A Survey of Convergence Results on Particle Filtering Methods for Practitioners

Dan Crisan and Arnaud Doucet

***Abstract*—Optimal filtering problems are ubiquitous in signal processing and related fields. Except for a restricted class of models, the optimal filter does not admit a closed-form expression. Particle filtering methods are a set of flexible and powerful sequential Monte Carlo methods designed to solve the optimal filtering problem numerically. The posterior distribution of the state is approximated by a large set of Dirac-delta masses (samples/particles) that evolve randomly in time according to the dynamics of the model and the observations. The particles are interacting; thus, classical limit theorems relying on statistically independent samples do not apply. In this paper, our aim is to present a survey of recent convergence results on this class of methods to make them accessible to practitioners.**

***Index Terms*—Bayesian estimation, optimal filtering, particle filtering, sequential Monte Carlo, state-space models.**

I. Introduction

MANY models in signal processing can be cast in a state-space form. In most applications, prior knowledge of the system is also available. This knowledge allows us to adopt a Bayesian approach, that is, to combine a prior distribution for the unknown quantities with a likelihood function relating these quantities to the observations. Within this setting, one performs inference on the unknown state according to the posterior distribution. Often, the observations arrive sequentially in time, and one is interested in *estimating recursively in time* the evolving posterior distribution. This problem is known as the Bayesian or *optimal filtering* problem. The posterior distribution only admits an analytical expression for few special models, including linear Gaussian state-space models (Kalman filter) and finite state-space hidden Markov models (HMM filters). However, in many realistic problems, state-space models include elements of nonlinearity and non-Gaussianity that preclude a closed-form expression for the optimal filter. For over 30 years, many approximation schemes have been proposed to tackle this problem, such as the extended Kalman filter and approximations using Gaussian sums; see [1] and [17]. Unfortunately, in many cases, these suboptimal methods are unreliable. Deterministic numerical integration methods have also been developed. Although they perform well when the state is low dimensional [20], they are very difficult to implement if the dimension of the state is, say, larger than 4. Moreover, the rate of convergence of the approximation error decreases as the state dimension increases. That is, these methods suffer from the so-called curse of dimensionality. Following the seminal paper by Gordon, Salmond, and Smith introducing the bootstrap filter/sampling importance resampling [19], there has been a surge of interest in particle filtering methods, which are also known as sequential Monte Carlo (SMC) methods. These methods utilize a large number N of random samples (or particles) to represent the posterior probability distributions. The particles are propagated over time using a combination of sequential importance sampling and resampling steps. The resampling step statistically multiplies and/or discards particles at each time step to adaptatively concentrate particles in regions of high posterior probability. These methods are very flexible and can be easily applied to nonlinear and non-Gaussian dynamic models.

In particle filtering methods, the particles (samples) interact and, thus, are statistically dependent. Consequently, classical convergence results on Monte Carlo methods, based on independent and identically distributed (i.i.d.) assumptions, do not apply. Therefore, it is useful to ask the following questions.

- Does the particle filter converge asymptotically (that is as $N \rightarrow \infty$) toward the optimal filter and in what sense?
- Do standard Monte Carlo rates for convergence apply?
- Is there an accumulation of error with time?
- Can we give any large deviation results?

To a certain degree, these questions have recently been answered. Of all the algorithms available, the most extensively studied is the bootstrap filter/SIR, which is also known as the interacting particle systems/resolution algorithm; see [8]–[10] for a recent overview of results. In [5] and [6], a rigorous treatment is given to a whole class of SMC methods. However, most of these results have been published in the probability literature. They can prove difficult for practitioners to read and, indeed, are often unknown to them. The aim of this paper is to present a survey of the results available in the literature to make them understandable and applicable to "real-life" problems.

The rest of the paper is organized as follows. In Section II, we specify the model and the optimal filtering problem. In Section III, a generic particle filtering algorithm is described, and its different steps are briefly detailed. Section IV discusses almost sure (weak) convergence of the empirical distributions toward the true ones. In Section V, we give simple sufficient conditions to ensure asymptotic convergence of mean square error to zero. We then discuss a few conditions to ensure uniform convergence in time. Finally, in Section VI, we present a few large deviation results.

Manuscript received February 24, 2001; revised November 2, 2001. The associate editor coordinating the review of this paper and approving it for publication was Prof. Simon J. Godsill.

D. Crisan is with the Department of Mathematics, Imperial College, London, U.K. (e-mail: d.crisan@ic.ac.uk).

A. Doucet is with the Department of Electrical Engineering, The University of Melbourne, Parkville, Victoria, Australia (e-mail: a.doucet@ee.mu.oz.au).

Publisher Item Identifier S 1053-587X(02)01328-4.

II. Optimal Filtering

A. General State-Space Models

Let (Ω, F, P) be a probability space on which we have defined two real vector-valued stochastic processes $X = \{X_t, t \in \mathbb{N}\}$ and $Y = \{Y_t, t \in \mathbb{N}\backslash\{0\}\}$. The process X is usually called the *signal* process, and the process Y is called the *observation* process. Let n_x and n_y be the dimensions of the state space of X and Y, respectively, and let $\mathcal{B}(\mathbb{R}^n)$ be the Borel σ-algebra on $\mathbb{R}^n$. The *signal* process X is a Markov process of initial distribution $X_0 \sim \mu(dx_0)$ and probability transition kernel $K(dx_t|x_{t-1})$ such that

$$\Pr(X_t \in A|X_{t-1} = x_{t-1}) = \int_A K(dx_t|x_{t-1}),\ A \in \mathcal{B}(\mathbb{R}^{n_x}).$$

The *observations* are conditionally independent of X and have marginal distribution

$$\Pr(Y_t \in B|X_t = x_t) = \int_B g(dy_t|x_t), \qquad B \in \mathcal{B}(\mathbb{R}^{n_y}).$$

For the sake of simplicity, we will assume here that $K(dx_t|x_{t-1})$ and $g(dy_t|x_t)$ admit densities with respect to the Lebesgue measure. This means that $\Pr(X_t \in dx_t|X_{t-1} = x_{t-1}) = K(dx_t|x_{t-1}) = K(x_t|x_{t-1})\,dx_t$, and $\Pr(Y_t \in dy_t|X_t = x_t) = g(dy_t|x_t) = g(y_t|x_t)\,dy_t$.

Example: Let us consider the scalar dynamic model

$$\begin{aligned} X_t &= f(X_{t-1}) + V_t,\\ Y_t &= g(X_t) + W_t \end{aligned}$$

where $\{V_t\}_{t\in\mathbb{N}}$ and $\{W_t\}_{t\in\mathbb{N}\backslash\{0\}}$ are both independent identically distributed (i.i.d.) sequences and are mutually independent with $\Pr(V_t \in C) = \int_C P_V(dv) = \int_C p_V(v)\,dv$ $(C \in \mathcal{B}(\mathbb{R}))$ and $\Pr(W_t \in D) = \int_D P_W(dw) = \int_D p_W(w)\,dw$ $(D \in \mathcal{B}(\mathbb{R}))$. Then, one has

$$\begin{aligned} K(x_t|x_{t-1}) &= p_V(x_t - f(x_{t-1}))\\ g(y_t|x_t) &= p_W(y_t - g(x_t)). \end{aligned}$$

B. Bayes' Recursions

We will denote by $X_{k:l} \triangleq (X_k, X_{k+1}, \ldots, X_l)$ and $Y_{k:l} \triangleq (Y_k, Y_{k+1}, \ldots, Y_l)$ the path of the signal and of the observation process from time k to time l, respectively. In addition, $x_{k:l} \triangleq (x_k, x_{k+1}, \ldots, x_l)$ and $y_{k:l} \triangleq (y_k, y_{k+1}, \ldots, y_l)$ are generic points in the space of paths of the signal and observation processes. Define the probability distribution as

$$\pi_{k:l|m}(dx_{k:l}) \triangleq P(X_{k:l} \in dx_{k:l}|Y_{1:m} = y_{1:m}).$$

Bayes' theorem allows us to propagate over time the joint and marginal filtering distributions $\pi_{0:t|t}(dx_{0:t})$ and $\pi_{t|t}(dx_t)$. That is, at time t, the joint distribution $\pi_{0:t|t}(dx_{0:t})$ satisfies

$$\pi_{0:t|t}(dx_{0:t}) \propto \mu(dx_0) \prod_{k=1}^{t} K(dx_k|x_{k-1})g(y_k|x_k)$$

and the recursion

$$\begin{aligned} &\pi_{0:t|t-1}(dx_{0:t})\\ &\quad= \pi_{0:t-1|t-1}(dx_{0:t-1})K(dx_t|x_{t-1}) \qquad \textit{Prediction}\\ &\pi_{0:t|t}(dx_{0:t})\\ &\quad= \left[\int_{\mathbb{R}^{n_x}} g(y_t|x_t)\pi_{0:t|t-1}(dx_{0:t})\right]^{-1}\\ &\qquad \cdot g(y_t|x_t)\pi_{0:t|t-1}(dx_{0:t}) \qquad \textit{Updating}. \end{aligned}$$

One typically focuses on the marginal distribution $\pi_{t|t}(dx_t)$. In this case, $\pi_{t|t}(dx_t)$ satisfies the recursion

$$\begin{aligned} &\pi_{t|t-1}(dx_t)\\ &\quad= \int_{\mathbb{R}^{n_x}} \pi_{t-1|t-1}(dx_{t-1})K(dx_t|x_{t-1}) \qquad \textit{Prediction}\\ &\pi_{t|t}(dx_t)\\ &\quad= \left[\int_{\mathbb{R}^{n_x}} g(y_t|x_t)\pi_{t|t-1}(dx_t)\right]^{-1}\\ &\qquad \cdot g(y_t|x_t)\pi_{t|t-1}(dx_t) \qquad \textit{Updating}. \end{aligned} \tag{1}$$

If ν is a measure, φ is a function, and Ξ is a Markov transition kernel,[1] we use the standard notation

$$\begin{aligned} (\nu, \varphi) &\triangleq \int \varphi\nu,\\ \nu\Xi(A) &\triangleq \int \nu(dx)\Xi(A|x)\\ \Xi\varphi(x) &\triangleq \int \Xi(dz|x)\varphi(z). \end{aligned}$$

Using this notation, it is easy to see that for any function φ: $\mathbb{R}^{n_x} \to \mathbb{R}$, the recurrence formula (1) implies that

$$\begin{aligned} (\pi_{t|t-1}, \varphi) &= (\pi_{t-1|t-1}, K\varphi) \qquad \textit{Prediction}\\ (\pi_{t|t}, \varphi) &= (\pi_{t|t-1}, g)^{-1}(\pi_{t|t-1}, \varphi g) \qquad \textit{Updating}. \end{aligned} \tag{2}$$

Except for a very restricted number of dynamic models, it is impossible to evaluate equations (1) or (2) in a closed-form expression.

