

# A Basic Convergence Result for Particle Filtering

Xiao-Li Hu, Thomas B. Schön and Lennart Ljung

**Abstract**—The basic nonlinear filtering problem for dynamical systems is considered. Approximating the optimal filter estimate by particle filter methods has become perhaps the most common and useful method in recent years. Many variants of particle filters have been suggested, and there is an extensive literature on the theoretical aspects of the quality of the approximation. Still a clear cut result that the approximate solution, for *unbounded* functions, converges to the true optimal estimate as the number of particles tends to infinity seems to be lacking. It is the purpose of this contribution to give such a basic convergence result for a rather general class of unbounded functions. Furthermore, a general framework, including many of the particle filter algorithms as special cases, is given.

## I. INTRODUCTION

The nonlinear filtering problem is formulated as follows. The objective is to recursively in time estimate the state in the dynamic model,

$$x_{t+1} = f_t(x_t, v_t), \quad (1a)$$

$$y_t = h_t(x_t, e_t), \quad (1b)$$

where  $x_t \in \mathbb{R}^{n_x}$  denotes the state,  $y_t \in \mathbb{R}^{n_y}$  denotes the measurement,  $v_t$  and  $e_t$  denote the stochastic process and measurement noise, respectively. Furthermore, the dynamic equations for the system are denoted by  $f_t : \mathbb{R}^{n_x} \times \mathbb{R}^{n_v} \rightarrow \mathbb{R}^{n_x}$  and the equations modelling the sensors are denoted by  $h_t : \mathbb{R}^{n_x} \times \mathbb{R}^{n_e} \rightarrow \mathbb{R}^{n_y}$ . Most applied signal processing problems can be written in the following special case of (1),

$$x_{t+1} = f_t(x_t) + v_t, \quad (2a)$$

$$y_t = h_t(x_t) + e_t, \quad (2b)$$

with  $v_t$  and  $e_t$  i.i.d. and mutually independent. Note that any deterministic input signal  $u_t$  is subsumed in the time-varying dynamics. The most commonly used estimate is an approximation of the conditional expectation,

$$\mathbb{E}(\phi(x_t)|y_{1:t}), \quad (3)$$

where  $y_{1:t} \triangleq (y_1, \dots, y_t)$  and  $\phi : \mathbb{R}^{n_x} \rightarrow \mathbb{R}$  is the function of the state that we want to estimate. We are interested in estimating a function of the state, such as  $\phi(x_t)$  from observed output data  $y_{1:t}$ . An especially common case is of course when we seek an estimate of the state itself  $\phi(x_t) = x_t[i], i = 1, \dots, n_x$ , where  $x_t = (x_t[1], \dots, x_t[n_x])^T$ . In order to compute (3) we need the filtering probability density function  $p(x_t|y_{1:t})$ . It is well known that this density function can be expressed using multidimensional integrals [1]. The

problem is that these integrals only permits analytical solutions in a few special cases. The most common special case is of course when the model (2) is linear and Gaussian and the solution is then given by the Kalman filter [2]. However, for the more interesting nonlinear/non-Gaussian case we are forced to approximations of some kind. Over the years there has been a large amount of ideas suggested on how to perform these approximations. The most popular being the extended Kalman filter (EKF) [3], [4]. Other popular ideas include the Gaussian-sum approximations [5], the point-mass filters [6], [7], the unscented Kalman filter (UKF) [8] and the class of multiple model estimators [9]. See, e.g., [10] for a brief overview of the various approximations. In the current work we will discuss a rather recent and popular family of methods, commonly referred to as *particle filters* (PF) or sequential Monte Carlo methods.

The *key idea* underlying the particle filter is to approximate the filtering density function using a number of particles  $\{x_t^i\}_{i=1}^N$  according to

$$\hat{p}_N(x_t|y_{1:t}) = \sum_{i=1}^N w_t^i \delta_{x_t^i}(x_t), \quad (4)$$

where each particle  $x_t^i$  has a weight  $w_t^i$  associated to it, and  $\delta_x(\cdot)$  denotes the delta-Dirac mass located in  $x$ . Due to the delta-Dirac form in (4), a finite sum is obtained when this approximation is passed through an integral and hence, multidimensional integrals are reduced to finite sums. All the details of the particle filter were first assembled by Gordon *et al.* in 1993 in their seminal paper [11]. However, the main ideas, save for the crucial resampling step, have been around since the 1940's [12].

