_{i,t}}(P_{i,t}^{\ell_{i,t}})\frac{\|G_{i,t}^{\ell_{i,t}}\|_{\mathrm{osc}}}{v(G_{i,t}^{\ell_{i,t}})}|(\nu - \mu)(h_i)|.$$

Using the fact that we have for some positive function φ,

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

and, furthermore, $\beta^{\ell_{i,t}}(P_{i,t}^{\ell_{i,t}}) \leq \beta_{i,t}(P)$ for any $i \leq t$ and $\ell_{i,t}$, we see

$$(7) \qquad \begin{aligned} &|[\Phi_{i,t}^{\ell_{i,t}}(\nu) - \Phi_{i,t}^{\ell_{i,t}}(\mu)](h_t)| \\ &\qquad \leq \beta_{i,t}(P)\frac{\|G_{i,t}^{\ell_{i,t}}\|}{v(G_{i,t}^{\ell_{i,t}})}\left[2 - \frac{\inf_{y_i \in E_i} G_{i,t}^{\ell_{i,t}}(y_i)}{\|G_{i,t}^{\ell_{i,t}}\|}\right]|(\nu - \mu)(h_i)|. \end{aligned}$$

Thus under assumptions $(G)_u$ and $(M)_u^{(m)}$ the error propagation in the leader node filter can be characterized as follows:

(8)
$$|[\Phi_{i,t}^{\ell_{i,t}}(\nu) - \Phi_{i,t}^{\ell_{i,t}}(\mu)](h_t)| \leq \left(1 - \epsilon_u^2(M)\epsilon_u^{(m-1)K_u}(G)\right)^{\lfloor(t-i)/m\rfloor}$$
$$\times \frac{2 - \epsilon_u(M)\epsilon_u^{mK_u}(G)}{\epsilon_u(M)\epsilon_u^{mK_u}(G)}|(\nu - \mu)(h_i)|.$$

These results describe the propagation of one-step approximation error through the nonlinear operator $\Phi_{i,t}^{\ell_{i,t}}$. 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 $G_{i,t}^{\ell_{i,t}}$ and the Dobrushin contraction coefficient $\beta_{i,t}(P)$.

1.4. *N-particle and parametric approximations.* Let the sampling operator $S^N : \mathcal{P}(E) \to \mathcal{P}(E^N)$ be defined as

(9)
$$S^N(\eta)(h) = \frac{1}{N}\sum_{k=1}^{N} h(\xi_k),$$

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 distribution update recursion, $\widehat{\eta}_{t+1} = S^N(\Phi_{t+1}(\widehat{\eta}_t))$.

The operation of the leader node with additional approximations, on the other hand, is more complex. In particular, the standard particle filter recursion is applied if $\Delta_{t+1} = 0$ (leader node does not change). If $\Delta_{t+1} = 1$, there is a change in leader node, and there must be a transfer of information from the current leader node to the next one. The communication of all N particles is prohibitively costly in terms of energy. The leader node particle filter therefore communicates a coarser approximation of its N-particle representation. In this paper, we consider two possibilities for this additional approximation step: (i) random subsampling (choosing N_b of the particles at random); and (ii) parametric approximation of the filtering distribution.

The subsampling leader node particle filter can then be expressed as

(10)
$$\widehat{\eta}_{t+1}^{\ell'_{0,t+1}} = S^N \circ S^{N_b}(\Phi_{t+1}^{\ell'_t}(\widehat{\eta}_t^{\ell'_{0,t}})) \qquad \text{if } \Delta'_{t+1} = 1,$$
$$\widehat{\eta}_{t+1}^{\ell'_{0,t+1}} = S^N(\Phi_{t+1}^{\ell'_t}(\widehat{\eta}_t^{\ell'_{0,t}})) \qquad \text{if } \Delta'_{t+1} = 0.$$

Here $\widehat{\eta}_t^{\ell'_{0,t}}$ is the distribution obtained via the sequence of the leader nodes $\ell'_{0,t}$ with the convention $\Phi_0^{\ell'_{-1}}(\widehat{\eta}_{-1}^{\ell'_{0,-1}}) = \eta_0$ and $\Phi_1^{\ell'_0}(\widehat{\eta}_0^{\ell'_{0,0}}) = S^N(\eta_0)M_1$. In this scenario, the sensor management step is accomplished via a suboptimal rule using the approximate prediction of the target state

$$(\ell'_{t+1}, \Delta'_{t+1}) = \Upsilon_t^{\ell'_t}(\Phi_{t+1}^{\ell'_t}(\widehat{\eta}_t^{\ell'_{0,t}}), \mathcal{I}_t).$$

There is also an additional subsampling operation (S^{N_b}) after the update of the predictive posterior using the operator $\Phi_{t+1}^{\ell_t'}$. Note that $N_b < N$ so that the communication cost of the leader node exchange is reduced, since only N_b particles are transmitted. This step is followed by communication of the subsampled particle set to the new leader node, and finally there is an upsampling operation to regenerate N particles from the N_b-particle approximation.

In order to express the parametric approximation particle filter in an analogous fashion, we introduce an operator $\mathbb{W}_{N_p} : \mathcal{P}(E) \to \mathcal{P}(E^{N_p})$, which, when applied to a measure $\nu \in \mathcal{P}(E)$, constructs a parametric mixture approximation comprised of N_p mixture components

$$(11) \qquad \mathbb{W}_{N_p}(\nu)(h) = \sum_{k=1}^{N_p} \alpha_k \mu_{\theta_k}(h).$$

Here $\mu_{\theta_k} \in \mathcal{P}(E)$ is parameterized by a set of parameters θ_k and α_k are weights satisfying $\alpha_k \geq 0$ and $\sum_k \alpha_k = 1$; θ_k and α_k are estimated from ν. Section 4 provides a concrete example of $\mathbb{W}_{N_p}$ based on the greedy maximum likelihood maximization. The parametric approximation particle filter can then be expressed as

$$(12) \qquad \begin{aligned} \widehat{\eta}_{t+1}^{\ell_{0,t+1}'} &= S^N \circ \mathbb{W}_{N_p} \circ S^N (\Phi_{t+1}^{\ell_t'}(\widehat{\eta}_t^{\ell_{0,t}'})) \qquad \text{if } \Delta_{t+1}' = 1, \\ \widehat{\eta}_{t+1}^{\ell_{0,t+1}'} &= S^N (\Phi_{t+1}^{\ell_t'}(\widehat{\eta}_t^{\ell_{0,t}'})) \qquad \text{if } \Delta_{t+1}' = 0. \end{aligned}$$

Here if there is a leader node exchange ($\Delta_{t+1}' = 1$) the output of the standard particle filter is fed into the parametric mixture approximation operator that outputs parameters θ_k and weights α_k, $k = 1, \ldots, N_p$. These parameters and weights are further transmitted to the new leader node to reduce the communication cost. An N-particle approximation is then regenerated by sampling from the mixture with parameters θ_k and weights α_k.

1.5. *Problem statement.* In this paper we study the additional approximation errors arising during the leader node exchanges. These additional approximation errors are the result of either additional random subsampling in the subsample approximation leader node particle filter, or the additional parametric approximation in the parametric leader node particle filter.

Let us denote $\eta_t^{\ell_{0,t}} = \Phi_{0,t}^{\ell_{0,t}}(\eta_0)$ the true leader node distribution flow and $\ell_{0,t}$ the associated sequence of leader nodes obtained via the optimal sensor management rule $(\ell_{t+1}, \Delta_{t+1}) = \Upsilon_t^{\ell_t}(\Phi_{t+1}^{\ell_t}(\eta_t^{\ell_{0,t}}), \mathcal{I}_t)$. The approximate leader node distribution flow $\widehat{\eta}_t^{\ell_{0,t}'}$ defined by either (10) or (12) uses the sequence of leader nodes obtained via the suboptimal sensor management rule, $(\ell_{t+1}', \Delta_{t+1}') = \Upsilon_t^{\ell_t'}(\Phi_{t+1}^{\ell_t'}(\widehat{\eta}_t^{\ell_{0,t}'}), \mathcal{I}_t)$.

The global error between the true filtering distribution, $\eta_t^{\ell_{0,t}}$, and $\widehat{\eta}_t^{\ell'_{0,t}}$ can be split into two components:

$$\mathbb{E}\{|[\eta_t^{\ell_{0,t}} - \widehat{\eta}_t^{\ell'_{0,t}}](h_t)|^p\}^{1/p} \leq \mathbb{E}\{|[\eta_t^{\ell'_{0,t}} - \widehat{\eta}_t^{\ell'_{0,t}}](h_t)|^p\}^{1/p}$$

$$+ \mathbb{E}\{|[\eta_t^{\ell_{0,t}} - \eta_t^{\ell'_{0,t}}](h_t)|^p\}^{1/p}.$$

Here the first term represents the error accumulated in the leader node recursion because of the additional distribution approximations during the leader node exchanges, and the second term represents the errors arising due to the sub-optimality of the sensor management rule. In this paper we study the errors of the first kind.