Example: For a dynamic model such that $\pi_{t|t-1}(dx_t)$ and $\pi_{t|t}(dx_t)$ admit some densities denoted $p(x_t|y_{1:t-1})$ and $p(x_t|y_{1:t})$, then (1) reads

$$\begin{aligned} p(x_t|y_{1:t-1}) &= \int_{\mathbb{R}^{n_x}} p(x_{t-1}|y_{1:t-1})K(x_t|x_{t-1})\,dx_{t-1},\\ p(x_t|y_{1:t}) &= \frac{g(y_t|x_t)p(x_t|y_{1:t-1})}{\displaystyle\int_{\mathbb{R}^{n_x}} g(y_t|x_t)p(x_t|y_{1:t-1})\,dx_t} \end{aligned}$$

and (2) reads

$$\begin{aligned} &\int_{\mathbb{R}^{n_x}} p(x_t|y_{1:t-1})\varphi(x_t)\,dx_t\\ &\quad= \int_{\mathbb{R}^{n_x}} p(x_{t-1}|y_{1:t-1})\left[\int_{\mathbb{R}^{n_x}} K(x_t|x_{t-1})\varphi(x_t)\,dx_t\right] dx_{t-1}\\ &\int p(x_t|y_{1:t})\varphi(x_t)\,dx_t\\ &\quad= \frac{\displaystyle\int_{\mathbb{R}^{n_x}} g(y_t|x_t)p(x_t|y_{1:t-1})\varphi(x_t)\,dx_t}{\displaystyle\int_{\mathbb{R}^{n_x}} g(y_t|x_t)p(x_t|y_{1:t-1})\,dx_t}. \end{aligned}$$

[1] $\Xi(\cdot|\cdot)$: $(\mathbb{R}^{n_x}, \mathcal{B}(\mathbb{R}^{n_x})) \to [0, 1]$ is a Markov transition kernel on $\mathbb{R}^{n_x}$ if, for any $x \in \mathbb{R}^{n_x}$, $\Xi(\cdot|x)$ is a probability measure and, for any $A \in \mathcal{B}(\mathbb{R}^{n_x})$, $\Xi(A|\cdot)$ is a measurable function.

III. Particle Filtering

A particle filtering method is a recursive algorithm that produces, at each time t, a cloud of particles whose empirical measure closely "follows" the distribution $\pi_{t|t}$. In the following subsections, we describe a general algorithm that generates, at time t (for all $t > 0$), N particles/paths $\{x_t^{(i)}\}_{i=1}^N$ with an associated empirical measure $\pi_{t|t}^N$

$$\pi_{t|t}^N(dx_t) \triangleq \frac{1}{N}\sum_{i=1}^N \delta_{x_t^{(i)}}(dx_t)$$

that is "close" to $\pi_{t|t}$; $\delta_x(dx_t)$ denotes the delta-Dirac mass located in x. The algorithm is recursive in the sense that $\{x_t^{(i)}\}_{i=1}^N$ is produced using the observation obtained at time t and the set of particles $\{x_{t-1}^{(i)}\}_{i=1}^N$ produced at time $t-1$ (whose empirical measure $\pi_{t-1|t-1}^N$ was "close" to $\pi_{t-1|t-1}$).

A. Basics

We present here a slight modification of the standard bootstrap filter algorithm described in [19]. Given a set of particles $\{x_{t-1}^{(i)}\}_{i=1}^N$ distributed approximately according to $\pi_{t-1|t-1}(dx_{t-1})$, one samples $\tilde{x}_t^{(i)} \sim \pi_{t-1|t-1}^N K(dx_t) = (1/N)\sum_{k=1}^N K(dx_t|x_{t-1}^{(k)})$.[2] The new particles $\{\tilde{x}_t^{(i)}\}_{i=1}^N$ are distributed approximately according to $\pi_{t|t-1}(dx_t)$ [see (1)]. Their empirical distribution

$$\tilde{\pi}_{t|t-1}^N(dx_t) \triangleq \frac{1}{N}\sum_{i=1}^N \delta_{\tilde{x}_t^{(i)}}(dx_t)$$

is an approximation of $\pi_{t|t-1}$. If one plugs this measure into (1), we get the Monte Carlo approximation of $\pi_{t|t-1}(dx_t)$

$$\tilde{\pi}_{t|t}^N(dx_t) \triangleq \frac{g(y_t|x_t)\tilde{\pi}_{t|t-1}^N(dx_t)}{\int_{\mathbb{R}^{n_x}} g(y_t|x_t)\tilde{\pi}_{t|t-1}^N(dx_t)} = \frac{\sum_{i=1}^N g\left(y_t\middle|\tilde{x}_t^{(i)}\right)\delta_{\tilde{x}_t^{(i)}}(dx_t)}{\sum_{i=1}^N g\left(y_t\middle|\tilde{x}_t^{(i)}\right)}$$

that is

$$\tilde{\pi}_{t|t}^N(dx_t) = \sum_{i=1}^N w_t^{(i)}\delta_{\tilde{x}_t^{(i)}}(dx_t), \qquad \sum_{i=1}^N w_t^{(i)} = 1 \tag{3}$$

where $w_t^{(i)} \propto g(y_t|\tilde{x}_t^{(i)})$ are the so-called *importance weights*. The distribution $\tilde{\pi}_{t|t}^N(dx_t)$ is a weighted sum of delta-Dirac masses.

The aim of the resampling/selection step is to obtain an "unweighted" empirical distribution approximation

$$\pi_{t|t}^N(dx_t) = \frac{1}{N}\sum_{i=1}^N \delta_{x_t^{(i)}}(dx_t)$$

by duplicating particles $\tilde{x}_t^{(i)}$ having high weights and discarding the others to focus on the zones of high posterior probabilities. This is typically achieved by resampling N times from the empirical distribution $\tilde{\pi}_{t|t}^N(dx_t)$. The resulting particles are approximately distributed according to $\pi_{t|t}$, and one can thus iterate the procedure to obtain an approximation of $\pi_{t+1|t+1}$.

B. Algorithm

We also assume that we can sample exactly from $\pi_{0|0} \triangleq \mu$ at $t = 0$. The algorithm proceeds as follows.

```
At time $t = 0$.
  Step 0: Initialization
    • For $i = 1, \ldots, N$, sample $x_0^{(i)} \sim \pi_{0|0}(dx_0)$ and set $t = 1$.
At time $t \geq 1$
  Step 1: Importance Sampling step
    • For $i = 1, \ldots, N$, sample $\tilde{x}_t^{(i)} \sim \pi_{t-1|t-1}^N K(dx_t)$.
    • For $i = 1, \ldots, N$, evaluate the normalized importance weights $w_t^{(i)}$
          $w_t^{(i)} \propto g\left(y_t\middle|\tilde{x}_t^{(i)}\right); \quad \sum_{i=1}^N w_t^{(i)} = 1.$    (4)
  Step 2: Resampling step
    • For $i = 1, \ldots, N$, sample $x_t^{(i)} \sim \tilde{\pi}_{t|t}^N(dx_t)$.
```

This particle filter is thus nothing but a simulation-based approximation of the recursion (1). In the sampling step, one obtains a set of particles $\{\tilde{x}_t^{(i)}\}_{i=1}^N$ whose "unweighted" empirical distribution $\tilde{\pi}_{t|t-1}^N(dx_t)$ is a Monte Carlo approximation of $\pi_{t|t-1}(dx_t)$. The weighted empirical distribution $\tilde{\pi}_{t|t}^N(dx_t)$ approximates $\pi_{t|t}(dx_t)$. The resampling step is a (crucial) algorithmic step that produces an unweighted approximation $\pi_{t|t}^N(dx_t)$ of $\tilde{\pi}_{t|t}^N(dx_t)$.

C. Extensions

The algorithm we have described is very intuitive and easy to use. As we will show later, it produces an approximation that converges (in a given sense) toward the "true" optimal filter under minimal assumptions. However, this algorithm suffers from several drawbacks in practice.

1) *Variation of the Importance Weights*: It can be inefficient if the distribution of the particles $\{\tilde{x}_t^{(i)}\}_{i=1}^N$, given approximately by $\pi_{t|t-1} = \pi_{t-1|t-1}K$, is "far" from $\pi_{t|t}$ in the sense that the ratio (i.e., the Radon–Nykodym derivative) of these two distributions

$$\frac{\pi_{t|t}(dx_t)}{\pi_{t|t-1}(dx_t)} \propto g(y_t|x_t)$$

generates importance weights $\{w_t^{(i)}\}_{i=1}^N$ ($w_t^{(i)} \propto g(y_t|\tilde{x}_t^{(i)})$) with a high variance.

