

# Particle Filters with Approximation Steps

Boris Oreshkin and Mark Coates

Department of Electrical and Computer Engineering  
McGill University, Montreal, Quebec, Canada

Email: boris.oreshkin@mail.mcgill.ca, mark.coates@mcgill.ca

**Abstract**—This paper examines the impact of approximation steps that become necessary when particle filters are implemented on resource-constrained platforms. We consider particle filters that perform intermittent approximation, either by subsampling the particles or by generating a parametric approximation. For such algorithms, we derive time-uniform bounds on the weak-sense  $L_p$  error. 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.

## I. INTRODUCTION

Particle filters have proven to be an effective approach for addressing difficult tracking problems. 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. We will motivate the theoretical analysis by considering the example of the “leader-node” particle filter, an algorithm for distributed tracking in sensor networks.

### A. Leader-node particle filter for sensor network tracking

The *leader node particle filter*, proposed in [1], [2] and refined and analyzed in [3], [4], strives to reduce the communication overhead. It is a local, distributed collaborative algorithm. At any time-step only one node performs the particle filtering (the *leader* node), and this node changes over time. The leader node is responsible for performing local tracking based on the data acquired by a set of *satellite* sensor nodes. The satellite nodes have sensing capabilities and locally transmit the acquired data to the nearest leader node. Sensor management strategies are used to determine when to change leader node [3]. When there is a change of leader-node, 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 (which can involve thousands of bits), the filtering distribution is more coarsely approximated, either by transmitting only a subset of the particles or by training a parametric model.

### B. Paper Organization

Section II overviews the analysis framework, describes the algorithms, and states the regularity conditions. Section III

presents the main results. Section IV describes numerical experiments and Section V concludes the paper.

## II. ANALYSIS FRAMEWORK

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

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

$$Y_t = g_t(X_t, \zeta_t). \quad (2)$$

Here  $X_t$  is the target state vector at time  $t$ ,  $Y_t$  is the measurement vector,  $\varrho_t$  and  $\zeta_t$  are system excitation and measurement noises;  $f_t, f_t : \mathbb{R}^{d_x} \rightarrow \mathbb{R}^{d_x}$ , and  $g_t, g_t : \mathbb{R}^{d_x} \rightarrow \mathbb{R}^{d_y}$ , are nonlinear system and measurement maps.

In order to conduct stability (error propagation) analysis, we need to introduce slightly more rigorous mathematical notation. Let  $(E_t, \mathcal{E}_t)$ ,  $t \in \mathbb{N}$  be a sequence of measurable spaces. The target state vector evolves according to a non-homogeneous (discrete-time) Markov chain  $X_t$  with transitions  $M_{t+1}$  from  $E_t$  into  $E_{t+1}$ . 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 finite supremum norm. We define a convex set  $\text{Osc}_1(E)$  of  $\mathcal{E}$ -measurable test functions with finite oscillations:

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

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

### A. Feynman-Kac models

The results we present in this paper are based on the representation of particle filters as  $N$ -particle approximations of a Feynman-Kac model. We now briefly review such models; see [5] for detailed discussion.

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

$$\Pr\{X_t \in dx_t | X_{t-1} = x_{t-1}\} = M_t(x_{t-1}, dx_t) \quad (3)$$

and the initial signal distribution  $\mu_0$ :

$$\mathbb{P}_{\mu, t}(d(x_0, \dots, x_t)) = \mu(dx_0)M_1(x_0, dx_1) \dots M_t(x_{t-1}, dx_t).$$

Bounded and non-negative potential functions  $G_t : E_t \rightarrow [0, \infty)$  characterize the properties of the observation process in (2). This leads to the following definition of the unnormalized *prediction* Feynman-Kac model, for  $h_t \in \mathcal{B}_b(E_t)$  and  $t \in \mathbb{N}$ .

$$\gamma_t(h_t) = \int_{E_{[0:t]}} h_t(X_t) \prod_{i=0}^{t-1} G_i(X_i) \mathbb{P}_{\mu, t}(d(x_0, \dots, x_t))$$

The normalized prediction Feynman-Kac model is then:

$$\eta_t(h_t) = \gamma_t(h_t)/\gamma_t(1)$$

The Boltzmann-Gibbs transformation  $\Psi_t$  reflects the effect of the likelihood function  $G_t$  (cf. Bayes update for some  $\nu$ ):