The global error of the first kind, $\widehat{\eta}_t^{\ell'_{0,t}} - \eta_t^{\ell'_{0,t}}$, can be related to the sequence of local approximation errors $\widehat{\eta}_i^{\ell'_{0,i}} - \Phi_i^{\ell'_{i-1}}(\widehat{\eta}_{i-1}^{\ell'_{0,i-1}}), i = 0, \ldots, t$ [3]

$$(13) \qquad \widehat{\eta}_t^{\ell'_{0,t}} - \eta_t^{\ell'_{0,t}} = \sum_{i=0}^{t}[\Phi_{i,t}^{\ell'_{i,t}}(\widehat{\eta}_i^{\ell'_{0,i}}) - \Phi_{i,t}^{\ell'_{i,t}}(\Phi_i^{\ell'_{i-1}}(\widehat{\eta}_{i-1}^{\ell'_{0,i-1}}))].$$

To simplify the notation in the rest of the article, we will use the following convention, suppressing the explicit identification of the leader-node sequences; we will write $\eta'_t \equiv \eta_t^{\ell'_{0,t}}$ and $\eta_t \equiv \eta_t^{\ell_{0,t}}$, with associated mappings $\Phi_t \equiv \Phi_t^{\ell_{t-1}}$ and $\Phi'_t \equiv \Phi_t^{\ell'_{t-1}}$. Similarly we express the particle approximations as $\widehat{\eta}'_t \equiv \widehat{\eta}_t^{\ell'_{0,t}}$ and $\widehat{\eta}_t \equiv \widehat{\eta}_t^{\ell_{0,t}}$.

1.6. *Paper organization.* The rest of the paper is organized as follows. Section 2 presents some foundational results that serve as the basis for our analysis. In Section 3 we present error bounds and exponential inequalities for the leader node particle filter that performs intermittent subsampling, and in Section 4 we analyze the performance of this filter when it employs parametric approximation. Section 5 describes numerical experiments that illustrate the performance of the algorithms we analyze and the relationship to the bounds. Section 6 discusses related work, and Section 7 summarizes the contribution and makes concluding remarks.

2. 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. It is used to characterize the one-step approximation errors in the leader node particle filter.

LEMMA 1. *Suppose $v \in \mathcal{P}(E)$, then for any $p \geq 1$ and an $\mathcal{E}$-measurable function h with finite oscillations we have*

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

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

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

We note that

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

and we have, from the Chernov–Hoeffding inequality,

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

Next we recall the following property:

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

And finally we obtain

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

Applying Lemma 7.3.3 of [3] 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.   □

Lemma 1 tightens Lemma 7.3.3 from [3] and extends it to include noninteger p. It is relatively straightforward to see why the sequence of constants $c(p)$ provides improvement over Lemma 7.3.3 from [3] that uses the sequence of constants $d(p)$. For example, for even $p = 2n$, $d(2n) = (2n)!/n!2^{-n}$ and the ratio of the two sequences is

$$(14) \qquad \frac{d(2n)}{c(2n)} = \frac{(2n)!2^{-n}}{n!(2n)\Gamma(n)2^{-n}} = \frac{(2n-1)!}{n(n-1)!\Gamma(n)} = \frac{\Gamma(2n)}{n\Gamma(n)\Gamma(n)} = \frac{1}{n\mathrm{B}(n,n)}.$$

Here B is the Beta function. $B(n, n)$ is a quickly decaying function. For large n, Stirling's approximation gives a simple expression for the Beta function, $B(n, n) \sim \sqrt{2\pi}n^{-1/2}2^{-2n+1/2}$, yielding the large n Stirling's approximation for (14),

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

This shows that $c(p)$ grows much slower with p than $d(p)$.

The following theorem provides a bound on the moment-generating function of the empirical process $\sqrt{N}[\nu - S^N(\nu)](h)$. The result employs Lemma 1 to tighten Theorem 7.3.1 of [3].

THEOREM 1. *For any $\mathcal{E}$-measurable function h such that $\sigma(h) < \infty$, we have for any ε*

$$\mathbb{E}\{e^{\varepsilon\sqrt{N}|[\nu-S^N(\nu)](h)|}\} \leq 1 + \varepsilon\sigma(h)\left(1 - \sqrt{\frac{\pi}{2}} + \sqrt{\frac{\pi}{2}}e^{(\varepsilon^2/8)\sigma^2(h)}\left[1 + \text{Erf}\left[\frac{\varepsilon\sigma(h)}{\sqrt{8}}\right]\right]\right).$$

PROOF. We first utilize the power series representation of the exponential

$$\mathbb{E}\{e^{\varepsilon|[\nu-S^N(\nu)](h)|}\} = 1 + \varepsilon\mathbb{E}\{|[\nu - S^N(\nu)](h)|\} + \sum_{n\geq 2}\frac{\varepsilon^n}{n!}\mathbb{E}\{|[\nu - S^N(\nu)](h)|^n\}.$$

Utilizing Lemma 1 we have

$$\mathbb{E}\{e^{\varepsilon|[\nu-S^N(\nu)](h)|}\}$$

$$\leq 1 + \frac{\varepsilon\sigma(h)}{\sqrt{N}} + \sum_{n\geq 2}\left[\frac{\varepsilon\sigma(h)}{(2N)^{1/2}}\right]^n\frac{\Gamma[n/2]}{(n-1)!}$$

$$= 1 + \frac{\varepsilon\sigma(h)}{\sqrt{N}} - \frac{\varepsilon\sigma(h)\sqrt{\pi}}{\sqrt{2N}} + \frac{\varepsilon\sigma(h)\sqrt{\pi}}{\sqrt{2N}}e^{\varepsilon^2\sigma^2(h)/(8N)}\left[1 + \text{Erf}\left[\frac{\varepsilon\sigma(h)}{\sqrt{8N}}\right]\right].$$

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 $\mathcal{E}$-measurable function h such that $\sigma(h) < \infty$, we have for any ε*

$$\mathbb{E}\{e^{\varepsilon\sqrt{N}|[\nu-S^N(\nu)](h)|}\} \leq (1 + \sqrt{2\pi}\varepsilon\sigma(h))e^{(\varepsilon^2/8)\sigma^2(h)}.$$

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

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 [3] for asymptotically large deviations ε while the more complex bound in Theorem 1 is uniformly tighter over the range of ε.

3. Particle filters with intermittent subsampling. This section presents an analysis of the error propagation in the leader node particle filter that performs 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. Our main results are a time-uniform bound on the weak-sense L_p-error and an associated exponential inequality.

3.1. *Time-uniform error bounds and exponential inequalities.* We now analyze the global approximation error for the leader node particle filter with intermittent subsampling defined by recursion (10). We first present a theorem that specifies a time-uniform bound on the weak-sense L_p error.

THEOREM 2. *Suppose $\widehat{\eta}'_t$ is defined by* (10) *and 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*

$$\sup_{t \geq 0} \mathbb{E}\{|[\widehat{\eta}'_t - \eta'_t](h_t)|^p\}^{1/p} \leq \frac{\epsilon_{u,m} c^{1/p}(p)}{\sqrt{N}}(q_u^{1/p}\sqrt{\chi} + (1 - q_u)^{1/p}),$$

where the constant $\epsilon_{u,m}$ *is*

$$(15) \qquad \epsilon_{u,m} = m\big(2 - \epsilon_u(M)\epsilon_u^{mK_u}(G)\big)/\epsilon_u^3(M)\epsilon_u^{(2m-1)K_u}(G).$$

PROOF. This and other technical proofs can be found in Section 8. □

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*

$$\sup_{t \geq 0} \mathbb{E}\{|[\widehat{\eta}'_t - \eta'_t](h_t)|^p\}^{1/p}$$

$$\leq \epsilon_{u,m} c^{1/p}(p)\Big(q_u^{1/p}\Big[\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{N_b}}\Big] + (1 - q_u)^{1/p}\frac{1}{\sqrt{N}}\Big).$$

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*

$$(16) \quad \sup_{t \geq 0} \mathbb{E}\{|[\widehat{\eta}'_t - \eta'_t](h_t)|^p\}^{1/p} \leq \frac{\epsilon_{\mathrm{u},m} c^{1/p}(p)}{\sqrt{N}} (q_{\mathrm{u}} \chi^{p/2} + (1 - q_{\mathrm{u}}))^{1/p}.$$

The intuitive implication of Theorem 2 and Corollary 3 is that rare approximation events have limited effect on the average approximation error of the subsample leader node particle filter. The L_2 error bound for the standard particle filter is the same as (16) 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 approximation error 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.