2) *Variation in the Resampling Step*: To produce the unweighted measure approximation $\pi_{t|t}^N(dx_t)$ from $\tilde{\pi}_{t|t}^N(dx_t)$, the algorithm proposed above samples N times from $\tilde{\pi}_{t|t}^N(dx_t)$. In effect, it generates $N_t^{(i)}$ copies of the ith particle, where the $N_t^{(i)}$ are distributed according to a multinomial distribution with parameters $(N; w_t^{(1)}, \ldots, w_t^{(N)})$. Consequently, $E(N_t^{(i)}) = Nw_t^{(i)}$

[2] In the bootstrap filter, one samples from $\tilde{x}_t^{(i)} \sim K(dx_t|x_{t-1}^{(i)})$.

and var$(N_t^{(i)}) = Nw_t^{(i)}(1 - w_t^{(i)})$. Although this produces an "unbiased" approximation of $\tilde{\pi}_{t|t}^N(dx_t)$, i.e., for any function φ

$$E\left[\left(\pi_{t|t}^N, \varphi\right)\middle|\left\{\tilde{x}_t^{(i)}\right\}_{i=1}^N\right] = \left(\tilde{\pi}_{t|t}^N, \varphi\right)$$

it also introduces a large Monte Carlo variation.

We now present the sequential importance sampling/resampling algorithm described in [14]. This algorithm addresses both problems.

1) Alternative Sampling Distributions: To address the first problem, one idea is to sample the particles $\{\tilde{x}_t^{(i)}\}_{i=1}^N$ from $\pi_{t-1|t-1}^N\tilde{K}$ instead of $\pi_{t-1|t-1}^N K$, i.e.,

$$\tilde{x}_t^{(i)} \sim \pi_{t-1|t-1}^N \tilde{K}$$

where the new kernel $\tilde{K}$ is chosen such that the distribution of these particles (which approximates $\pi_{t-1|t-1}\tilde{K}$) is "closer" to $\pi_{t|t}$ than is $\pi_{t-1|t-1}K$. Several choices are discussed in [15] and [16]. To account for the effect of the discrepancy between $\pi_{t-1|t-1}\tilde{K}$ and $\pi_{t|t}$, we use the expression

$$\pi_{t-1:t|t}(dx_{t-1:t}) = \frac{w(x_{t-1}, x_t, y_t)\tilde{K}(dx_t|x_{t-1})\pi_{t-1|t-1}(dx_{t-1})}{\displaystyle\int w(x_{t-1}, x_t, y_t)\tilde{K}(dx_t|x_{t-1})\pi_{t-1|t-1}(dx_{t-1})} \tag{5}$$

where

$$w(x_{t-1}, x_t, y_t) \propto \frac{g(y_t|x_t)K(dx_t|x_{t-1})}{\tilde{K}(dx_t|x_{t-1}, y_t)}. \tag{6}$$

Thus, replacing $\tilde{K}(dx_t|x_{t-1})\pi_{t-1|t-1}(dx_{t-1})$ in (5) by its empirical approximation

$$\frac{1}{N}\sum_{i=1}^N \delta_{x_{t-1}^{(i)}, \tilde{x}_t^{(i)}}(dx_{t-1}, dx_t)$$

and, marginalizing over x_{t-1}, we get the expression (3) for $\tilde{\pi}_{t|t}^N(dx_t)$, where $w_t^{(i)} \propto w(x_{t-1}^{(i)}, \tilde{x}_t^{(i)}, y_t)$.

A different interpretation of this algorithm that aids understanding why some particular algorithms perform better than others can be obtained by defining a *new dynamic model* such that $X_0 \sim \mu(dx_0)$ and

$$\begin{aligned}
&\Pr(X_t \in A|X_{t-1} = x_{t-1}, Y_t = y_t)\\
&\quad = \int_A \tilde{K}(dx_t|x_{t-1}, y_t), \qquad A \in \mathcal{B}(\mathbb{R}^{n_x}) \qquad (7)\\
&\Pr(Y_t \in B|X_{t-1} = x_{t-1}, X_t = x_t)\\
&\quad = \int_B w(x_{t-1}, x_t, dy_t), \qquad B \in \mathcal{B}(\mathbb{R}^{n_y}). \qquad (8)
\end{aligned}$$

In (7), $\tilde{K}(dx_t|x_{t-1}, y_t)$ is a Markov transition kernel dependent on y_t, whereas in (8), $w(x_{t-1}, x_t, y_t)$ is given by (6) [where we assume that this ratio is well defined and that $\int w(x_{t-1}, x_t, y_t)dy_t < \infty$]. Let us define

$$\rho_{k:l|m}(dx_{k:l}) \triangleq \Pr(X_{k:l} \in dx_{k:l}|Y_{1:m} = y_{1:m}).$$

At time t, the joint distribution $\rho_{0:t|t}(dx_{0:t})$ satisfies

$$\begin{aligned}
&\rho_{0:t|t}(dx_{0:t})\\
&\quad\propto \mu(dx_0)\prod_{k=1}^t \tilde{K}(dx_k|x_{k-1}, y_k)w(x_{k-1}, x_k, y_k)\\
&\quad\propto \mu(dx_0)\prod_{k=1}^t K(dx_k|x_{k-1})g(y_k|x_k) \text{ [from (6)]}.
\end{aligned}$$

Thus, although $\rho_{0:t|t-1} \neq \pi_{0:t|t-1}$, one has $\rho_{0:t|t} = \pi_{0:t|t}$ for any t, and thus, in particular, one has $\rho_{t|t} = \pi_{t|t}$. We will give a number of proofs for the standard algorithm involving conditions on $K(dx_t|x_{t-1})$ and $g(y_t|x_t)$. The above shows that they are valid for the algorithm presented here if similar conditions are imposed on $\tilde{K}(dx_t|x_{t-1}, y_t)$ and $w(x_{t-1}, x_t, y_t)$.[3]

If $\tilde{K}(dx_t|x_{t-1}, y_t)$ and $w(x_{t-1}, x_t, y_t)$ have "better" theoretical properties than $K(dx_t|x_{t-1})$ and $g(y_t|x_t)$, such as better mixing properties of $\tilde{K}(dx_t|x_{t-1}, y_t)$ or flatter likelihood $w(x_{t-1}, x_t, y_t)$, then the algorithm will perform better. That is, designing efficient particle filtering methods is equivalent to finding an appropriate dynamic model that has good theoretical properties while keeping the same filtering distributions. The resampling step is a generic step that is independent of the dynamic model considered.

Remark 1: Recently, the introduction of Markov chain Monte Carlo steps in particle filtering algorithms has been suggested; see [18]. This fits in this framework as it can simply be interpreted as the introduction of a new evolution equation in the dynamic model; see [3] for details.

2) Resampling Schemes: To address the second problem, we note that the aim of the resampling/selection step is just to obtain an "unweighted" empirical distribution approximation $\pi_{t|t}^N(dx_t)$ of the weighted measure $\tilde{\pi}_{t|t}^N(dx_t)$ by associating a number of copies/offspring $N_t^{(i)} \in \mathbb{N}$ with each particle $\{\tilde{x}_t^{(i)}\}_{i=1}^N$. That is, one wants

$$\begin{aligned}
\pi_{t|t}^N(dx_t) &= N^{-1}\sum_{i=1}^N \delta_{x_t^{(i)}}(dx_t)\\
&= N^{-1}\sum_{i=1}^N N_t^{(i)}\delta_{\tilde{x}_t^{(i)}}(dx_t)\\
&\approx \sum_{i=1}^N w_t^{(i)}\delta_{\tilde{x}_t^{(i)}}(dx_t) = \tilde{\pi}_{t|t}^N(dx_t)
\end{aligned}$$

with $\sum_{i=1}^N N_t^{(i)} = N$. Recently, many schemes have been proposed in the literature to deal with these problems. Most of these

[3] This is true because we assume the observations to be fixed in this paper. For random observations, further integrability assumptions need to be imposed on $\tilde{K}(dx_t|x_{:t-1}, y_t)$ and $w(x_{t-1}, x_t, y_t)$.

algorithms ensure that $E(N_t^{(i)}) = Nw_t^{(i)}$, as for the multinomial sampling procedure, but have a lower variance var($N_t^{(i)}$). Algorithms achieving the minimum variance are presented in [4], [6], and [21]. It also possible to use a deterministic algorithm such at the one described in [21].

IV. Almost Sure Convergence

From now on, we will assume that the observation process is fixed to a given observation record $Y_t = y_t$, $t > 0$. All the convergence results will be given under this condition.

A. *Preliminary Remark*

Before we analyze the convergence of the algorithms presented previously, we make a few preliminary remarks that will enable us to understand very quickly why they converge and what conditions need to be imposed. We start with an abstract formulation, but gradually, we identify its elements with those comprising the filtering problem. Let (E, d) be a metric space, and on this space, let $(a_t)_{t=1}^{\infty}$ and $(b_t)_{t=1}^{\infty}$ be two sequences of continuous functions $a_t, b_t \colon E \to E$. In addition, let k_t and $k_{1:t}$ be two other sequences of functions defined as

$$k_t \triangleq a_t \circ b_t, \qquad k_{1:t} \triangleq k_t \circ k_{t-1} \circ \cdots \circ k_1$$

where the operation "$\circ$" denotes the composition of functions, i.e., $a_t \circ b_t(e) = a_t(b_t(e))$. Obviously, both k_t and $k_{1:t}$ are continuous.

For the stochastic filtering setup, the space E will be $\mathcal{P}(\mathbb{R}^{n_x})$, the space of all probability measures over the n_x-dimensional Euclidean space $\mathbb{R}^{n_x}$, b_t will be the map that takes $\pi_{t-1|t-1}$ into $\pi_{t|t-1}$, and a_t the map that takes $\pi_{t|t-1}$ into $\pi_{t|t}$. Thus, k_t will be the transformation $\pi_{t-1|t-1} \longrightarrow \pi_{t|t}$, and $k_{1:t}$ will be the transformation $\pi_{0|0} \longrightarrow \pi_{t|t}$.

We perturb k_t and $k_{1:t}$ using a (not necessarily continuous) function c^N, c^N $E \to E$ in the following way.

$$k_t^N = c^N \circ a_t \circ c^N \circ b_t, \qquad k_{1:t}^N = k_t^N \circ k_{t-1}^N \circ \cdots \circ k_1^N.$$

In the context of stochastic filtering, c^N will be the map that takes a measure to a random sample of size N of the measure. We next assume that as N increases, the perturbations become increasingly smaller. In other words, we assume that c^N converges to i, which is the identity function on E [$i(\alpha) = \alpha$, for all $\alpha \in E$]. A natural question to ask is whether k_t^N converges to k_t and $k_{1:t}^N$ converges to $k_{1:t}$? It turns out that the answer is "no," as the following example clearly demonstrates.