Whenever an approximation is used it is very important to address the issue of its convergence to the true solution and more specifically, under what conditions this convergence is valid. An extensive treatment of the currently existing convergence results can be found in the book [13] and the excellent survey papers [14], [15]. They consider stability, uniform convergence (see also [16], [17]), central limit theorems (see also [18]) and large deviations (see also [19], [20]). The previous results prove convergence of probability measures and only treat bounded functions  $\phi$ , effectively excluding the most commonly used state estimate, the mean value. To the best of our knowledge there are no results available for unbounded functions  $\phi$ . The main contribution of this paper is that we prove convergence of the particle filter for a rather general class of unbounded functions, applicable in many practical situations. This contribution will also describe a general framework for particle filtering algorithms.

It is worth stressing the *key mechanisms* that enables us to study unbounded functions in the particle filtering context.

X-L. Hu is with the Department of Mathematics, College of Science, China Jiliang University 310018, Hangzhou China, e-mail: xlhu@amss.ac.cn

T. B. Schön and L. Ljung are with the Division of Automatic Control, Department of Electrical Engineering, Linköpings universitet, SE-581 83 Linköping, Sweden, e-mail: {schon, ljung}@isy.liu.se, Phone: +46 13 281373, Fax: +46 13 282622

- 1) The most important idea, enabling the contribution in the present paper, is that we consider the relation between the function  $\phi$  and the density functions for noises. This implies that the class of functions  $\phi$  will depend on the involved noise densities.
- 2) We have also introduced a slight algorithm modification, required to complete the proof. It is worth mentioning that this modification is motivated from the mathematics in the proof. However, it is a useful and reasonable modification of the algorithm in its own right. Indeed, it has previously been used to obtain a more efficient algorithm [21].

In Section II we provide a formal problem formulation and introduce the notation we need for the results to follow. A brief introduction to particle filters is given in Section III. In an attempt to make the results as available as possible the particle filter is discussed both in an application oriented fashion and in a more general setting. The algorithm modification is discussed and illustrated in Section IV. Section V provides a general account of convergence results and in Section VI we state the main result and discuss the conditions that are required for the result to hold. The result is then proved in Section VII. Finally, the conclusions are given in Section VIII.

## II. PROBLEM FORMULATION

The problem under consideration in this work is the following. For a fixed time  $t$ , under what conditions and for which functions  $\phi$  does the approximation offered by the particle filter converge to the true estimate,

$$\mathbb{E}(\phi(x_t)|y_{1:t}). \quad (5)$$

In order to give the results in the most simple form possible we are only concerned with  $L^4$ -convergence in this paper. The more general case of  $L^p$ -convergence for  $p > 1$  is also under consideration, using a Rosenthal-type inequality [22].

### A. Dynamic Systems

We will now represent model (1) in a slightly different framework, more suitable for a theoretical treatment. Let  $(\Omega, \mathcal{F}, P)$  be a probability space on which two real vector-valued stochastic processes  $X = \{X_t, t = 0, 1, 2, \dots\}$  and  $Y = \{Y_t, t = 1, 2, \dots\}$  are defined. The  $n_x$ -dimensional process  $X$  describes the evolution of the hidden state of a dynamic system, and the  $n_y$ -dimensional process  $Y$  denotes the available observation process of the same system.

The state process  $X$  is a Markov process with initial state  $X_0$  obeying an initial distribution  $\pi_0(dx_0)$ . The dynamics, describing the state evolution over time, is modelled by a Markov transition kernel  $K(dx_{t+1}|x_t)$  such that

$$P(X_{t+1} \in A | X_t = x_t) = \int_A K(dx_{t+1}|x_t), \quad (6)$$

for all  $A \in \mathcal{B}(\mathbb{R}^{n_x})$ , where  $\mathcal{B}(\mathbb{R}^{n_x})$  denotes the Borel  $\sigma$ -algebra on  $\mathbb{R}^{n_x}$ . Given the states  $X$ , the observations  $Y$  are conditionally independent and have the following marginal distribution,

$$P(Y_t \in B | X_t = x_t) = \int_B \rho(dy_t|x_t), \quad \forall B \in \mathcal{B}(\mathbb{R}^{n_y}). \quad (7)$$