THEOREM 3. *Suppose assumptions* $(G)_{\mathrm{u}}$ *and* $(M)_{\mathrm{u}}^{(m)}$ *hold. Suppose further that* $\mathbb{P}\{\Delta'_i = 1\} \leq q_{\mathrm{u}}$ *for* $i \geq 0$ *and* $0 \leq q_{\mathrm{u}} \leq 1$. *Then for any* $N_{\mathrm{b}} < N$, $t \geq 0$ *and* $h_t \in \mathrm{Osc}_1(E_t)$ *we have*

$$\sup_{t \geq 0} \mathbb{P}\{|[\widehat{\eta}'_t - \eta'_t](h_t)| \geq \epsilon\} \leq \left(1 + 4\sqrt{2\pi} \frac{\varepsilon\sqrt{N}}{\epsilon_{\mathrm{u},m}}\right) e^{-N\varepsilon^2/(2\epsilon_{\mathrm{u},m}^2)}$$

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

The implication of this theorem is that the tail probabilities of the approximation error can be significantly affected by the rare hand-off events. Although the average approximation error bounds obtained in Corollary 3 appear encouraging, care should be exercised when selecting approximation parameters to prevent the explosion of the tails of the approximation error distribution. These tails characterize the probabilities of relatively rare, but catastrophic events.

4. Particle filtering with intermittent parametric approximations. In this section we analyze the error behavior of the leader node particle filter described by recursion (12). This filter incorporates intermittent parametric mixture estimation of the filtering probability density. The probability density estimation problem consists of estimating an unknown probability density given the i.i.d. sample $\{\xi_i\}_{1 \le i \le N}$ 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, $\Phi'_{i+1}(\widehat{\eta}'_i)$, there exists an associated and well-behaved sequence of approximate filtering densities $\frac{\mathrm{d}}{\mathrm{d}x_{i+1}} \Phi'_{i+1}(\widehat{\eta}'_i)$ 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 behavior of the parametric approximation leader node particle filter.

4.1. *Parametric approximation.* Within the Greedy Maximum Likelihood (GML) framework proposed by Li and Barron [16], the discrepancy between the target density f and its estimate is measured by the Kullback–Leibler divergence. For any two measures ν and μ on E, KL-divergence can be defined as follows:

$$(17) \qquad D(\nu \| \mu) = \int \log \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \, \mathrm{d}\nu.$$

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

$$(18) \qquad D(f \| g) = \int \log \frac{f(x)}{g(x)} f(x) \, \mathrm{d}x.$$

Consider the following class of bounded parametric probability densities:

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

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

$$\mathcal{C}_{N_p, i} = \mathrm{conv}_{N_p}(\mathcal{H}_i)$$

$$= \left\{ g : g(x) = \sum_{j=1}^{N_p} \alpha_{i,j} \phi_{\theta_{i,j}}(x), \phi_{\theta_{i,j}} \in \mathcal{H}_i, \sum_{j=1}^{N_p} \alpha_{i,j} = 1, \alpha_{i,j} \ge 0 \right\}.$$

Algorithm 1: GML

---

1 Given $g_1 \in \mathcal{H}$,
2 **for** $k = 2$ *to* N_{p} **do**
3   Find $\phi_{\theta_k} \in \mathcal{H}$ and $0 \leq \alpha_k \leq 1$ to maximize the function
4     $(\theta_k^*, \alpha_k^*) = \arg\max_{\alpha_k, \theta_k} \sum_{j=1}^{N} \log((1 - \alpha_k)g_{k-1}(\xi_j) + \alpha_k \phi_{\theta_k}(\xi_j))$.
5   Let $g_k = (1 - \alpha_k^*)g_{k-1} + \alpha_k^* \phi_{\theta_k^*}$.
6 **endfor**

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

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

The general framework for the greedy approximation of arbitrary cost functions is discussed in [24]. The particular instance of this more general framework is the GML for mixture approximation (see [16]). The corresponding computational routine, a sequential greedy maximum likelihood, associated with this procedure and based on the sample $(\xi_i)_{1 \leq i \leq N}$ from the target density f is summarized in the form of Algorithm 1. The optimization step in this algorithm can be performed with any standard numerical nonlinear optimization technique.

4.2. *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 [16]. Second, there are several bounds on approximation and sampling errors of Algorithm 1 in terms of KL-divergence (see [16] and [21]). In this section we extend the existing results and perform the L_p error analysis. Third, it was shown [16] that the performance of the greedy algorithm converges to the performance of the optimal mixture estimation algorithm as N and N_{p} become large.

Here we state the relevant results from [16] 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 (bias) 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 $i = 2, \ldots, N_{\mathrm{p}}$ (see Algorithm 1):

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

Alternatively, $\widehat{g}^{N_p} \in \mathcal{C}_{N_p}$ is an empirical N_p-mixture estimator constructed using Algorithm 1 based on a sample from the target density, f.

The following theorem (see [16]) reveals an important general property of the GML algorithm. It bounds the divergence between the target density and the analytical estimator g^{N_p}. The bound is the sum of two terms. The first is the divergence between the target density and an arbitrary approximating density $g_\mathcal{C} \in \mathcal{C}$. The second term involves γ, the upper bound on the log-ratio of two arbitrary functions from class $\mathcal{C}$, and $c_{f,\mathcal{C}}^2$, a class dependent constant (see [16] for more detail). For example, for the class of densities bounded below by a and above by b we have $c_{f,\mathcal{C}}^2 \le (b/a)^2$. This second term features N_p as a denominator, so it tends toward zero as the number of components in the mixture grows.

THEOREM 4 (Li and Barron [16], 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 of Theorem 4 is the following relationship between an arbitrary $g_\mathcal{C}(x) \in \mathcal{C}$ and the empirical GML algorithm output $\widehat{g}^{N_p} \in \mathcal{C}_{N_p}$ [16]:

$$(19) \qquad \frac{1}{N}\sum_{i=1}^{N}\log\widehat{g}^{N_p}(\xi_i) \ge \frac{1}{N}\sum_{i=1}^{N}\log g_\mathcal{C}(\xi_i) - \frac{\gamma c_{f,\mathcal{C}}^2}{N_p}.$$

Clearly, it also follows directly from Theorem 4 that $D(f\|g^{N_p}) \le D(f\|\mathcal{C}) + \frac{\gamma c_{f,\mathcal{C}}^2}{N_p}$. 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.

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. The bound consists of two components which arise because we split the total error into approximation error (the distance between the best analytical distribution g^{N_p} and f) and sampling error (the additional error arising because the empirical estimator $\widehat{g}^{N_p}$ is derived from a sample from f, rather than f itself). The approximation error bound follows directly from Theorem 4.

The bound on the sampling error is expressed in terms of the packing number $\mathcal{D}(\varepsilon, \mathcal{H}, d_N)$, which is the maximum number of ε-separated points in $\mathcal{H}$ (the class of parametric density functions) and the entropy integral

$$\int_0^b \sqrt{\log(1 + \mathcal{D}(\varepsilon, \mathcal{H}, d_N))} \, d\varepsilon,$$

both defined with respect to the empirical semimetric d_N, which, in its turn, is defined for $h_1, h_2 \in \mathcal{H}$ as follows:

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

Examples of classes of functions with converging entropy integral can be found in [21] and [22].

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 $\|h\|_{osc} \leq 1$, $p \geq 1$, and $N, N_p \in \mathbb{N}$ we have*

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

where $c(p) = 1$ if $1/2 \leq p < 1$ and C is a universal constant.[4]

The following corollary addresses the special case when the target density f lies within the class of continuous convex combinations, $\mathcal{C}$. In this case, the approximation error bound approaches 0 as the number of mixture components grows.

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

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

---

[4]See [22] for details.

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 $\gamma = 4\log(3\sqrt{e}(b/a))$. □

4.3. *Time-uniform error bounds.* In this section we present a result specifying time-uniform error bounds for the leader node particle filter performing parametric approximation. The result links the properties of Markov transitions M_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.

In the context of the GML algorithm and the leader node recursion (12) the operator $\mathbb{W}_{N_p}$ can be described as follows:

$$\frac{d}{dx_{i+1}}\mathbb{W}_{N_p} \circ S^N(\Phi'_{i+1}(\widehat{\eta}'_i)) = \sum_{j=1}^{N_p} \alpha_{i,j}\phi_{\theta_{i,j}}.$$

This means that in this context our target density is $\frac{d}{dx_{i+1}}\Phi'_{i+1}(\widehat{\eta}'_i)$ and we obtain an i.i.d. sample from this density through the particle filtering step. Based on the i.i.d. sample we estimate the weights, $\alpha_{i,j}$, and parameters, $\theta_{i,j}$, of a mixture using Algorithm 1. In the following we study the conditions for the unbiased estimation of our target density and then formulate our main result for the parametric approximation leader node particle filter.