Example 1: Let $E = [0, 1]$ and d be the usual metric on $[0, 1]$, $d(\alpha, \beta) = |\alpha - \beta|$. Let a_t and b_t be equal to the identity function i on E, $a_t = b_t = i$. Hence, k_t is the identity function as well. We modify k_t as above using the following continuous piecewise linear perturbation c^N.

$$c^N(\alpha) = \begin{cases} \alpha + \dfrac{\alpha}{N}, & \text{if } \alpha \in \left[0, \dfrac{1}{2}\right] \\ 1-(N-1)\left|\dfrac{1}{2} + \dfrac{1}{2N} - \alpha\right|, & \text{if } \alpha \in \left(\dfrac{1}{2}, \dfrac{1}{2} + \dfrac{1}{N}\right) \\ \alpha + \dfrac{\alpha - 1}{N-2}, & \text{if } \alpha \in \left[\dfrac{1}{2} + \dfrac{1}{N}, 1\right]. \end{cases}$$

It is easy to check that although $\lim_{N\to\infty} c^N(\alpha) = \alpha$, for all $\alpha \in [0, 1]$, one has

$$\begin{aligned} \lim_{N\to\infty} k_t^N\left(\tfrac{1}{2}\right) &= \lim_{N\to\infty} c^N\left(c^N\left(\tfrac{1}{2}\right)\right) \\ &= \lim_{N\to\infty} c^N\left(\frac{1}{2} + \frac{1}{2N}\right) \\ &= 1 \neq \tfrac{1}{2} = k_t\left(\tfrac{1}{2}\right). \end{aligned}$$

Hence, it is not true that successive small perturbations of a function still amount to a small perturbation. In order to have $\lim_{N\to\infty} k_{1:t}^N = k_{1:t}$, we need a stronger type of convergence for c^N. We need c^N to converge in a uniform manner[4] to the identity function i. In particular, we need c^N to satisfy

$$\begin{gathered} \text{For all } e_N,\ e \in E \text{ such that} \\ \lim_{N\to\infty} e_N = e \Longrightarrow \lim_{N\to\infty} c^N(e_N) = e. \end{gathered} \tag{9}$$

Condition (9) is equivalent to

$$\begin{gathered} \text{For all } e_N,\ e \in E \text{ such that} \\ \lim_{N\to\infty} e_N = e \Longrightarrow \lim_{N\to\infty} d(c^N(e_N), e_N) = 0. \end{gathered} \tag{10}$$

Hence, we have the following lemma.

Lemma 1: Let a_t, b_t, k_t, $k_{1:t}$, and c^N be defined as above. Then, if c^N satisfies (9), we have

$$\lim_{N\to\infty} k_t^N = k_t \quad \text{and} \quad \lim_{N\to\infty} k_{1:t}^N = k_{1:t}. \tag{11}$$

Moreover, both k_t^N and $k_{1:t}^N$ satisfy

$$\begin{aligned} \lim_{N\to\infty} e_N = e \Longrightarrow \lim_{N\to\infty} k_t^N(e_N) &= k_t(e) \\ \lim_{N\to\infty} k_{1:t}^N(e_N) &= k_{1:t}(e). \end{aligned} \tag{12}$$

Proof: Since (12) implies (11) [take $e_N = e$ for all N in (12)], we only need to prove (12). As b_t is continuous, we have

$$\lim_{N\to\infty} e_N = e \Longrightarrow \lim_{N\to\infty} b_t(e_N) = b_t(e). \tag{13}$$

Then, using (9), we get that

$$\lim_{N\to\infty} b_t(e_N) = b_t(e) \Longrightarrow \lim_{N\to\infty} c^N(b_t(e_N)) = b_t(e) \tag{14}$$

and, since a_t is continuous

$$\begin{aligned} &\lim_{N\to\infty} c^N(b_t(e_N)) = b_t(e) \\ &\Longrightarrow \lim_{N\to\infty} a_t(c^N(b_t(e_N))) = a_t(b_t(e)) \end{aligned} \tag{15}$$

and, again using (9), we get that

$$\begin{aligned} &\lim_{N\to\infty} a_t(c^N(b_t(e_N))) = a_t(b_t(e)) \\ &\Longrightarrow \lim_{N\to\infty} c^N(a_t(c^N(b_t(e_N)))) = a_t(b_t(e)). \end{aligned} \tag{16}$$

Finally, by putting together (13)–(16), we prove that $\lim_{N\to\infty} k_t^N(e_N) = k_t(e)$, which, in turn, implies by induction (over t) that $\lim_{N\to\infty} k_{1:t}^N(e_N) = k_{1:t}(e)$. ■

[4] c^N converges uniformly to the identity function i if, by definition, for all $\varepsilon > 0$ there exists $N(\varepsilon)$ such that $d(c^N(e), i(e)) < \varepsilon$ for all $N \geq N(\varepsilon)$. Uniform convergence is stronger than (9), but we only need (9) for Lemma 1 to be valid.

B. Application to Optimal Filtering

In the following, we will relate the previous proof to the stochastic filtering problem. The convergence of the particle filter algorithm will be shown to be a direct corollary of Lemma 1.

1) Space of Probability Measures Over $\mathbb{R}^{n_x}$*:* Let $E = \mathcal{P}(\mathbb{R}^{n_x})$ be the set of probability measures over the n_x-dimensional Euclidean space $\mathbb{R}^{n_x}$ endowed with the topology of weak convergence. In this topology, if $(\mu_N)_{N=1}^{\infty}$ is a sequence of probability measures, then we say that μ_N converges (weakly) to $\mu \in \mathcal{P}(\mathbb{R}^{n_x})$ and write $\lim_{N\to\infty} \mu_N = \mu$ if, for any $\varphi \in C_b(\mathbb{R}^{n_x})$

$$\lim_{N\to\infty} (\mu_N, \varphi) = (\mu, \varphi)$$

where $C_b(\mathbb{R}^{n_x})$ is the set of all continuous bounded functions on $\mathbb{R}^{n_x}$. One can choose a countable subset $\mathcal{A} = \{\varphi_i\}_{i>0} \in C_b(\mathbb{R}^{n_x})$ of continuous bounded functions that completely determines convergence. In other words

$$\lim_{N\to\infty} \mu_N = \mu \text{ weakly} \Longleftrightarrow \lim_{N\to\infty} (\mu_N, \varphi_i) = (\mu, \varphi_i) \quad \forall\, \varphi_i \in \mathcal{A}.$$

Using this set, we can define the following distance on $\mathcal{P}(\mathbb{R}^{n_x})$, which generates the weak topology.

$$d(\mu, \nu) = \sum_{i=1}^{\infty} \frac{|(\mu, \varphi_i) - (\nu, \varphi_i)|}{2^i \|\varphi_i\|}$$

where $\|\cdot\|$ is the supremum norm on $C_b(\mathbb{R}^{n_x})$, $\|\varphi\| \triangleq \sup_{x\in\mathbb{R}^{n_x}} |\varphi(x)|$. It is easy to prove that

$$\lim_{N\to\infty} \mu_N = \mu \text{ weakly} \Longleftrightarrow \lim_{N\to\infty} d(\mu_N, \mu) = 0.$$

Hence, d generates the weak topology on $\mathcal{P}(\mathbb{R}^{n_x})$. Of course, d depends on the choice of the set $\mathcal{A}$. However, the topology itself is independent of $\mathcal{A}$.

2) Continuous Functions Over $\mathcal{P}(\mathbb{R}^{n_x})$*:* We define b_t: $\mathcal{P}(\mathbb{R}^{n_x}) \to \mathcal{P}(\mathbb{R}^{n_x})$ to be the mapping

$$b_t(\nu)(dx_t) \triangleq \nu K(dx_t) = \int_{\mathbb{R}^{n_x}} K(dx_t|x_{t-1})\nu(dx_{t-1})$$

for arbitrary $\nu \in \mathcal{P}(\mathbb{R}^{n_x})$. Hence, for $\varphi \in C_b(\mathbb{R}^{n_x})$

$$\begin{aligned}(b_t(\nu), \varphi) &= \int_{\mathbb{R}^{n_x}} \int_{\mathbb{R}^{n_x}} \varphi(x_t) K(dx_t|x_{t-1})\nu(dx_{t-1}) \\ &= (\nu, K\varphi). \end{aligned} \tag{17}$$

We have

$$\pi_{t|t-1} = b_t(\pi_{t-1|t-1}). \tag{18}$$

We want to ensure that b_t is continuous. This is quite natural. In the context of filtering, it simply means (heuristically) that the signal moves in a continuous manner and that two realizations of the signal that start from "close" positions will remain "close" at subsequent times. Mathematically, one way to ensure this happening is to assume that the transition kernel of the signal is Feller, i.e., it has the property that for φ a continuous bounded function, $K\varphi$ is also a continuous bounded function

$$\forall\, \varphi \in C_b(\mathbb{R}^{n_x}) \Longrightarrow K\varphi \in C_b(\mathbb{R}^{n_x}). \tag{19}$$

If $\lim_{N\to\infty} \nu_N = \nu$, then, by definition, $\lim_{N\to\infty}(\nu_N, \varphi) = (\nu, \varphi)$, $\forall\, \varphi \in C_b(\mathbb{R}^{n_x})$. Hence, $\lim_{N\to\infty}(\nu_N, K\varphi) = (\nu, K\varphi)$, $\forall\, \varphi \in C_b(\mathbb{R}^{n_x})$ and

$$\begin{aligned}\lim_{N\to\infty} (b_t(\nu_N), \varphi) &= \lim_{N\to\infty} (\nu_N, K\varphi) \\ &= (\nu, K\varphi) = (b_t(\nu), \varphi), \quad \forall\, \varphi \in C_b(\mathbb{R}^{n_x}).\end{aligned}$$