For convenience we assume that  $K(dx_{t+1}|x_t)$  and  $\rho(dy_t|x_t)$  have densities with respect to a Lebesgue measure, allowing us to write

$$\begin{aligned} P(X_{t+1} \in dx_{t+1} | X_t = x_t) &= K(dx_{t+1}|x_t) \\ &= K(x_{t+1}|x_t)dx_{t+1}, \end{aligned} \quad (8a)$$

$$P(Y_t \in dy_t | X_t = x_t) = \rho(dy_t|x_t) = \rho(y_t|x_t)dy_t. \quad (8b)$$

In the following example it is explained how a model in the form (2) relates to the more general framework introduced above.

*Example 2.1:* Let the model be given by (2), where the probability density functions of  $v_t$  and  $e_t$  are denoted by  $p_{v_t}(\cdot)$  and  $p_{e_t}(\cdot)$ , respectively. Then we have the following relations,

$$K(x_{t+1}|x_t) = p_{v_t}(x_{t+1} - f_t(x_t)), \quad (9a)$$

$$\rho(y_t|x_t) = p_{e_t}(y_t - h(x_t)). \quad (9b)$$

### B. Conceptual Solution

In practice, we are most interested in the marginal distribution  $\pi_{t|t}(dx_t)$ , since the main objective is usually to estimate  $\mathbb{E}(x_t|y_{1:t})$  and the corresponding conditional covariance. This section is devoted to describing the generally intractable form of  $\pi_{t|t}(dx_t)$ . By the total probability formula and Bayes' formula, we have the following recursive form for the evolution of the marginal distribution,

$$\pi_{t|t-1}(dx_t) = \int_{\mathbb{R}^{n_x}} \pi_{t-1|t-1}(dx_{t-1})K(dx_t|x_{t-1}) \quad (10a)$$

$$\triangleq b_t(\pi_{t-1|t-1}),$$

$$\pi_{t|t}(dx_t) = \frac{\rho(y_t|x_t)\pi_{t|t-1}(dx_t)}{\int_{\mathbb{R}^{n_x}} \rho(y_t|x_t)\pi_{t|t-1}(dx_t)} \triangleq a_t(\pi_{t|t-1}), \quad (10b)$$

where we have defined  $a_t$  and  $b_t$  as transformations between probability measures on  $\mathbb{R}^{n_x}$ .

Let us now introduce some additional notation, commonly used in this context. Given a measure  $\nu$ , a function  $\phi$ , and a Markov transition kernel  $K$ , denote

$$(\nu, \phi) \triangleq \int \phi(x)\nu(dx), \quad K\phi(x) = \int K(dz|x)\phi(z). \quad (11)$$

Hence,  $\mathbb{E}(\phi(x_t)|y_{1:t}) = (\pi_{t|t}, \phi)$ . Using this notation, by (10), for any function  $\phi: \mathbb{R}^{n_x} \rightarrow \mathbb{R}$ , we have the following recursive form for the optimal filter  $\mathbb{E}(\phi(x_t)|y_{1:t})$ ,

$$(\pi_{t|t-1}, \phi) = (\pi_{t-1|t-1}, K\phi), \quad (12a)$$

$$(\pi_{t|t}, \phi) = \frac{(\pi_{t|t-1}, \phi\rho)}{(\pi_{t|t-1}, \rho)}. \quad (12b)$$