Suppose we can write the Markov kernel M_i via its density function $p_i(x_i|x_{i-1})$ [3]

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

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

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

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

$$\frac{d\Phi'_{i+1}(\widehat{\eta}'_i)}{dx_{i+1}} = \int p_{i+1}(x_{i+1}|x_i)\frac{G_i^{S_{\ell'_i}}(x_i)}{\eta'_i(G_i^{S_{\ell'_i}})}\,d\widehat{\eta}'_i.$$

Thus for a Markov kernel with $p_{\vartheta_{i+1}}(x_{i+1}) \in \mathcal{M}_{i+1}$ we can rewrite the previous equation with a suitable change of measure

$$\frac{d\Phi'_{i+1}(\widehat{\eta}'_i)}{dx_{i+1}} = \int_{\Theta_{i+1}} p_{\vartheta_{i+1}}(x_{i+1})\mathbb{P}(d\vartheta_{i+1}).$$

This implies that for an N-particle approximation $\widehat{\eta}_i'$ [see (12)] we have that $\frac{\mathrm{d}\Phi_{i+1}'(\widehat{\eta}_i')}{\mathrm{d}x_{i+1}} \in \mathrm{conv}_N(\mathcal{M}_{i+1})$ and, as N grows without bound, we have $\frac{\mathrm{d}\Phi_{i+1}'(\widehat{\eta}_i')}{\mathrm{d}x_{i+1}} \in \mathrm{conv}(\mathcal{M}_{i+1})$. Therefore the bias of the GML algorithm in the leader node particle filter setting is determined by the properties of Markov transition kernel M_{i+1} and the class of approximating densities $\mathcal{H}_{i+1}$. In particular, for the Markov kernel with $p_{\vartheta_{i+1}}(x_{i+1}) \in \mathcal{M}_{i+1}$ and a sufficiently rich class $\mathcal{H}_{i+1}$, such that $\mathcal{M}_{i+1} \subseteq \mathcal{H}_{i+1}$ we have asymptotically unbiased approximation [recall that $\mathcal{C}_{i+1} = \mathrm{conv}(\mathcal{H}_{i+1})$]

$$D\left(\frac{\mathrm{d}\Phi_{i+1}'(\widehat{\eta}_i')}{\mathrm{d}x_{i+1}} \Big\| \mathcal{C}_{i+1}\right) = 0.$$

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 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 $\mathcal{M}_i = \{p_{\vartheta_i}(x_i) : \vartheta_i \in \Theta_i \subset \mathbb{R}^{d_i}\}$. Algorithm 1 exploits such classes $\mathcal{H}_i$ that there exist strictly positive numbers $a_{\mathrm{u}} = \inf_{i \geq 0} a_i$, $b_{\mathrm{u}} = \sup_{i \geq 0} b_i$ satisfying $0 < a_{\mathrm{u}} < b_{\mathrm{u}} < \infty$ and for any $i \geq 0$ we have

$$\mathcal{M}_i \subseteq \mathcal{H}_i.$$

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 $\widehat{\eta}_t'$ is defined by* (12) *and assumptions* $(G)_{\mathrm{u}}$, $(M)_{\mathrm{u}}^{(m)}$ *and* $(\mathcal{H})_{\mathrm{u}}$ *hold. Suppose further that* $\mathbb{P}\{\Delta_i' = 1\} \leq q_{\mathrm{u}}$ *for any* $i \geq 0$ *and* $0 \leq q_{\mathrm{u}} \leq 1$. *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*

$$\sup_{t \geq 0} \mathbb{E}\{|[\widehat{\eta}_t' - \eta_t'](h_t)|^p\}^{1/p}$$

$$\leq \epsilon_{\mathrm{u},m}\left[\frac{c^{1/p}(p)}{\sqrt{N}} + q_{\mathrm{u}}^{1/p}\left[\frac{16}{a_{\mathrm{u}}\sqrt{N}}\left(2c^{2/p}(p/2) + C(p/4)!\right.\right.\right.$$

$$\times \sup_{i \geq 0} \mathbb{E} \int_0^{b_i} \log^{1/2}\left(1 + \mathcal{D}(\varepsilon, \mathcal{H}_i, d_N)\right) \mathrm{d}\varepsilon\Bigg)$$

$$\left.\left.+ 8\log\left(\frac{3\sqrt{e}b_{\mathrm{u}}}{a_{\mathrm{u}}}\right)\frac{b_{\mathrm{u}}^2}{a_{\mathrm{u}}^2 N_{\mathrm{p}}}\right]^{1/2}\right].$$

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 ($\frac{c^{1/p}(p)}{\sqrt{N}}$) correspond-
ing to the normal operation of the filter and the other capturing the impact of the
additional parametric approximation. The bound on this second contribution is de-
rived directly from the bound expressed in Corollary 4. The theorem establishes
a 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 mem-
ber of the class $\mathcal{H}_i$. This condition is reminiscent of the modeling assumptions
that underpin Gaussian sum particle filtering (see, e.g., [10]), where the premise
is that the filtering density can asymptotically be represented as an infinite sum of
Gaussians.

5. Numerical experiments. In this section we present the results of numer-
ical experiments exploring the performance of the leader node particle filter. The
experiments provide an example of how the subsampling and parametric approxi-
mation 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 behavior 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 [9]

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

Here r_0 is a constant (we set $r_0 = 0.02$), and φ_t, v_t are independent and uniformly
distributed $v_t \sim U[0, 1]$, $\varphi_t \sim U[-\pi, \pi]$. $K_l = 20$ 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
$r_c = \sqrt{2\log(K_s)/K_s}$. We assume that any active leader node can route an approx-
imation of its posterior representation to any other potential leader node.

The measurement equation of every satellite sensor is the binary detector [1]
capable of detecting a target within radius r_d with probability p_d and false alarm
rate p_f

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

where the detection region $\mathcal{X}_d^j$ of satellite sensor j is defined as $\mathcal{X}_d^j = \{x : \|x - s_j\|_2 \le r_d\}$. To perform sensor selection step we use the mutual information (MI)
criterion [17]

$$(20) \qquad \ell_{t+1} = \arg\max_{\ell_{t+1}\in\mathcal{L}} I(X_{t+1}, Y_{t+1}^{\mathcal{S}_{\ell_{t+1}}} | y_1^{\mathcal{S}_{\ell_1}}, \ldots, y_t^{\mathcal{S}_{\ell_t}}).$$

Here the mutual information is defined as

$$I(X_{t+1}, Y_{t+1}^{\mathcal{S}_{\ell_{t+1}}} | y_1^{\mathcal{S}_{\ell_1}}, \ldots, y_t^{\mathcal{S}_{\ell_t}})$$

$$= \int p(x_{t+1}, y_{t+1}^{\mathcal{S}_{\ell_{t+1}}} | y_1^{\mathcal{S}_{\ell_1}}, \ldots, y_t^{\mathcal{S}_{\ell_t}})$$

$$\times \log\left(\frac{p(x_{t+1}, y_{t+1}^{\mathcal{S}_{\ell_{t+1}}} | y_1^{\mathcal{S}_{\ell_1}}, \ldots, y_t^{\mathcal{S}_{\ell_t}})}{p(x_{t+1} | y_1^{\mathcal{S}_{\ell_1}}, \ldots, y_t^{\mathcal{S}_{\ell_t}}) p(y_{t+1}^{\mathcal{S}_{\ell_{t+1}}} | y_1^{\mathcal{S}_{\ell_1}}, \ldots, y_t^{\mathcal{S}_{\ell_t}})} \right) dx \, dy,$$

$y_1^{\mathcal{S}_{\ell_1}}, \ldots, y_t^{\mathcal{S}_{\ell_t}}$ 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}.

Williams et al. pointed out in [23] 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 finite-time horizon dynamic program [23]. 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 applies (20) to determine if the current particle representation should be transferred to a new leader node that is more likely to make informative measurements. The leader node exchange rule can then be represented as

$$\Delta_{t+1} = \begin{cases} 1, & \text{if } (u_t \geq \lambda) \text{ and } (\ell_{t+1} \neq \ell_t), \\ 0, & \text{otherwise,} \end{cases}$$

where u_t is uniformly distributed in $[0, 1]$. 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. In our experiments we fix $\lambda = 1/5$. Note that equations for Δ_{t+1} and ℓ_{t+1} define the structure of an example sensor management rule $\Upsilon_t^{\ell_t}$, which was formulated in a more general form in Section 1.