We now define the application a_t. Let a_t: $\mathcal{P}(\mathbb{R}^{n_x}) \to \mathcal{P}(\mathbb{R}^{n_x})$ be a mapping such that for arbitrary $\nu \in \mathcal{P}(\mathbb{R}^{n_x})$, $a_t(\nu)$ is a probability measure defined as

$$(a_t(\nu), \varphi) = (\nu, g)^{-1}(\nu, \varphi g), \quad \text{for any } \varphi \in C_b(\mathbb{R}^{n_x}). \tag{20}$$

Then

$$\pi_{t|t} = a_t(\pi_{t|t-1}) = a_t \circ b_t(\pi_{t-1|t-1}). \tag{21}$$

Again, in the context of filtering, it is natural to assume that a_t is continuous. This means (heuristically) that a slight variation in the (starting) conditional distribution of the signal X_t will not result in a big variation in the conditional distribution of the signal when the new observation y_t is taken into account. Mathematically, one of the ways to ensure that this happens is to assume that $g(y_t|\cdot)$ is a continuous bounded strictly positive function

$$g(y_t|\cdot) \in C_b(\mathbb{R}^{n_x}),\ g(y_t|x_t) > 0, \quad \forall\, x_t \in \mathbb{R}^{n_x}. \tag{22}$$

The positivity assumption is necessary to ensure that (ν, g) is never 0 and thereby allowing division by it in (20). Indeed, if $g(y_t|\cdot)$ satisfies (22), then from (20), we have that $\lim_{N\to\infty} \nu_N = \nu$ implies

$$\lim_{N\to\infty} (a_t(\nu_N), \varphi) = \frac{\lim_{N\to\infty} (\nu_N, \varphi g)}{\lim_{N\to\infty} (\nu_N, g)} = \frac{(\nu, \varphi g)}{(\nu, g)} = (a_t(\nu), \varphi)$$

for all test functions $\varphi \in C_b(\mathbb{R}^{n_x})$. Hence, $\lim_{N\to\infty} a_t(\nu_N) = a_t(\nu)$, and therefore, a_t is continuous. Obviously, if a_t and b_t are continuous, so are k_t and $k_{1:t}$, and

$$\pi_{t|t} = k_t(\pi_{t-1|t-1}) = k_{1:t}(\mu). \tag{23}$$

3) Perturbation: In the context of particle filtering, the perturbation c^N will be a random one. However, with probability 1, it will still have all the properties required by the general setup. Let $c^{N,\,\omega}$, $N > 0$, $\omega \in \Omega$ be the following (random) perturbation. For all $\nu \in \mathcal{P}(\mathbb{R}^{n_x})$, $c^{N,\,\omega}(\nu)$ is equal to

$$c^{N,\,\omega}(\nu) = \frac{1}{N} \sum_{j=1}^{N} \delta_{\{V_j(\omega)\}} \tag{24}$$

where V_j: $\Omega \to \mathbb{R}^d$ are i.i.d. random variables with common distribution ν.

Lemma 2: If $c^{N,\omega}$ is defined as above, then for almost all $\omega \in \Omega$, $c^{N,\omega}$ satisfies (9).

Proof: Let ν_N, $\nu \in \mathcal{P}(\mathbb{R}^{n_x})$ be such that $\lim_{N\to\infty} \nu_N = \nu$. $\forall\,\varphi \in \mathcal{A}$, we have, using the independence of the V_js

$$\begin{aligned} E\left[\left((c^{N,\cdot}(\nu_N), \varphi_i) - (\nu_N, \varphi_i)\right)^4\right] &= \frac{1}{N^4} E\left[\left(\sum_{j=1}^{N} (\varphi_i(V_j) - (\nu_N, \varphi_i))\right)^4\right] \\ &= \frac{1}{N^4} \sum_{j=1}^{N} E\left[(\varphi_i(V_j) - (\nu_N, \varphi_i))^4\right] \\ &= \frac{6}{N^4} \sum_{j_1, j_2=1, j_1\neq j_2}^{N} \\ &\quad \cdot E\left[(\varphi_i(V_j) - (\nu_N, \varphi_i))^2 (\varphi_i(V_j) - (\nu_N, \varphi_i))^2\right] \\ &\le \frac{2^4\|\varphi_i\|^4 N + 2^4\|\varphi_i\|^4 \times 3N(N-1)}{N^4} \\ &\le \frac{48\|\varphi_i\|^4}{N^2}. \end{aligned}$$

It follows that

$$E\left[\sum_{N=1}^{\infty} \left((c^{N,\cdot}(\nu_N), \varphi_i) - (\nu_N, \varphi_i)\right)^4\right] \le 48\|\varphi_i\|^4 \sum_{N=1}^{\infty} \frac{1}{N^2} < \infty$$

and hence

$$\sum_{N=1}^{\infty} ((c^{N,\omega}(\nu_N), \varphi_i) - (\nu_N, \varphi_i))^4 < \infty, \text{ for almost all } \omega \in \Omega$$

which implies

$$\lim_{N\to\infty} |(c^{N,\omega}(\nu_N), \varphi_i) - (\nu_N, \varphi_i)| = 0 \text{ for almost all } \omega \in \Omega.$$

Thus, there exists a subset $\overline{\Omega} \subset \Omega$ of full measure $P(\overline{\Omega}) = 1$ such that

$$\forall\,\omega \in \overline{\Omega},\ \forall\, i \in I,\ \lim_{N\to\infty} |(c^{N,\omega}(\nu_N), \varphi_i) - (\nu_N, \varphi_i)| = 0$$

which implies, for all $\omega \in \overline{\Omega}$, that $\lim_{N\to\infty} d(c^{N,\omega}(\nu_N), \nu_N) = 0$, and hence, $c^{N,\omega}$ satisfies (9) for all $\omega \in \overline{\Omega}$. In the following section, we will ignore the dependence on ω. However, all the results stated should be regarded as being true with probability 1, i.e., for almost all $\omega \in \Omega$. ■

4) Particle Filter: Let us now consider $\pi^N_{t|t}$, which is the empirical measure associated with the set of particles obtained at the end of the resampling step in the bootstrap filter described in Section III. It is easy to see that after the resampling step [11] (μ is the initial distribution of the signal)

$$\begin{aligned} \pi^N_{t|t} &= c^N \circ a_t \circ c^N \circ b_t(\pi^N_{t-1|t-1}) = k^N_t(\pi^N_{t-1|t-1}) \\ \pi^N_{t|t} &= k^N_{1:t} \circ c^N(\mu) = k^N_{1:t}(\mu^N) \end{aligned}$$

where $\mu^N = c^N(\mu)$. In addition, observe that $\tilde{\pi}^N_{t|t-1} = c^N \circ b_t(\pi^N_{t-1|t-1})$.

Theorem 1: Assuming that the transition kernel K is Feller and that the likelihood function g is bounded, continuous, and strictly positive, then $\lim_{N\to\infty} \pi^N_{t|t} = \pi_{t|t}$ almost surely.

Proof: This result follows from Lemma 1 and (23) since $\lim_{N\to\infty} \mu^N = \mu$; then

$$\lim_{N\to\infty} \pi^N_{t|t} = \lim_{N\to\infty} k^N_{1:t}(\mu^N) = k_{1:t}(\mu) = \pi_{t|t}.$$

■

Let us consider the case where resampling is achieved by an algorithm different other than multinomial sampling. Then, the algorithm has the form

$$k^N_t = \overline{c}^N \circ a_t \circ c^N \circ b_t$$

where $\overline{c}^N$ is the perturbation introduced by the resampling step, for example, stratified sampling [21] or minimum variance sampling [6]. In this case, we need to apply the same condition (9) to $\overline{c}^N$ as to c^N.

A way to ensure that the resampling procedure satisfies the required condition is to check that it satisfies

$$E\left[\left((\overline{c}^{N,\cdot}(\nu), \varphi) - (\nu, \varphi)\right)^4\right] \le \frac{C}{N^2}$$

for all arbitrary bounded functions φ. If this is not possible, one can ask for

$$E\left[\left((\overline{c}^{N,\cdot}(\nu), \varphi) - (\nu, \varphi)\right)^2\right] \le \frac{C}{N}$$

but in this case, one has to take a subsequence of $\pi^N_{t|t}$, $\pi^{N_k}_{t|t}$ so that

$$E\left[\left((\overline{c}^{N_k,\cdot}(\nu), \varphi) - (\nu, \varphi)\right)^2\right] \le C \sum_k \frac{1}{N_k} < \infty.$$

The approximations $\pi^N_{t|t}$ introduced in [5] are not necessarily probability measures. However, the same analysis applies, only now, one takes as the underlying space E, which is the set $\mathcal{M}(\mathbb{R}^{n_x})$ of finite measures over $\mathbb{R}^{n_x}$, and one defines a distance d similar to that defined above.

Reference [6] makes the point that the conditions on $\overline{c}^N$ and c^N are, in some sense, not only sufficient but also necessarily. It is proved that the following two assertions are equivalent.

1) For all $t > 0$, $\lim_{N\to\infty} \pi^N_{t|t} = \pi_{t|t}$ and $\lim_{N\to\infty} \tilde{\pi}^N_{t|t} = \tilde{\pi}_{t|t}$.
2) For all $t > 0$, $\lim_{N\to\infty} d(c^N(e^N_t), e^N_t) = 0$ [where $e^N_t = b_t(\pi^N_{t|t})$] and $d(\overline{c}^N(\overline{e}^N_t), \overline{e}^N_t) = 0$ [where $\overline{e}^N_t = a_t \circ c^N \circ b_t(\pi^N_{t|t})$].

The sampling perturbation c^N can be replaced to include the case of the bootstrap filter [19].