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

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

$$\begin{aligned} \Phi_t(\eta_{t-1}) &= \Psi_{t-1}(\eta_{t-1}) M_t \\ \eta_t &= \Phi_t(\eta_{t-1}) \end{aligned}$$

We can define a particle filter by developing an  $N$ -particle approximation to the Feynman-Kac model. Let the sampling operator  $S^N : \mathcal{P}(E) \rightarrow \mathcal{P}(E^N)$  be defined as:

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

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

### B. Regularity Conditions

The analysis we present relies on certain assumptions about the regularity and mixing properties of the Markov kernels and likelihood potential functions. For deriving time-uniform bounds, we adopt assumptions similar to the ones used in [5].

We define the following condition on the Markov kernels.

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

$$M_{i,i+m}(x_i, \cdot) = M_{i+1} \dots M_{i+m}(x_i, \cdot) \geq \epsilon_u(M) M_{i,i+m}(y_i, \cdot).$$

The following regularity condition is defined for the likelihood potentials:

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

$$G_t(x_t) \geq \epsilon_u(G) G_t(y_t)$$

We now 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:

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

where  $0 < a_i < b_i < \infty$  and  $\Theta_i \subset \mathbb{R}^{d_i}$  defines the parameter space, and inf and sup are taken over  $\Theta_i$  and  $E_i$ .

The following condition on the Markov kernels is employed when developing the bound on the error propagation of the parametric-approximation particle filter.

$(\mathcal{H})_u$ : The Markov kernels associated with target dynamics are absolutely continuous and can be expressed in the form

$M_i(x_{i-1}, dx_i) = p_{\vartheta_i}(x_i) dx_i$ . The class of densities associated with  $M_i$  is defined as  $\mathcal{M}_i = \{p_{\vartheta_i}(x_i) : \vartheta_i \in \Theta_i \subset \mathbb{R}^{d_i}\}$ . For each  $\mathcal{M}_i$  there exists an approximation class  $\mathcal{H}_i$  and strictly positive numbers  $a_u = \inf_{i \geq 0} a_i$ ,  $b_u = \sup_{i \geq 0} b_i$  satisfying  $0 < a_u < b_u < \infty$  such that for any  $i \geq 0$  we have

$$\mathcal{M}_i \subseteq \mathcal{H}_i \quad \text{and hence} \quad M_i(x_{i-1}, \cdot) \geq \frac{a_u}{b_u} M_i(y_{i-1}, \cdot)$$

### C. Description of Algorithms

Denote by  $\delta_t$  a binary variable indicating whether approximation (subsampling or parametric) is performed at time-step  $t$ . In our analysis, we will assume that this variable is the outcome of a decision function based on the set of particles  $\{\xi_{t-1}^j\}_{j=1}^N$  and observations  $Y_{t-1}$ . We define  $\delta_0 = 0$ , and we assume there exists probability characterizing the frequency of approximation at time-step  $t$ :  $\mathbb{E}\{\delta_t\} = \mathbb{P}\{\delta_t = 1\} = q_t$ . The expectation is with respect to the Monte-Carlo sampling, measurement noise and the possible target trajectories.

#### Subsample approximation particle filter

The subsample-approximation particle filter involves approximating the  $N$ -sample particle filter by a subsample of size  $N_b$  whenever  $\delta_t = 1$ . In the case of the leader-node particle filter, this step is performed whenever there is a change in leader-node. The new leader-node then re-builds an  $N$ -sample particle representation by sampling<sup>1</sup> from the  $N_b$ -sample approximation. The algorithm can be expressed as:

$$\begin{aligned} \eta_t^N &= S^N \circ S^{N_b}(\Phi_t(\eta_{t-1}^N)) \quad \text{if } \delta_t = 1, \\ \eta_t^N &= S^N(\Phi_t(\eta_{t-1}^N)) \quad \text{if } \delta_t = 0 \end{aligned} \quad (5)$$

#### Parametric approximation particle filter

The parametric-approximation particle filter involves generating a parametric approximation to the filtering density if  $\delta_t = 1$ . After transmission of this parametric approximation, an  $N$ -particle representation is generated by sampling from it. Denote by  $\mathbb{W}_{N_p} : \mathcal{P}(E) \rightarrow \mathcal{C}_{N_p}(E)$  an operator that represents a parametric mixture approximation procedure that involves  $N_p$  mixture components. The parametric-approximation particle filter can then be expressed as:

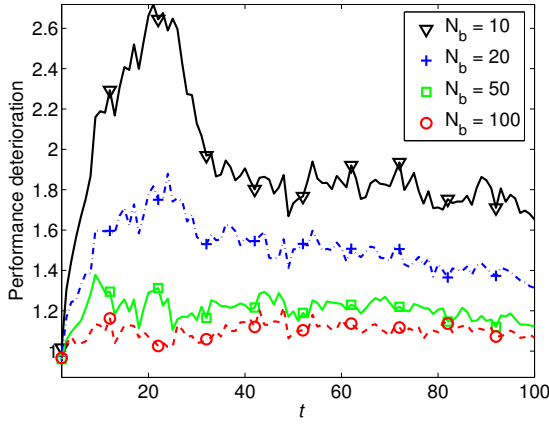
$$\begin{aligned} \eta_t^N &= S^N \circ \mathbb{W}_{N_p}(\Phi_t(\eta_{t-1}^N)) \quad \text{if } \delta_t = 1, \\ \eta_t^N &= S^N(\Phi_t(\eta_{t-1}^N)) \quad \text{if } \delta_t = 0 \end{aligned} \quad (6)$$

The approximation is restricted to a class of discrete  $N_p$ -component convex combinations of the form  $g_i(x) = \sum_{j=1}^{N_p} \alpha_{i,j} \phi_{\theta_{i,j}}(x)$ , where  $\phi_{\theta_{i,j}} \in \mathcal{H}_i$ . To analyze the potential performance of the parametric-approximation particle filter we consider the scenario when  $g_i$  is estimated using the greedy maximum likelihood (GML) approach proposed in [6].

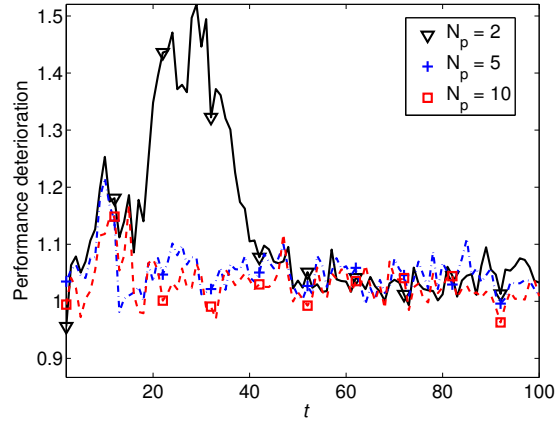
## III. MAIN RESULTS

This section presents our main results — time-uniform bounds on the weak-sense  $L_p$  error for the subsample- and parametric-approximation particle filters. The theorems are

<sup>1</sup>Throughout the paper, we assume that the random sampling without replacement is used.



(a) Subsample leader-node particle filter



(b) Parametric leader-node particle filter

Fig. 1. The deterioration of the dynamic approximation performance for varying number of transmitted particles (averaged over 5000 trials).

stated here without proofs (see details and discussion in [7]). We first present the result for the subsampling particle filter:

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

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

where the constant  $\epsilon_{u,m}$  is given by:

$$\epsilon_{u,m} = m(2 - \epsilon_u(M)\epsilon_u^m(G))/\epsilon_u^3(M)\epsilon_u^{(2m-1)}(G). \quad (7)$$

The following theorem specifies a time-uniform bound on the weak-sense  $L_p$  error for the parametric-approximation particle filter employing the GML algorithm.

**Theorem 2.** *Suppose assumptions  $(G)_u$ , and  $(\mathcal{H})_u$  hold. Suppose further that  $\mathbb{P}\{\delta_i = 1\} \leq q_u$  for any  $i \geq 0$  and  $0 \leq q_u \leq 2/3$ . Then for any  $N_p, N \geq 1$ ,  $t \geq 0$ ,  $p \geq 1$  and  $h_t \in \text{Osc}_1(E_t)$  we have time uniform estimate*

$$\begin{aligned} \sup_{t \geq 0} \mathbb{E} \left\{ \left| [\eta_t^N - \eta_t](h_t) \right|^p \right\}^{1/p} &\leq \epsilon_u \left[ \frac{c^{1/p}(p)}{\sqrt{N}} \right. \\ &+ q_u^{1/p} \left[ \frac{16}{a\sqrt{N}} \left( 2c^{2/p}(p/2) + C(p/4) \mathbb{E}[F(\mathcal{H}, d_N)] \right) \right. \\ &\left. \left. + 8 \log(3\sqrt{\epsilon}(b/a)) \frac{(b/a)^2}{N_p} \right]^{1/2} \right] \end{aligned}$$