We consider two leader node particle filtering algorithms, with one employing nonparametric approximation (subsampling) and the other using parametric approximation. To create a subsample for transmission in the nonparametric framework we use the general residual resampling scheme [7]. 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 [19] for details of the implementation).
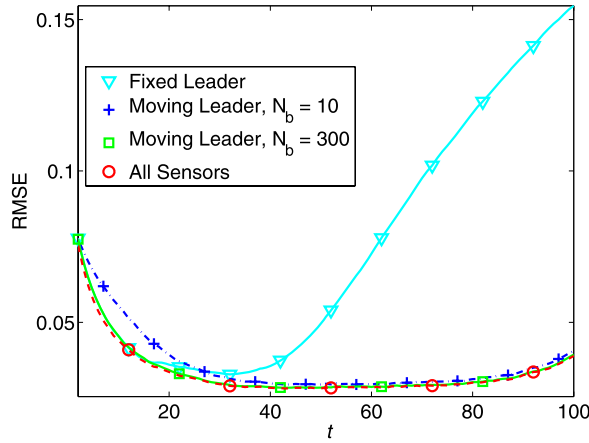
FIG. 2. *Performance* (*RMSE*) *of different fusion schemes versus time*: ▽ *denotes the scheme with fixed leader node selected at initialization*; + *denotes the scheme with leader node selected using approximate Mutual Information* (*MI*) *criterion and nonparametric* (*subsampling*) *approximation with $N_b = 10$;* □ *denotes the scheme with leader node selected using approximate MI criterion but no subsampling approximation* ($N_b = 300$); *and* ○ *denotes the centralized scheme using the entire set of measurements from all sensors.*

In the following we report the simulation results obtained using the set-up discussed above. All results are achieved using 5,000 Monte Carlo trials, and in each trial a new trajectory of the target is generated.

Figure 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 ▽ 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 □ respectively correspond to nonparametric leader node algorithms using $N_b = 10$ and $N_b = 300$ particles for communications, respectively. The fourth scheme denoted by ○ 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 (this value was selected after experimentation with multiple values of N because it provides sufficient accuracy without inducing unnecessary computational overhead) in all scenarios (so the $N_b = 300$ case corresponds to no subsampling). We can see from Figure 2 that the centralized scheme is only marginally better than the leader node scenario without compression ($N = N_b = 300$). This confirms that our leader node selection based on the approximate mutual information is a valid approach.

The leader node particle filter that uses a very small number of transmitted particles ($N_b = 10$) 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.

In the next set of results, we explore the approximation error, that is, 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 $N = 3{,}000$ particles, with no approximation during hand-off. For each of the 5,000 Monte Carlo trials, we apply this reference filter to generate location estimates. The approximation error for our test filters is measured relative to these estimates rather than the true locations.

Figure 3 depicts how the approximation performance is affected as the number of particles in the subsampling step (N_b) changes and 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 additional approximation.

Figure 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 behavior [see Figure 3(a), $N_b = 10$ in particular]. The transient in these curves arises because the particle representation of the target location density is initially highly dispersed and multi-modal. However, as time progresses ($t \in [51, 100]$) the particle representation of the target becomes more localized and
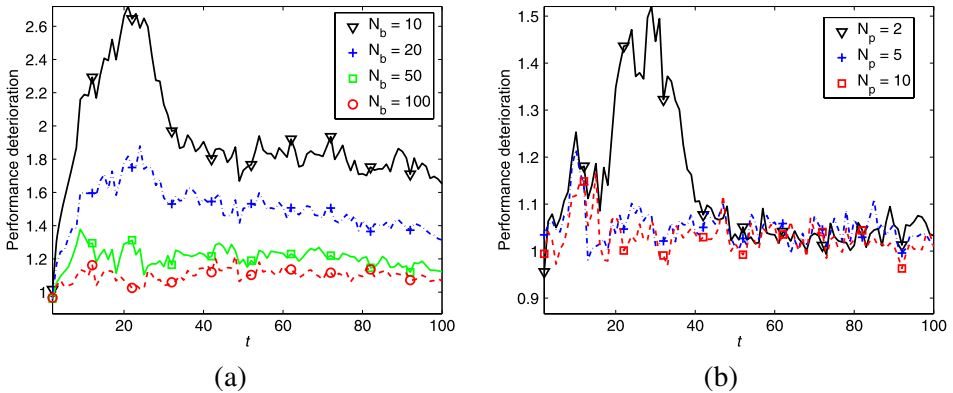

FIG. 3. *Deterioration of performance as a function of* (a) *varying number of transmitted particles and* (b) *varying number of transmitted mixture components for the leader node particle filter. The performance deterioration is measured as the ratio of the RMSAE of a leader node particle filter* (*with subsampling or parametric approximation*) *to that of a leader node particle filter with no approximation* ($N_b = 300$). (a) *Nonparametric leader node particle filter.* (b) *Parametric leader node particle filter.*
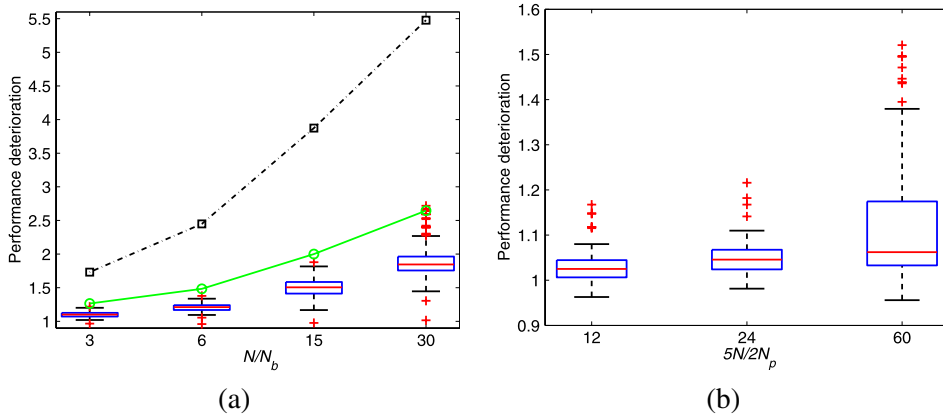
FIG. 4. *The relationship between deterioration of approximation performance and compression factor. The performance deterioration is measured as the ratio of the RMSAE of a leader node particle filter (subsampling or parametric approximation) to that of a leader node particle filter with no approximation ($N_b = 300$). The compression factor is the ratio of N to the number of values transmitted during leader node exchange. The boxes show lower quartile, median and upper quartile. Whiskers depict 1.5 times the interquartile range, and the + values denote outliers. (a) Nonparametric leader node particle filter. □ denotes the naive performance deterioration characterization, $\sqrt{N/N_b}$. ○ denotes the proposed characterization captured by Corollary 3. (b) Parametric leader node particle filter.*

closer to unimodal, 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 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 $2N/5N_p$. 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, Corollary 3 provides an analytical bound on the expected approximation error. The curve based on this result [depicting the factor $(q_u \chi + (1 - q_u))^{1/2}$; experimentally measured q_u never exceeds $\lambda/2$] is shown in Figure 4(a) and provides a meaningful characterization of the expected performance deterioration. For comparison purposes, we include a bound derived based on a simple worst-case assumption that the subsample approximation particle fil-

ter performs only as well as a particle filter that uses N_b particles at all times. The bound developed in this paper clearly provides a better indication of the performance deterioration.

6. Related work. In [11] Kunita studied the asymptotic behavior of the error and stability of the filter that has an ergodic signal transition semigroup with respect to the initial distribution. Ocone and Pardoux [18] addressed the stability of linear filters with respect to a non-Gaussian initial condition and examined the stability of nonlinear filters in the case where the signal diffusion is convergent. Although interesting, the results in [11, 18] 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 [2–5, 7, 12, 13, 15].

The Feynman–Kac semigroup approach to the stability analysis of particle filters has been described and developed by Del Moral, Miclo and Guionnet in [3–5]. The authors study the stability properties of general nonlinear 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 [5], 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 [3], 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). There have been some efforts to relax these types of assumptions. In [13, 15], Le Gland and Oudjane study the stability and convergence rates for particle filters using the Hilbert projective metric. In [15], they relax the signal mixing assumptions by employing a specific, "robust" particle filter architecture with truncated likelihood functions.

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 [2]; in these systems the population size also varies because at each time step a particle generates a random number of offspring. 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 [7].

Le Gland et al. provide uniform convergence results for the regularized particle filters [12, 13]. 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). From an algorithmic standpoint, there are also similarities with the Gaussian sum particle filter [10], but the theoretical analysis of this filter is less developed.

There has been some work addressing the analysis of the leader node particle filter [9]. 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 [9].

7. Concluding remarks. We have presented the analysis of the leader node particle filter that performs intermittent approximation. Our main results have the form of upper bounds on the expected L_p approximation error of the leader node particle filter that occasionally employs 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 significant (a subset of subsamples or a few mixture components are used during leader node exchange), and the error remains reasonable. Numerical experiments indicate that the bound for the subsample approximation particle filter provides a meaningful characterization of practical approximation performance.