V. Convergence of the Mean Square Error

We have given conditions to ensure weak convergence of the empirical distributions toward their true values. Now, let us assume now that we still keep $E = \mathcal{P}(\mathbb{R}^{n_x})$, but instead of using weak convergence, we use the following: If $(\mu^\omega_N)^\infty_{N=1}$ is a sequence of (random) probability measures, then we say that μ^ω_N converges to $\mu \in \mathcal{P}(\mathbb{R}^{n_x})$ if, for any $\varphi \in B(\mathbb{R}^{n_x})$ (the set of Borel bounded measurable functions on $\mathbb{R}^{n_x}$)

$$\lim_{N\to\infty} E\left[((\mu_N, \varphi) - (\mu, \varphi))^2\right] = 0$$

where the expectation is over all the realizations of the random particle method. We not only show that this result holds for $\pi_{t|t}^N$, but we also show that the rate of convergence toward zero of this quantity is proportional to $1/N$. It is independent of the state dimension n_x.

A. Simple Convergence

In this subsection, we give a short proof for convergence of the particle filters described in Section III.

1) Bootstrap Filter: We make the following assumption.

Assumption: $g(y_t|\cdot)$ is a bounded function in argument $x_t \in \mathbb{R}^{n_x}$, i.e., $\|g\| < \infty$. The following lemmas essentially state that at each step of the particle filtering algorithm, the approximation admits a mean square error of order $1/N$.

Lemma 3: Let us assume that for any $\varphi \in B(\mathbb{R}^{n_x})$

$$E\left[\left(\left(\pi_{t-1|t-1}^N, \varphi\right) - (\pi_{t-1|t-1}, \varphi)\right)^2\right] \le c_{t-1|t-1}\frac{\|\varphi\|^2}{N}$$

then, after **Step 1** of the algorithm, for any $\varphi_t \in B((\mathbb{R}^{n_x}))$

$$E\left[\left(\left(\pi_{t|t-1}^N, \varphi\right) - (\pi_{t|t-1}, \varphi)\right)^2\right] \le c_{t|t-1}\frac{\|\varphi\|^2}{N}.$$

Proof: One has

$$\begin{aligned}
&\left|\left(\pi_{t|t-1}^N, \varphi\right) - (\pi_{t|t-1}, \varphi)\right| \\
&\quad \le \left|\left(\pi_{t|t-1}^N, \varphi\right) - \left(\pi_{t-1|t-1}^N, K\varphi\right)\right| \\
&\qquad + \left|\left(\pi_{t-1|t-1}^N, K\varphi\right) - (\pi_{t-1|t-1}, K\varphi)\right|.
\end{aligned}$$

Let $\mathcal{G}_{t-1}$ be the σ-field generated by $\{x_{t-1}^{(i)}\}_{i=1}^N$; then

$$E\left[\left(\pi_{t|t-1}^N, \varphi\right)\middle|\mathcal{G}_{t-1}\right] = \left(\pi_{t-1|t-1}^N, K\varphi\right)$$

and, as $\|K\varphi\| \le \|\varphi\|$,

$$\begin{aligned}
&E\left[\left(\left(\pi_{t|t-1}^N, \varphi\right) - E\left[\left(\pi_{t|t-1}^N, \varphi\right)\middle|\mathcal{G}_{t-1}\right]\right)^2\middle|\mathcal{G}_{t-1}\right] \\
&\quad = E\left[\left(\left(\pi_{t|t-1}^N, \varphi\right) - \left(\pi_{t-1|t-1}^N, K\varphi\right)\right)^2\middle|\mathcal{G}_{t-1}\right] \\
&\quad = \frac{1}{N}\left(\left(\pi_{t-1|t-1}^N, K\varphi^2\right) - \left(\pi_{t-1|t-1}^N, K\varphi\right)^2\right) \\
&\quad \le \frac{\|\varphi\|^2}{N}.
\end{aligned}$$

Thus, using Minkowski's inequality, one obtains

$$\begin{aligned}
&E\left[\left(\left(\pi_{t|t-1}^N, \varphi\right) - (\pi_{t|t-1}, \varphi)\right)^2\right]^{1/2} \\
&\quad \le E\left[\left(\left(\pi_{t|t-1}^N, \varphi\right) - \left(\pi_{t-1|t-1}^N, K\varphi\right)\right)^2\right]^{1/2} \\
&\qquad + E\left[\left(\left(\pi_{t-1|t-1}^N, K\varphi\right) - (\pi_{t-1|t-1}, K\varphi)\right)^2\right]^{1/2} \\
&\quad \le \sqrt{c_{t|t-1}}\frac{\|\varphi\|}{\sqrt{N}}
\end{aligned}$$

where $c_{t|t-1} = (1 + \sqrt{c_{t-1|t-1}})^2$. ■

Lemma 4: Let us assume that for any $\varphi \in B(\mathbb{R}^{n_x})$

$$E\left[\left(\left(\pi_{t|t-1}^N, \varphi\right) - (\pi_{t|t-1}, \varphi)\right)^2\right] \le c_{t|t-1}\frac{\|\varphi\|^2}{N}.$$

Then, for any $\varphi \in B(\mathbb{R}^{n_x})$

$$E\left[\left(\left(\tilde{\pi}_{t|t}^N, \varphi\right) - (\pi_{t|t}, \varphi)\right)^2\right] \le \tilde{c}_{t|t}\frac{\|\varphi\|^2}{N}.$$

Proof: One has

$$\begin{aligned}
&\left(\tilde{\pi}_{t|t}^N, \varphi\right) - (\pi_{t|t}, \varphi) \\
&\quad = \frac{\left(\pi_{t|t-1}^N, g\varphi\right)}{\left(\pi_{t|t-1}^N, g\right)} - \frac{(\pi_{t|t-1}, g\varphi)}{(\pi_{t|t-1}, g)} \\
&\quad = \frac{\left(\pi_{t|t-1}^N, g\varphi\right)}{\left(\pi_{t|t-1}^N, g\right)} - \frac{\left(\pi_{t|t-1}^N, g\varphi\right)}{(\pi_{t|t-1}, g)} + \frac{\left(\pi_{t|t-1}^N, g\varphi\right)}{(\pi_{t|t-1}, g)} \\
&\qquad - \frac{(\pi_{t|t-1}, g\varphi)}{(\pi_{t|t-1}, g)}
\end{aligned}$$

where

$$\begin{aligned}
&\left|\frac{\left(\pi_{t|t-1}^N, g\varphi\right)}{\left(\pi_{t|t-1}^N, g\right)} - \frac{\left(\pi_{t|t-1}^N, g\varphi\right)}{(\pi_{t|t-1}, g)}\right| \\
&\quad = \frac{\left(\pi_{t|t-1}^N, g\varphi\right)\left|(\pi_{t|t-1}, g) - \left(\pi_{t|t-1}^N, g\right)\right|}{\left(\pi_{t|t-1}^N, g\right)(\pi_{t|t-1}, g)} \\
&\quad \le \frac{\|\varphi\|}{(\pi_{t|t-1}, g)}\left|(\pi_{t|t-1}, g) - \left(\pi_{t|t-1}^N, g\right)\right|.
\end{aligned}$$

Thus, one obtains, using Minkowski's inequality again

$$\begin{aligned}
&E\left[\left(\left(\tilde{\pi}_{t|t}^N, \varphi\right) - (\pi_{t|t}, \varphi)\right)^2\right]^{1/2} \\
&\quad \le E\left[\left(\frac{\left(\pi_{t|t-1}^N, g\varphi\right)}{\left(\pi_{t|t-1}^N, g\right)} - \frac{\left(\pi_{t|t-1}^N, g\varphi\right)}{(\pi_{t|t-1}, g)}\right)^2\right]^{1/2} \\
&\qquad + E\left[\left(\frac{\left(\pi_{t|t-1}^N, g\varphi\right)}{(\pi_{t|t-1}, g)} - \frac{(\pi_{t|t-1}, g\varphi)}{(\pi_{t|t-1}, g)}\right)^2\right]^{1/2} \\
&\quad \le \frac{\|\varphi\|}{(\pi_{t|t-1}, g)}E\left[\left((\pi_{t|t-1}, g) - \left(\pi_{t|t-1}^N, g\right)\right)^2\right]^{1/2} \\
&\qquad + \frac{E\left[\left(\left(\pi_{t|t-1}^N, g\varphi\right) - (\pi_{t|t-1}, g\varphi)\right)^2\right]^{1/2}}{(\pi_{t|t-1}, g)} \\
&\quad \le \frac{2\sqrt{c_{t|t-1}}\|g\|}{(\pi_{t|t-1}, g)}\frac{\|\varphi\|}{\sqrt{N}}
\end{aligned}$$

as $\|g\| < \infty$ by assumption. ■

Lemma 5: Let us assume that for any $\varphi \in B(\mathbb{R}^{n_x})$

$$E\left[\left(\left(\tilde{\pi}_{t|t}^N, \varphi\right) - (\pi_{t|t}, \varphi)\right)^2\right] \leq \tilde{c}_{t|t} \frac{\|\varphi\|^2}{N}.$$

Then, after **Step 2** of the algorithm, there exists a constant $c_{t|t}$ such that for any $\varphi \in B(\mathbb{R}^{n_x})$

$$E\left[\left(\left(\pi_{t|t}^N, \varphi\right) - (\pi_{t|t}, \varphi)\right)^2\right] \leq c_{t|t} \frac{\|\varphi\|^2}{N}.$$

Proof: One has

$$\begin{aligned}\left(\pi_{t|t}^N, \varphi\right) - (\pi_{t|t}, \varphi) \\ = \left(\pi_{t|t}^N, \varphi\right) - \left(\tilde{\pi}_{t|t}^N, \varphi\right) + \left(\tilde{\pi}_{t|t}^N, \varphi\right) - (\pi_{t|t}, \varphi).\end{aligned}$$

Then, Minkowski's inequality gives

$$\begin{aligned}E&\left[\left(\left(\pi_{t|t}^N, \varphi\right) - (\pi_{t|t}, \varphi)\right)^2\right]^{1/2} \\ &\leq E\left[\left(\left(\pi_{t|t}^N, \varphi\right) - \left(\tilde{\pi}_{t|t}^N, \varphi\right)\right)^2\right]^{1/2} \\ &\quad + E\left[\left(\left(\tilde{\pi}_{t|t}^N, \varphi\right) - (\pi_{t|t}, \varphi)\right)^2\right]^{1/2}.\end{aligned}$$

Let $\mathcal{F}_t$ be the σ-field generated by $\{\tilde{x}_t^{(i)}\}_{i=1}^N$. It is easy to see that the multinomial procedure is such that

$$E\left[\left(\pi_{t|t}^N, \varphi\right) \middle| \mathcal{F}_t\right] = \left(\tilde{\pi}_{t|t}^N, \varphi\right)$$

and

$$E\left[\left(\left(\pi_{t|t}^N, \varphi\right) - \left(\tilde{\pi}_{t|t}^N, \varphi\right)\right)^2 \middle| \mathcal{F}_t\right] \leq \frac{C}{N} \|\varphi\|^2$$

giving

$$E\left[\left(\left(\pi_{t|t}^N, \varphi\right) - (\pi_{t|t}, \varphi)\right)^2\right]^{1/2} \leq \frac{\sqrt{C} + \sqrt{\tilde{c}_{t|t}}}{\sqrt{N}} \|\varphi\|.$$

By putting together Lemmas 3–5, we obtain the following theorem.