where  $C$  is an absolute constant,  $\epsilon_u$  is given by:

$$\epsilon_u = (2 - (a_u/b_u)\epsilon_u(G))/(a_u/b_u)^3 \epsilon_u(G),$$

$$\text{and } F(\mathcal{H}, d_N) = \sup_{t \geq 0} \int_0^{bt} \sqrt{\log(1 + \mathcal{D}(\epsilon, \mathcal{H}_t, d_N))} d\epsilon$$

Here  $d_N$  is the empirical semi-metric defined for  $h_1, h_2 \in \mathcal{H}$

$$d_N^2(h_1, h_2) = \frac{1}{N} \sum_{i=1}^N (h_1(x_i) - h_2(x_i))^2$$

and  $\mathcal{D}(\epsilon, \mathcal{H}, d_N)$  is the packing number — the maximum number of  $\epsilon$ -separated points in  $\mathcal{H}$  under semi-metric  $d_N$ .

#### IV. NUMERICAL EXPERIMENTS

In this section we present the results of numerical experiments. We adopt the following information acquisition and target movement models (see [7] for a detailed description of the simulation scenario and the discussion of results). The state of the target is two-dimensional with dynamics [4]

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

Here  $r_0$  is fixed  $r_0 = 0.02$  and  $\varphi_t, u_t$  are independent and uniformly distributed  $u_t \sim U[0, 1]$ ,  $\varphi_t \sim U[-\pi, \pi]$ .  $K_l = 20$  leader nodes and  $K_s = 200$  satellite nodes are distributed uniformly in the unit square. A satellite 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  $r_c = \sqrt{2 \log(K_s)/K_s}$ . The measurement equation of every satellite sensor is the binary detector 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  $\mathcal{X}_d^j$  is the  $j$ th sensor detection region. To perform sensor selection step we use the mutual information (MI) criterion [2] (see [7] for implementation details). Rapid fluctuations of leader node are avoided by only checking the MI criterion, on average, every 5 time steps (a biased coin toss determines whether checking is performed, resulting in  $q_u \leq 1/5$ ). Subsample generation is performed using the residual resampling scheme [8]. The parametric-approximation particle filter is implemented using the GML algorithm with  $N_p$  components. Each component consists of a 2D Gaussian density with diagonal covariance matrix. The mean vector and covariance matrix are estimated using the particle representation available at the current leader-node (see details in [7]).

Figure 1(a) depicts how the  $L_2$  approximation performance changes with respect to  $N_b$ , the size of the subsample. The y-axis depicts the ratio of the  $L_2$  approximation error of the leader-node particle filter using  $N_b$  subsamples and the  $L_2$  approximation error of the leader-node particle filter performing