8. Proofs of theorems.

PROOF OF THEOREM 2. We begin by applying Minkowski's inequality to (13):

$$\mathbb{E}\{|[\widehat{\eta}_t' - \eta_t'](h_t)|^p\}^{1/p} \leq \sum_{i=0}^{t} \mathbb{E}\{|[\Phi_{i,t}'(\widehat{\eta}_i') - \Phi_{i,t}'(\Phi_i'(\widehat{\eta}_{i-1}'))](h_t)|^p\}^{1/p},$$

and then using (7), (3) and (4) we have

$$\sum_{i=0}^{t} \mathbb{E}\{|[\Phi_{i,t}'(\widehat{\eta}_i') - \Phi_{i,t}'(\Phi_i'(\widehat{\eta}_{i-1}'))](h_i)|^p\}^{1/p}$$

$$\leq \frac{2 - \epsilon_{\mathrm{u}}(M)\epsilon_{\mathrm{u}}^{mK_{\mathrm{u}}}(G)}{\epsilon_{\mathrm{u}}(M)\epsilon_{\mathrm{u}}^{mK_{\mathrm{u}}}(G)}$$

$$\times \sum_{i=0}^{t} \left(1 - \epsilon_{\mathrm{u}}^2(M)\epsilon_{\mathrm{u}}^{(m-1)K_{\mathrm{u}}}(G)\right)^{\lfloor (t-i)/m \rfloor} \mathbb{E}\{|[\widehat{\eta}_i' - \Phi_i'(\widehat{\eta}_{i-1}')](h_i)|^p\}^{1/p}.$$

Next we analyze each individual expectation under the sum above. In particular, using the structure of the algorithm defined in (10) and the definition of sampling operator introduced in (9) we can rewrite the terms comprising the sum in the following explicit way:

$$\mathbb{E}\{|[\widehat{\eta}'_i - \Phi'_i(\widehat{\eta}'_{i-1})](h_i)|^p\}^{1/p}$$

$$(21) \qquad = \mathbb{E}\{|[\Delta'_i S^N \circ S^{N_b}(\Phi'_i(\widehat{\eta}'_{i-1})) + (1 - \Delta'_i)S^N(\Phi'_i(\widehat{\eta}'_{i-1}))$$
$$- \Phi'_i(\widehat{\eta}'_{i-1})](h_i)|^p\}^{1/p}.$$

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

$$\mathbb{E}\{|[\widehat{\eta}'_i - \Phi'_i(\widehat{\eta}'_{i-1})](h_i)|^p\}^{1/p}$$

$$\leq \mathbb{E}\{|\Delta'_i[S^N \circ S^{N_b}(\Phi'_i(\widehat{\eta}'_{i-1})) - \Phi'_i(\widehat{\eta}'_{i-1})](h_i)|^p\}^{1/p}$$

$$+ \mathbb{E}\{|(1 - \Delta'_i)[S^N(\Phi'_i(\widehat{\eta}'_{i-1})) - \Phi'_i(\widehat{\eta}'_{i-1})](h_i)|^p\}^{1/p}.$$

Adding and subtracting $\Delta'_i S^{N_b}(\Phi'_i(\widehat{\eta}'_{i-1}))$ in the first term, we have

$$\mathbb{E}\{|[\widehat{\eta}'_i - \Phi'_i(\widehat{\eta}'_{i-1})](h_i)|^p\}^{1/p}$$

$$(22) \qquad \leq \mathbb{E}\{|\Delta'_i[S^N \circ S^{N_b}(\Phi'_i(\widehat{\eta}'_{i-1})) - S^{N_b}(\Phi'_i(\widehat{\eta}'_{i-1}))](h_i)|^p\}^{1/p}$$

$$+ \mathbb{E}\{|\Delta'_i[S^{N_b}(\Phi'_i(\widehat{\eta}'_{i-1})) - \Phi'_i(\widehat{\eta}'_{i-1})](h_i)|^p\}^{1/p}$$

$$+ \mathbb{E}\{|(1 - \Delta'_i)[S^N(\Phi'_i(\widehat{\eta}'_{i-1})) - \Phi'_i(\widehat{\eta}'_{i-1})](h_i)|^p\}^{1/p}.$$

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 $N = \chi N_b$, 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 (22) is zero.

The analysis of the remaining two terms is similar. We first concentrate on the second term. Recall that $(\ell'_i, \Delta'_i) = \Upsilon^{\ell'_{i-1}}_{i-1}(\Phi'_i(\widehat{\eta}^{\ell'_{0:i-1}}_{i-1}), \mathcal{I}_{i-1})$. Thus given the σ-algebra $\mathcal{F}_{i-1}$ and the realization of the measurement taken by leader node ℓ'_{i-1}, $Y^{\mathcal{S}_{\ell'_{i-1}}}_{i-1} = y^{\mathcal{S}_{\ell'_{i-1}}}_{i-1}$, the output of the decision rule is independent of the sampling error, $[S^N(\Phi'_i(\widehat{\eta}'_{i-1})) - \Phi'_i(\widehat{\eta}'_{i-1})](h_i)$. We exploit this Markovian nature of the decision rule and apply Lemma 1 to the conditional expectation rendering the following bound:

$$\mathbb{E}\{|\Delta'_i[S^{N_b}(\Phi'_i(\widehat{\eta}'_{i-1})) - \Phi'_i(\widehat{\eta}'_{i-1})](h_i)|^p\}^{1/p}$$

$$(23) \quad = \mathbb{E}\{\Delta'_i \mathbb{E}\{|[S^{N_b}(\Phi'_i(\widehat{\eta}'_{i-1})) - \Phi'_i(\widehat{\eta}'_{i-1})](h_i)|^p|\mathcal{F}_{i-1}, Y^{\mathcal{S}_{\ell'_{i-1}}}_{i-1} = y^{\mathcal{S}_{\ell'_{i-1}}}_{i-1}\}\}^{1/p}$$

$$\leq \frac{c^{1/p}(p)}{\sqrt{N_b}} q_i^{1/p}.$$

Combining the analysis results for all three terms, we obtain

$$\mathbb{E}\{|[\widehat{\eta}_i' - \Phi_i'(\widehat{\eta}_{i-1}')](h_t)|^p\}^{1/p} \leq c^{1/p}(p)\Big(q_i^{1/p}\frac{1}{\sqrt{N_b}} + (1-q_i)^{1/p}\frac{1}{\sqrt{N}}\Big).$$

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 $q_i = q_{max}$,

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

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

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

Thus if $q_u \leq 2/3 \leq q_{max}$, then for any $i \geq 0$ we have

$$\mathbb{E}\{|[\widehat{\eta}_i' - \Phi_i'(\widehat{\eta}_{i-1}')](h_t)|^p\}^{1/p} \leq c^{1/p}(p)\Big(q_u^{1/p}\frac{1}{\sqrt{N_b}} + (1-q_u)^{1/p}\frac{1}{\sqrt{N}}\Big).$$

Finally, noting [3] that

$$(24) \qquad \sum_{i=0}^{t}(1 - \epsilon_u^2(M)\epsilon_u^{(m-1)K_u}(G))^{\lfloor(t-i)/m\rfloor} \leq \frac{m}{\epsilon_u^2(M)\epsilon_u^{(m-1)K_u}(G)},$$

we complete the proof of theorem. $\square$

PROOF OF COROLLARY 2. The corollary follows by allowing for sampling error to arise in the first term in (22):

$$\mathbb{E}\{|\Delta_i'[S^N \circ S^{N_b}(\Phi_i'(\widehat{\eta}_{i-1}')) - S^{N_b}(\Phi_i'(\widehat{\eta}_{i-1}'))](h_i)|^p\}^{1/p} \leq \frac{c^{1/p}(p)}{\sqrt{N}}q_i^{1/p}$$

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

PROOF OF COROLLARY 3. Starting with (21), we perform a different error decomposition expanding the power. We observe that $\Delta_i'(1 - \Delta_i') = 0$ and that if $N = \chi N_b$ for integer χ, we can reconstruct an N-sample representation from the N_b sample with no additional error. Thus

$$\mathbb{E}\{|[\widehat{\eta}_i' - \Phi_i'(\widehat{\eta}_{i-1}')](h_i)|^p\}^{1/p}$$
$$\leq \mathbb{E}\{\Delta_i'|[S^{N_b}(\Phi_i'(\widehat{\eta}_{i-1}')) - \Phi_i'(\widehat{\eta}_{i-1}')](h_i)|^p$$
$$+ (1 - \Delta_i')|[S^N(\Phi_i'(\widehat{\eta}_{i-1}')) - \Phi_i'(\widehat{\eta}_{i-1}')](h_i)|^p\}^{1/p}.$$

Applying the same conditioning as in (23) and utilizing Lemma 1,

$$(25) \qquad \mathbb{E}\{|[\widehat{\eta}_i' - \Phi_i'(\widehat{\eta}_{i-1}')](h_i)|^p\}^{1/p} \leq \frac{c(p)^{1/p}}{\sqrt{N}} (q_i \chi^{p/2} + (1 - q_i))^{1/p}.$$