Theorem 2: Under assumption 1, for all $t \geq 0$, there exists $c_{t|t}$ independent of N such that for any $\varphi \in B(\mathbb{R}^{n_x})$

$$E\left[\left(\left(\pi_{t|t}^N, \varphi\right) - (\pi_{t|t}, \varphi)\right)^2\right] \leq c_{t|t} \frac{\|\varphi\|^2}{N}. \tag{25}$$

In other words, particle filtering methods beat the *curse of dimensionality* as the rate of convergence is independent of the state dimension n_x. However, to ensure a given precision on the mean square error given by (25), the number of particles N also depends from $c_{t|t}$, which can depend on n_x. Note that the result has only been established for bounded functions. This excludes $\varphi(x) = x$ and, thus, the standard minimum mean square estimate (MMSE) of the state $E[X_t | Y_{1:t} = y_{1:t}]$.

2) Extensions: Assume the general sequential importance sampling/resampling algorithm described previously. It is clear that if one uses a kernel $\tilde{K} \neq K$, the assumption $\|g\| < \infty$ must be replaced by $\|w\| < \infty$. Similarly, if one uses a resampling scheme other than the multinomial resampling, one needs to ensure that $\{N_t^{(i)}\}_{i=1}^N$ are integer-valued random variables such that

$$E\left[\left|\sum_{i=1}^N \left(N_t^{(i)} - N\overline{w}_t^{(i)}\right) q^{(i)}\right|^2\right] \leq C_t N \max_{i=1,\dots N} \left|q^{(i)}\right|^2$$

for all N-dimensional vectors $q = (q^{(1)}, q^{(2)}, \ldots, q^{(N)}) \in \mathbb{R}^N$ and $\sum_{i=1}^N N_t^{(i)} = N$. This assumption is satisfied by the resampling schemes described in [4], [6], and [21].

To sum up, as long as the importance weights are upper bounded and one uses a standard resampling scheme, then convergence of the mean square error toward zero is ensured, and the rate of convergence is in $1/N$.

3) Uniform Convergence: Theorem 2 ensures that under minimal conditions, $(\pi_{t|t}^N, \varphi)$ converges toward $(\pi_{t|t}, \varphi)$ in the mean square sense for any $\varphi \in B(\mathbb{R}^{n_x})$ and that the rate of convergence of the approximation error $E[((\pi_{t|t}^N, \varphi) - (\pi_{t|t}, \varphi))^2]$ is in $1/N$. However, we have not paid attention to the growth of the sequence $c_{t|t}$. Indeed, there is no reason why $c_{t|t}$ should not increase over time. Actually, without any additional assumption, it does. Assuming that the "true" optimal filter associated with the dynamic model one simulates does not forget its initial condition, then the (approximation) errors committed at any time t accumulate over time. As a consequence, $c_{t|t}$ increases over time. This implies that to ensure a given precision of the estimate $(\pi_{t|t}^N, \varphi)$, one needs an increasingly larger number of particles as time t increases. This is not really satisfactory in applications where one faces a large number of data.

To ensure that $c_{t|t}$ does not increase over time, one needs to have some mixing assumptions on the dynamic model (and thus the "true" optimal filter) that ensure that any error is forgotten (exponentially) with time. Several results have been established recently in the literature after the pioneering work in [11]. A general overview of this problem and new results are presented in [22]. We present here a result established in [22].

Let us consider the kernel

$$R_t(dx_t | x_{t-1}) \triangleq g(y_t | x_t) K(dx_t | x_{t-1}).$$

Assumption (Mixing Kernel): There exist ε and a positive measure λ such that

$$\varepsilon \lambda(dx_t) \leq R_t(dx_t | x_{t-1}) \leq \varepsilon^{-1} \lambda(dx_t)$$

for any $x_{t-1} \in \mathbb{R}^{n_x}$.

This assumption means that the kernel is very weakly dependent on the past value x_{t-1}. This is a strong assumption. It can typically only be established when X_t lies in a compact subset of $\mathbb{R}^{n_x}$. However, it might be possible to relax this strong assumption.

Assumption: One has

$$\rho_t = \frac{\sup_{x \in \mathbb{R}^{n_x}} g(y_t | x_t)}{\inf_{\mu \in \mathcal{P}(\mathbb{R}^{n_x})} (\mu K, g(y_t | \cdot))} < \rho < \infty.$$

Then, under these two assumptions, the following uniform (in time) convergence result holds.

Theorem 3 [22]: For all $t \geq 0$, there exists a constant $c(\varepsilon)$ independent of N such that for any $\varphi \in B(\mathbb{R}^{n_x})$

$$E\left[\left(\left(\pi_{t|t}^N, \varphi\right) - (\pi_{t|t}, \varphi)\right)^2\right] \leq c(\varepsilon)\,\frac{\|\varphi\|^2}{N}.$$

This result roughly means that if the "true" optimal filter is quickly mixing, then uniform convergence in time of the particle filtering method is ensured. On the contrary, it is expected that if the optimal filter has a "long memory," then there will be an accumulation of errors over time that prevents uniform convergence.

Remark 2: In the case where a (random) fixed parameter is part of the state, the dynamic model is not ergodic, and it is thus expected that whatever the particle filtering one uses, one cannot obtain uniform convergence results. In practice, it has been observed that as time increases, such algorithms indeed diverge [2].

VI. Large Deviations

We state here a result concerning the large deviations analysis of two types of particle filters (for details, see [7] and [12]). We start with the definition of a large deviation principle (LDP) [13, p. 35].

Definition 1: Let X be a separable metric space equipped with the Borel σ-field $\mathcal{B}(X)$, and let $\{\mu_N\}_{N\in\mathbb{N}}$ be a sequence of probability measures on $\mathcal{B}(X)$. We say that the sequence $\{\mu_N\}_{N\in\mathbb{N}}$ satisfies a full LDP with the *rate function* I: $X \to [0, \infty]$ if the following conditions hold.

- The rate function I is lower semi-continuous, that is, for every sequence $x_N \to x \in X$, we have $\liminf_{N\to\infty} I(x_N) \geq I(x)$ or, equivalently, that $\{I \leq \alpha\} \subset X$ is a closed set for every $\alpha \geq 0$.
- For every open set $U \subset X$, we have the lower bound
$$-\inf_{x\in U} I(x) \leq \liminf_{N\to\infty} \frac{1}{N}\ln \mu_N(U).$$
- For every closed set $F \subset X$, we have the upper bound
$$\limsup_{N\to\infty} \frac{1}{N}\ln \mu_N(F) \leq -\inf_{x\in F} I(x).$$

  If, in addition, $\{I \leq \alpha\} \subset X$ is a compact set for every $\alpha \geq 0$, we say that the LDP holds with the *good rate function* I.

The probability measures μ_N are usually the laws of a sequence of random variables x_N that converge to a certain value x. Heuristically, the rate function I tells us how quickly x_N converges to x. The higher the value of I is on a certain set A (for which $x \notin A$), the quicker the sequence leaves that set

$$P(x_N \in A) \simeq e^{-N\inf_{y\in A} I(y)}.$$

Using the notation introduced in the previous sections, let $\pi_{t-1:t|t-1}^N$, q_t^N be the measures

$$\pi_{t-1:t|t-1}^N \triangleq \frac{1}{N}\sum_{i=1}^N \delta_{(x_{t-1}^{(i)},\, \tilde{x}_t^{(i)})};$$
$$q_t^N \triangleq \left(\pi_{0|0}^N, \pi_{0:1|0}^N, \pi_{1:2|1}^N, \ldots, \pi_{t-1:t|t-1}^N, \pi_{t-1|t-1}^N\right).$$

Then, q_t^N converges almost surely to $q_t \triangleq (\pi_{0|0}, \pi_{0:1|0}, \pi_{1:2|1}, \ldots, \pi_{t-1:t|t-1}, \pi_{t-1|t-1})$, where

$$\pi_{s-1:s|s-1}(dx_{s-1}, d\tilde{x}_s) \triangleq \pi_{s-1|s-1}(dx_{s-1})K(d\tilde{x}_s|x_{s-1}).$$

Obviously, the first marginal of $\pi_{s-1:s|s-1}$ is $\pi_{s-1|s-1}$, and the second marginal of $\pi_{s-1:s|s-1}$ is $\pi_{s|s-1}$

$$\pi_{s-1:s|s-1}|_1 = \pi_{s-1|s-1}, \quad \pi_{s-1:s|s-1}|_2 = \pi_{s|s-1}.$$

Let μ and ν be two probability measures. We define $H(\mu|\nu)$ to be the relative entropy of μ with respect to ν

$$H(\mu|\nu) \triangleq \int \log\frac{d\mu}{d\nu}\,d\mu$$

if μ is absolutely continuous with respect to ν, $\mu \ll \nu$; otherwise, $H(\mu|\nu) = \infty$. We also define $I_{bootstrap}^s(\rho;\, \nu) \triangleq H(\nu|\tilde{\rho}_s)$, where $\tilde{\rho}_s$ is the measure given by $\tilde{\rho}_s(dx_s) \triangleq \overline{g}_s(x_s)\rho(dx_s)$ and $\overline{g}_s^\rho(x_s) = (g(y_s|x_s)/(\rho,\, g(y_s|\cdot)))$ and $I_{\min}^s(\rho;\, \nu)$ to be the function

$$I_{\min}^s(\rho,\, \nu) \triangleq \begin{cases} \displaystyle\int \delta^s(\nu)\,d\rho, & \text{if } \nu \ll \rho \text{ and } \left[\dfrac{d\nu}{d\rho}\right] = [\overline{g}_s^\rho(x_s)]\,\rho\text{-a.s.} \\ \infty, & \text{otherwise} \end{cases}$$

where

$$\delta^s(\nu) \triangleq \left(1 - \left\{\frac{d\nu}{d\rho}\right\}\right)\ln\frac{1-\left\{\dfrac{d\nu}{d\rho}\right\}}{1-\{\overline{g}_s^\rho(x_s)\}} + \left\{\frac{d\nu}{d\rho}\right\}\ln\frac{\left\{\dfrac{d\nu}{d\rho}\right\}}{\{\overline{g}_s^\rho(x_s)\}}.$$

Then, we have the following theorem.