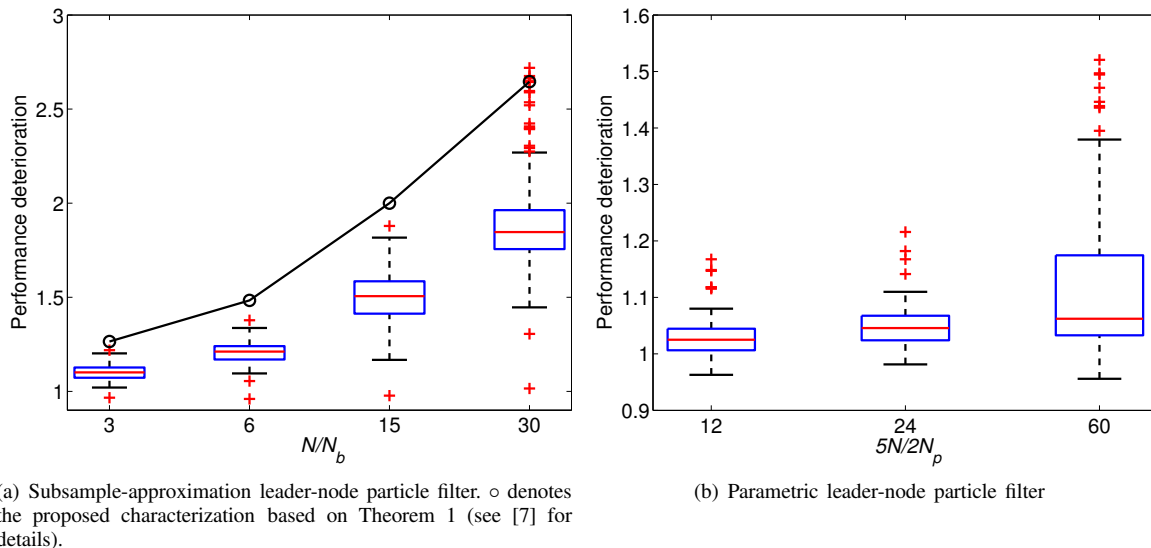


Fig. 2. Box-plots showing the relationship between deterioration of approximation performance and compression factor.

no subsampling ( $N_b = N = 300$ ). Figure 1(b) provides similar results for the parametric-approximation particle filter as  $N_p$  varies. Initially ( $t \in [1, 50]$ ), the particle representation of the target location density is highly dispersed and multi-modal, making it challenging to approximate. As time progresses,  $t \in [51, 100]$ , the particle representation of the target becomes more localized and approximation performance improves. Qualitatively, the performance deteriorates gracefully with respect to reductions in  $N_b$  and  $N_p$ , in line with the theoretical performance bounds.

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, so the factor is  $5N/2N_p$ . Figure 2 presents a box-plot depicting performance deterioration (ratio of the  $L_2$  approximation error of the leader-node with  $N_b < N$  and the leader-node with  $N_b = N$ ) versus the compression factor. Both the median and the maximal deviations of the performance deterioration scale smoothly with changing compression factor. Parametric approximation clearly outperforms subsampling.

For the subsampling case, Theorem 1 provides an analytical bound on the expected approximation error. The curve based on a corollary for this result (see [7], Corollary 3 and Sections IV, VI for details) is depicted in Figure 2(a) and provides a meaningful characterization of the expected performance deterioration. Indeed, the theoretical prediction closely coincides with the maximal performance deterioration observed for each compression factor.

## V. CONCLUSION

We have presented two results bounding the weak-sense  $L_p$  error for particle filters that perform intermittent approximation steps, using either subsampling or parametric

mixture modeling. The important conclusion is that these approximation steps do not induce instability, and moreover, the frequency of the approximation steps significantly affects the extent of performance degradation. If the approximation steps are rare, then the compression can be very high (very few subsamples or very few mixture components) and the error remains reasonable. Numerical experiments indicate that the bound for the subsample-approximation particle filter provides a meaningful characterization of performance.

## ACKNOWLEDGMENT

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

## REFERENCES

- [1] F. Zhao, J. Shin, and J. Reich, "Information-driven dynamic sensor collaboration," *IEEE Signal Process. Magazine*, vol. 19, pp. 61–72, Mar. 2002.
- [2] J. Liu, J. Reich, and F. Zhao, "Collaborative in-network processing for target tracking," *J. Appl. Signal Process.*, vol. 2003, pp. 378–391, 2003.
- [3] J. L. Williams, J. W. Fisher, and A. S. Willsky, "Approximate dynamic programming for communication-constrained sensor network management," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4300–4311, 2007.
- [4] A. Ihler, J. Fisher, and A. Willsky, "Particle filtering under communication constraints," in *Proc. IEEE Workshop Stat. Signal Process.*, Bordeaux, France, May 2005.
- [5] P. Del Moral, *Feynman–Kac formulae. Genealogical and interacting particle approximations*. New York: Springer, 2004.
- [6] J. Q. Li and A. R. Barron, "Mixture density estimation," in *Advances in Neural Information Processing Systems 12*. MIT Press, 1999, pp. 279–285.
- [7] B. N. Oreshkin and M. J. Coates, "Analysis of error propagation in particle filters with approximation," Department of Electrical and Computer Engineering, McGill University, Montreal, QC, Canada, Tech. Rep., Jul. 2009, available at <http://www.tsp.ece.mcgill.ca/Networks/publications-techreport.html>.
- [8] R. Douc and O. Cappe, "Comparison of resampling schemes for particle filtering," in *Proc. 4th Int. Symp. Image and Signal Process. and Analysis*, Sept. 2005, pp. 64–69.