We note that $\chi \geq 1$ and $q_i \chi + (1 - q_i) \leq q_u \chi + (1 - q_u)$ under the assumption $q_i \leq q_u$. The final step in the proof involves applying (24) as in the proof of Theorem 2. $\square$

PROOF OF THEOREM 3. Using the triangle inequality in (13), following the methodology presented in Theorem 2 and denoting $\omega_i = (1 - \epsilon_u^2(M) \times \epsilon_u^{(m-1)}(G))^{\lfloor (t-i)/m \rfloor}$ and $a = \frac{2 - \epsilon_u(M)\epsilon_u^m(G)}{\epsilon_u(M)\epsilon_u^m(G)}$ we have

$$|[\widehat{\eta}_t' - \eta_t'](h_t)| \leq a \sum_{i=0}^{t} \omega_i |[\widehat{\eta}_i' - \Phi_i'(\widehat{\eta}_{i-1}')](h_i)|.$$

Using the structure of the algorithm defined in (10) and the definition of sampling operator introduced in (9), we obtain the following (similarly to Theorem 2):

$$|[\widehat{\eta}_t' - \eta_t'](h_t)| \leq Z_1 + Z_2,$$

where

$$Z_1 = a \sum_{i=0}^{t} \omega_i \Delta_i' |[S^N \circ S^{N_b}(\Phi_i'(\widehat{\eta}_{i-1}')) - S^{N_b}(\Phi_i'(\widehat{\eta}_{i-1}'))](h_i)|$$

$$+ a \sum_{i=0}^{t} \omega_i (1 - \Delta_i') |[S^N(\Phi_i'(\widehat{\eta}_{i-1}')) - \Phi_i'(\widehat{\eta}_{i-1}')](h_i)|,$$

$$Z_2 = a \sum_{i=0}^{t} \omega_i \Delta_i' |[S^{N_b}(\Phi_i'(\widehat{\eta}_{i-1}')) - \Phi_i'(\widehat{\eta}_{i-1}')](h_i)|.$$

Noting that

$$\sup_{t \geq 0} \mathbb{P}\{|[\widehat{\eta}_t' - \eta_t'](h_t)| \geq \epsilon\} \leq \sup_{t \geq 0} \mathbb{P}\{Z_1 + Z_2 \geq \epsilon\}$$

and using the fact that for any couple of random variables Z_1, Z_2 we have $(Z_1 + Z_2 \geq 1) \rightarrow ((Z_1 \geq 1/2)$ or $(Z_2 \geq 1/2))$ and thus $\mathbb{P}\{X + Y \geq \varepsilon\} \leq \mathbb{P}\{X \geq \varepsilon/2\} + \mathbb{P}\{Y \geq \varepsilon/2\}$, we have

$$\sup_{t \geq 0} \mathbb{P}\{|[\widehat{\eta}_t' - \eta_t'](h_t)| \geq \epsilon\} \leq \sup_{t \geq 0} \mathbb{P}\{Z_1 \geq \epsilon/2\} + \sup_{t \geq 0} \mathbb{P}\{Z_2 \geq \epsilon/2\}.$$

Now applying Markov's inequality we conclude

$$\sup_{t \geq 0} \mathbb{P}\{|[\widehat{\eta}_t' - \eta_t'](h_t)| \geq \epsilon\} \leq \sup_{t \geq 0} e^{-\tau_1 \epsilon/2} \mathbb{E}\{e^{\tau_1 Z_1}\} + \sup_{t \geq 0} e^{-\tau_2 \epsilon/2} \mathbb{E}\{e^{\tau_2 Z_2}\}.$$

Next we apply the exponential series expansion,

$$(26) \qquad \mathbb{E}\{e^{\tau_1 Z_1}\} = \sum_{n \geq 0} \frac{\tau_1^n}{n!} \mathbb{E}\{Z_1^n\},$$

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

$$\mathbb{E}\{Z_1^n\}^{1/n} = (\mathbb{E}\{Z_1^n | \Delta_i' = 1\}\mathbb{P}\{\Delta_i' = 1\} + \mathbb{E}\{Z_1^n | \Delta_i' = 0\}\mathbb{P}\{\Delta_i' = 0\})^{1/n}$$

$$\leq a \sum_{i=0}^{t} \omega_i \big(q_i c(n) N^{-n/2} + (1 - q_i) c(n) N^{-n/2}\big)^{1/n} = \frac{c^{1/n}(n)}{\sqrt{N}} a \sum_{i=0}^{t} \omega_i.$$

Noting that $a \sum_{i=0}^{t} \omega_i \leq \epsilon_{\mathrm{u},m}$ we have that $\mathbb{E}\{Z_1^n\} \leq \epsilon_{\mathrm{u},m}^n c(n) N^{-n/2}$. Substituting this into (26) and employing the same simplifications as in the proofs of Theorem 1 and Corollary 1 we obtain

$$e^{-\varepsilon \tau_1/2} \mathbb{E}\{e^{\tau_1 Z_1}\} \leq \left(1 + \sqrt{2\pi} \frac{\tau_1 \epsilon_{\mathrm{u},m}}{\sqrt{N}}\right) e^{\tau_1^2 \epsilon_{\mathrm{u},m}^2/(8N) - \varepsilon \tau_1/2}.$$

Applying similar analysis to $e^{-\varepsilon \tau_2/2} \mathbb{E}\{e^{\tau_2 Z_2}\}$ and choosing $\tau_1 = \frac{2\varepsilon N}{\epsilon_{\mathrm{u},m}^2}$ and $\tau_2 = \frac{2 N_{\mathrm{b}} \varepsilon}{\epsilon_{\mathrm{u},m}^2}$ completes the proof. $\quad\square$

PROOF OF THEOREM 5. Using Pinsker's inequality, $\int |f - g| \leq \sqrt{2 D(f \| g)}$, [6] we have

$$\mathbb{E}\{|[\widehat{\mathcal{G}}^{N_{\mathrm{p}}} - F](h)|^p\}^{1/p} \leq \sqrt{2} [\mathbb{E}\{D(f \| \widehat{g}^{N_{\mathrm{p}}})^{p/2}\}^{2/p}]^{1/2}.$$

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

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

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 [21]:

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

$$+ \frac{1}{N} \sum_{i=1}^{N} \log g^*(\xi_i) - \frac{1}{N} \sum_{i=1}^{N} \log \widehat{g}^{N_{\mathrm{p}}}(\xi_i)$$

$$+ \int \log g^*(x) F(\mathrm{d}x) - \frac{1}{N} \sum_{i=1}^{N} \log g^*(\xi_i).$$

Applying (19) to the middle term we see

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

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

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

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

This allows us to split the effect of approximation and sampling errors by applying Minkowski's inequality (since $p \geq 2$),

$$\mathbb{E}\{D(f\|\widehat{g}^{N_{\mathrm{p}}})^{p/2}\}^{2/p} \leq 2\mathbb{E}\Big\{\Big[\sup_{g\in\mathcal{C}} |[F - S^N(F)](\log g)|\Big]^{p/2}\Big\}^{2/p}$$

$$+ \frac{\gamma c_{f,\mathcal{C}}^2}{N_{\mathrm{p}}} + D(f\|\mathcal{C}).$$

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 ξ_k being the samples from μ)

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

Using the symmetrization lemma (see, e.g., Lemma 2.3.1 in [22] or Lemma 6.3 in [14]), we deduce

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

Denoting $\kappa(x) = g(x) - 1$ and using the fact [20] that $\varphi(\kappa(x)) = a\log(\kappa(x) + 1)$ is a contraction,[5] we apply the comparison inequality (Theorem 4.12 in [14]), observing that $[\cdot]^{p/2}$ is convex and increasing for $p \geq 2$, and κ is a bounded function

$$\mathbb{E}\Big\{\Big[\sup_{g\in\mathcal{C}} |S_\varepsilon^N(F)(\log g)|\Big]^{p/2}\Big\}^{2/p} \leq \frac{2}{a}\mathbb{E}\Big\{\Big[\sup_{g\in\mathcal{C}} |S_\varepsilon^N(F)(g)|\Big]^{p/2}\Big\}^{2/p}$$

$$+ \frac{2}{a}\mathbb{E}\{|S_\varepsilon^N(F)(1)|^{p/2}\}^{2/p}.$$

---

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

Applying the same technique used to prove Lemma 1 we have

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

On the other hand, using the representation of $g \in \mathcal{C}$ and exchanging the order of integration and summation

$$|S_\varepsilon^N(F)(g)| = \left|\frac{1}{N}\sum_{i=1}^N \varepsilon_i \int_{\theta\in\Theta} \phi_\theta(\xi_i)\mathbb{P}(\mathrm{d}\theta)\right|$$

$$\leq \sup_{\theta\in\Theta}\left|\frac{1}{N}\sum_{i=1}^N \varepsilon_i\phi_\theta(\xi_i)\right|,$$

and we conclude

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

The Orlicz norm [3, 22] $\pi_{\psi_p}(Y)$ 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 [22]. 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 [3] or [22], page 105, Problem 4), we deduce