Theorem 4: If q_t^N is obtained using the bootstrap filter, then the law of q_t^N satisfies a full LDP with the good rate function

$$I_t(\nu_0,\, \mu_1,\, \ldots,\, \mu_t,\, \rho_t) = H(\nu_0|\pi_{0|0}) + \sum_{s=0}^t I^s(\mu_s|_2;\mu_{s+1}|_1) + \sum_{s=1}^t H(\mu_s|\hat{\mu}_s) \qquad (26)$$

for all $\nu_0, \rho_t \in \mathcal{P}(\mathbb{R}^{n_x})$, $\mu_1, \ldots, \mu_t \in \mathcal{P}(\mathbb{R}^{n_x} \times \mathbb{R}^{n_x})$. In (26), we took $\mu_0|_2 \triangleq \pi_{0|0}$, $\mu_{t+1}|_1 \triangleq \rho_t$ and $\hat{\mu}_s(dx_{s-1},\, d\tilde{x}_s) \triangleq \mu_s|_1(dx_{s-1})K(d\tilde{x}_s|x_{s-1})$ and $I^s(\rho,\, \nu) = I_{bootstrap}^s(\rho,\, \nu)$. Moreover, if q_t^N is obtained using an algorithm with a minimal variance resampling scheme, then the law of q_t^N satisfies a full LDP with the good rate function $I_t(\nu_0,\, \mu_1,\, \ldots,\, \mu_t,\, \rho_t)$, where the function I^s in (26) is given by $I^s(\rho,\, \nu) = I_{\min}^s(\rho,\, \nu)$.

As corollaries to the above theorem, one can obtain large deviation results for more convenient path spaces. Since we have

$$I_{\min}^s(\rho,\, \nu) - I_{bootstrap}^s(\rho,\, \nu) \geq \int\left(\frac{d\nu}{d\rho} - \overline{g}_s^\rho\right)^2 d\rho \qquad (27)$$

the minimal variance resampling scheme converges faster than the bootstrap filter on the set of probability measures.

However, in the above theorem, we refer to a variant of the minimal variance resampling scheme described in [7, Sec. 3.3.1] for which $\pi_{t|t}^N$ has a random number of particles (though very close to N). Hence, $\pi_{t|t}^N$ is no longer a probability measure. As a result, $I^s_{\min}(\rho, \nu)$ is finite on some measures for which the mass is not necessarily 1, whereas $I^s_{bootstrap}(\rho, \nu)$ is ∞ on nonprobability measures.

VII. Discussion

In this survey, we have reviewed a few convergence results on particle filtering methods. This is by no way an exhaustive list of results; see, for example, [10] for further detailed results and their proofs. Under weak assumptions, we have shown that it is possible to ensure (almost sure) convergence of the empirical distributions generated by particle filtering methods toward the true ones, some bounds on the mean square errors, and some large deviations results. However, there are still many results to establish. In particular, from a practitioner viewpoint, it seems unsatisfactory to have to assume $\varphi(x)$ bounded above to obtain some convergence results. Similarly, the crucial uniform convergence results rely on strong assumptions on the dynamic models that make them unapplicable for most real-world problems. Nevertheless, this is a very new field, and it is likely that in the near future, stronger results will be established.

References

[1] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[2] C. Andrieu, N. De Freitas, and A. Doucet, "Sequential MCMC for Bayesian model selection," in *Proc. IEEE Higher Order Statist.*, Israel, 1999.

[3] C. Andrieu, A. Doucet, and E. Punskaya, "Sequential Monte Carlo for optimal filtering," in *Sequential Monte Carlo Methods in Practice*, A. Doucet, J. F. G. de Freitas, and N. J. Gordon, Eds. New York: Springer-Verlag, 2001, pp. 79–95.

[4] J. Carpenter, P. Clifford, and P. Fearnhead, "Building robust simulation-based filters for evolving data sets," Tech. Rep., Dept. Statist., Univ. Oxford, Oxford, U.K., 1999.

[5] D. Crisan, P. Del Moral, and T. Lyons, "Discrete filtering using branching and interacting particle systems," *Markov Proc. Rel. Fields*, vol. 5, pp. 293–318, 1999.

[6] D. Crisan, "Particle filters—A theoretical perspective," in *Sequential Monte Carlo Methods in Practice*, A. Doucet, J. F. G. de Freitas, and N. J. Gordon, Eds. New York: Springer-Verlag, 2001.

[7] D. Crisan and M. Grunwald, "Large deviation comparison of branching algorithms versus resampling algorithms: application to discrete time stochastic filtering," Statist. Lab., Cambridge Univ., Cambridge, U.K., Tech. Rep., TR1999-9, 1999.

[8] P. Del Moral, "Non linear filtering: Interacting particle solution," *Markov Proc. Rel. Fields*, vol. 2, pp. 555–580, 1996.

[9] ——, "Measure valued processes and interacting particle systems. Application to non linear filtering problems," *Ann. Appl. Probabil.*, vol. 8, pp. 438–495, 1997.

[10] P. Del Moral and L. Miclo, "Branching and interacting particle systems approximations of Feynman–Kac formulae with applications to nonlinear filtering," in *Séminaire de Probabilités XXXIV*, J. Azéma, M. Emery, M. Ledoux, and M. Yor, Eds. Berlin, Germany: Springer-Verlag, 2000, vol. 1729, pp. 1–145.

[11] P. Del Moral and A. Guionnet, "On the stability of interacting processes with applications to filtering and genetic algorithms," *Ann. Instit. Henri Poincaré*, 2001, to be published.

[12] ——, "Large deviations for interacting particle systems. applications to non linear filtering problems," *Stochast. Process. Applicat.*, vol. 78, pp. 69–95, 1998.

[13] J. Deuschel and D. W. Stroock, *Large Deviations*. Boston, MA: Academic, 1989.

[14] A. Doucet, J. F. G. de Freitas, and N. J. Gordon, *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag, 2001.

[15] A. Doucet, "On sequential simulation-based methods for Bayesian filtering," Tech. Rep., Cambridge Univ., CUED/F-INFENG TR310, Cambridge, U.K., 1998.

[16] A. Doucet, S. J. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statist. Comput.*, vol. 10, pp. 197–208, 2000.

[17] A. Gelb, *Applied Optimal Estimation*. Cambridge, MA: MIT Press, 1974.

[18] W. R. Gilks and C. Berzuini, "Following a moving target—Monte Carlo inference for dynamic Bayesian models," *J. R. Statist. Soc. B*, vol. 63, pp. 127–146, 2001.

[19] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *Proc. Inst. Elect. Eng. F*, vol. 140, pp. 107–113, 1993.

[20] G. Kitagawa, "Non-Gaussian state-space modeling of nonstationary time series," *J. Amer. Statist. Assoc.*, vol. 82, pp. 1032–1063, 1987.

[21] ——, "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models," *J. Comput. Graph. Statist.*, vol. 5, pp. 1–25, 1996.

[22] F. Legland and N. Oudjane, "Stability and uniform approximation of nonlinear filters using the Hilbert metric, and application to particle filters," INRIA, Paris, France, Res. Rep. RR-4215, 2001.

[23] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York: Springer-Verlag, 1999.

[24] M. West and P. J. Harrison, *Bayesian Forecasting and Dynamic Models*, 2nd ed. New York: Springer-Verlag, 1997.

Dan Crisan received the Ph.D. degree in mathematics from the University of Edinburgh, Edinburgh, U.K., in 1996.

From 1995 to 1998, he worked as a Research Associate at Imperial College, London, U.K. In 1998, he joined the Statistical Laboratory, University of Cambridge, Cambridge, U.K., for a two-year assistant lectureship. He returned to Imperial College in 2000 to assume a lectureship position in the Department of Mathematics. His research interests include the mathematics of sequential Monte Carlo methods, stochastics PDEs, and measure-valued branching processes.

Arnaud Doucet was born in Melle, France, on November 2, 1970. He received the M.S. degree from the Institut National des Telecommunications, Paris, France, in 1993 and the Ph.D. degree from University Paris XI, Orsay, France, in 1997.

During 1998, he was a visiting scholar with the Signal Processing Group of Cambridge University, Cambridge, U.K. From 1999 to 2001, he was a research associate in the same group. Since March 2001, he has been senior lecturer with the Department of Electrical Engineering, Melbourne University, Parkville, Victoria, Australia. His research interests include Markov chain Monte Carlo methods, sequential Monte Carlo methods, Bayesian statistics, and Hidden Markov models. He has co-edited, with J. F. G. de Freitas and N. J. Gordon, *Sequential Monte Carlo Methods in Practice* (New York: Springer-Verlag, 2001).