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

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 [22])

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

Combining the above we have

$$\mathbb{E}\{|[\widehat{\mathcal{G}}^{N_\mathrm{p}} - F](h)|^p\}^{1/p}$$

$$\leq \sqrt{2}\left[\frac{8}{a\sqrt{N}}\left(2c^{2/p}(p/2) + (p/4)!C\mathbb{E}\int_0^b \sqrt{\log(1 + \mathcal{D}(\varepsilon, \mathcal{H}, d_N))}\,\mathrm{d}\varepsilon\right)\right.$$

$$\left.+ \frac{\gamma c_{f,\mathcal{C}}^2}{N_\mathrm{p}} + D(f\|\mathcal{C})\right]^{1/2}.$$

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

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

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

PROOF OF THEOREM 6. Using the same argument as in Theorem 2 we have

$$\mathbb{E}\{|[\widehat{\eta}_t' - \eta_t'](h_t)|^p\}^{1/p}$$

$$\leq \frac{2 - \epsilon_u(M)\epsilon_u^{mK_u}(G)}{\epsilon_u(M)\epsilon_u^{mK_u}(G)} \sum_{i=0}^t \beta_{i,t}(P)\mathbb{E}\{|[\widehat{\eta}_i' - \Phi_i'(\widehat{\eta}_{i-1}')](h_i)|^p\}^{1/p}.$$

Based on (12), $\widehat{\mathcal{G}}_{i+1}^{N_p} = \mathbb{W}_{N_p} \circ S^N(\Phi_{i+1}'(\widehat{\eta}_i^{\ell_{0,i}'}))$ and Minkowski inequality we have the decomposition for each individual expectation under the sum above:

$$\mathbb{E}\{|[\widehat{\eta}_i' - \Phi_i'(\widehat{\eta}_{i-1}')](h_i)|^p\}^{1/p}$$

$$\leq \mathbb{E}\{|\Delta_i'[S^N(\widehat{\mathcal{G}}_i^{N_p}) - \widehat{\mathcal{G}}_i^{N_p}](h_i)$$

$$+ (1 - \Delta_i')[S^N(\Phi_i'(\widehat{\eta}_{i-1}')) - \Phi_i'(\widehat{\eta}_{i-1}')](h_i)|^p\}^{1/p}$$

$$+ \mathbb{E}\{|\Delta_i'[\widehat{\mathcal{G}}_i^{N_p} - \Phi_i'(\widehat{\eta}_{i-1}')](h_i)|^p\}^{1/p}.$$

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

$$\mathbb{E}\{|\Delta_i'[\widehat{\mathcal{G}}_i^{N_p} - \Phi_i'(\widehat{\eta}_{i-1}')](h_i)|^p\}^{1/p}$$

$$= \mathbb{E}\{\mathbb{E}\{\Delta_i'|[\widehat{\mathcal{G}}_i^{N_p} - \Phi_i'(\widehat{\eta}_{i-1}')](h_i)|^p|\mathcal{F}_{i-1}, Y_{i-1}^{\mathcal{S}_{\ell_{i-1}}} = y_{i-1}^{\mathcal{S}_{\ell_{i-1}}}\}\}^{1/p}$$

$$\leq q_i^{1/p}\sqrt{2}\left[\frac{8}{a_i\sqrt{N}}\left(2c^{2/p}(p/2)\right.\right.$$

$$\left.+ (p/4)!C\mathbb{E}\int_0^{b_i}\sqrt{\log(1 + \mathcal{D}(\varepsilon, \mathcal{H}_i, d_N))}\,d\varepsilon\right)$$

$$\left.+ 4\log(3\sqrt{e}(b_i/a_i))\frac{(b_i/a_i)^2}{N_p}\right]^{1/2}.$$

Next we apply Lemma 1 and the same conditioning argument as in Theorem 3 to the remaining term and conclude that since $q_i \leq q_u$, then for any $i \geq 0$ we have the time-uniform estimate

$$\mathbb{E}\{|[\widehat{\eta}_i' - \Phi_i'(\widehat{\eta}_{i-1}')](h_i)|^p\}^{1/p}$$

$$\leq \frac{c^{1/p}(p)}{\sqrt{N}}$$

$$+ q_u^{1/p}\sqrt{2}$$

$$\times \left[\frac{8}{a_u\sqrt{N}}\left(2c^{2/p}(p/2)\right.\right.$$

$$+ (p/4)! C \sup_{i \geq 0} \mathbb{E} \int_0^{b_i} \sqrt{\log\big(1 + \mathcal{D}(\varepsilon, \mathcal{H}_i, d_N)\big)}\, d\varepsilon \bigg)$$

$$+ 4\log\big(3\sqrt{e}(b_u/a_u)\big) \frac{(b_u/a_u)^2}{N_p} \bigg]^{1/2}.$$

This along with (3) and (24) completes the proof of theorem. $\quad\square$

REFERENCES

[1] 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.

[2] CRISAN, D., DEL MORAL, P. and LYONS, T. (1999). Discrete filtering using branching and interacting particle systems. *Markov Process. Related Fields* **5** 293–318. MR1710982

[3] DEL MORAL, P. (2004). *Feynman–Kac Formulae*: *Genealogical and Interacting Particle Systems with Applications*. Springer, New York. MR2044973

[4] DEL MORAL, P. and GUIONNET, A. (2001). On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann. Inst. H. Poincaré Probab. Statist.* **37** 155–194. MR1819122

[5] DEL MORAL, P. and MICLO, L. (2002). On the stability of nonlinear Feynman–Kac semigroups. *Ann. Fac. Sci. Toulouse Math.* (6) **11** 135–175. MR1988460

[6] DEVROYE, L. and LUGOSI, G. (2001). *Combinatorial Methods in Density Estimation*. Springer, New York. MR1843146

[7] 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. IEEE, New York.

[8] DOUCET, A., DE FREITAS, N. and GORDON, N. (eds.) (2001). *Sequential Monte Carlo Methods in Practice*. Springer, New York. MR1847783

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

[10] KOTECHA, J. H. and DJURIĆ, P. M. (2003). Gaussian sum particle filtering. *IEEE Trans. Signal Process.* **51** 2602–2612. MR2003067

[11] KUNITA, H. (1971). Asymptotic behavior of the nonlinear filtering errors of Markov processes. *J. Multivariate Anal.* **1** 365–393. MR0301812

[12] 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. Prague, Czech Republic* 167–174. IEEE, New York.

[13] LE GLAND, F. and OUDJANE, N. (2004). Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters. *Ann. Appl. Probab.* **14** 144–187. MR2023019

[14] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces*: *Isoperimetry and Processes. Ergebnisse der Mathematik und Ihrer Grenzgebiete* (3) [*Results in Mathematics and Related Areas* (3)] **23**. Springer, Berlin. MR1102015

[15] LEGLAND, 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 Process. Appl.* **106** 279–316. MR1989630

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

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

[18] OCONE, D. and PARDOUX, E. (1996). Asymptotic stability of the optimal filter with respect to its initial condition. *SIAM J. Control Optim.* **34** 226–243. MR1372912

[19] ORESHKIN, B. N. and COATES, M. J. (2009). Error propagation in particle filters with approximation steps. Technical report, McGill Univ. Available at http://www.tsp.ece.mcgill.ca/Networks/publications-techreport.html.

[20] RAKHLIN, A. (2006). *Applications of Empirical Processes in Learning Theory*: *Algorithmic Stability and Generalization Bounds*. ProQuest LLC, Ann Arbor, MI. MR2717338

[21] RAKHLIN, A., PANCHENKO, D. and MUKHERJEE, S. (2005). Risk bounds for mixture density estimation. *ESAIM Probab. Stat.* **9** 220–229 (electronic). MR2148968

[22] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*: *With Applications to Statistics*. Springer, New York. MR1385671

[23] WILLIAMS, J. L., FISHER, J. W., III and WILLSKY, A. S. (2007). Approximate dynamic programming for communication-constrained sensor network management. *IEEE Trans. Signal Process.* **55** 4300–4311. MR2464439

[24] ZHANG, T. (2003). Sequential greedy approximation for certain convex optimization problems. *IEEE Trans. Inform. Theory* **49** 682–691. MR1967192

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

DEPARTMENT OF ELECTRICAL
  AND COMPUTER ENGINEERING
MCGILL UNIVERSITY
MONTREAL, QUEBEC H3A-2A7
CANADA
E-MAIL: boris.oreshkin@mail.mcgill.ca
           mark.coates@mcgill.ca