Here it is worth noticing that we have to require that  $(\pi_{t|t-1}, \rho) > 0$ , otherwise the optimal filter (12) will not exist. Furthermore, note that

$$\begin{aligned} \mathbb{E}(\phi(x_t)|y_{1:t}) &= (\pi_{t|t}, \phi) \\ &= \frac{\int \dots \int \pi_0(x_0)K_{1\rho_1} \dots K_{t\rho_t}\phi(x_t)dx_{0:t}}{\int \dots \int \pi_0(x_0)K_{1\rho_1} \dots K_{t\rho_t}dx_{0:t}}, \end{aligned} \quad (13)$$

where  $K_s \triangleq K(x_s|x_{s-1})$ ,  $\rho_s \triangleq \rho(y_s|x_s)$ ,  $s = 1, \dots, t$ ,  $dx_{0:t} \triangleq \{dx_0 \cdots dx_t\}$ , and the integral areas have all been omitted, for the sake of brevity. In general it is, as previously mentioned, impossible to obtain an explicit solution for the optimal filter  $E(\phi(x_t)|y_{1:t})$  by (13). This implies that we have to resort to numerical methods, such as particle filters, to approximate the optimal filter.

### III. PARTICLE FILTERS

We start this section with a rather intuitive and application oriented introduction to the particle filter and then we move on to a general description, more suitable for the theoretical treatment that follows.

#### A. Introduction

Roughly speaking, particle filtering algorithms are numerical methods used to approximate the conditional filtering distribution  $\pi_{t|t}(dx_t)$  using an empirical distribution, consisting of a cloud of particles at each time  $t$ . The main reason for using particles to represent the distributions is that this allows us to approximate the integral operators by finite sums. Hence, the difficulty inherent in (10) has successfully been removed. The basic particle filter, as it was introduced by [11] is given in Algorithm 1 and it is briefly described below. For a more complete introduction, see e.g., [11], [23], [10], [21] where the latter contains a straightforward MATLAB implementation of the particle filter. There are also several books available on the particle filter [24], [25], [26], [13].

#### Algorithm 1: Particle filter

- 1) Initialize the particles,  $\{x_0^i\}_{i=1}^N \sim \pi_0(dx_0)$ .
- 2) Predict the particles by drawing independent samples according to

$$\tilde{x}_t^i \sim K(dx_t|x_{t-1}^i), \quad i = 1, \dots, N.$$

- 3) Compute the importance weights  $\{w_t^i\}_{i=1}^N$ ,

$$w_t^i = \rho(y_t|\tilde{x}_t^i), \quad i = 1, \dots, N,$$

and normalize  $\tilde{w}_t^i = w_t^i / \sum_{j=1}^N w_t^j$ .

- 4) Draw  $N$  new particles, with replacement (resampling), for each  $i = 1, \dots, N$ ,

$$P(x_t^i = \tilde{x}_t^j) = \tilde{w}_t^j \quad j = 1, \dots, N.$$

- 5) Set  $t := t + 1$  and repeat from step 2.

The particle filter is initialized at time  $t = 0$  by drawing a set of  $N$  particles  $\{x_0^i\}_{i=1}^N$  that are independently generated according to the initial distribution  $\pi_0(dx_0)$ . At time  $t - 1$  the estimate of the filtering distribution  $\pi_{t-1|t-1}(dx_{t-1})$  is given by the following empirical distribution

$$\pi_{t-1|t-1}^N(dx_{t-1}) \triangleq \frac{1}{N} \sum_{i=1}^N \delta_{x_{t-1}^i}(dx_{t-1}), \quad (14)$$

In step 2, the particles from time  $t - 1$  are predicted to time  $t$  using the dynamic equations in the Markov transition kernel

$K$ . When step 2 has been performed we have computed the empirical one-step ahead prediction distribution,

$$\tilde{\pi}_{t|t-1}^N(dx_t) \triangleq \frac{1}{N} \sum_{i=1}^N \delta_{\tilde{x}_t^i}(dx_t), \quad (15)$$

which constitutes an estimate of  $\pi_{t|t-1}(dx_t)$ . In step 3 the information in the present measurement  $y_t$  is used. This step can be understood simply by substituting (15) into (10b), resulting in the following approximation of  $\pi_{t|t}(dx_t)$

$$\begin{aligned} \tilde{\pi}_{t|t}^N(dx_t) &\triangleq \frac{\rho(y_t|x_t) \tilde{\pi}_{t|t-1}^N(dx_t)}{\int_{R^{n_x}} \rho(y_t|x_t) \tilde{\pi}_{t|t-1}^N(dx_t)} \\ &= \frac{\sum_{i=1}^N \rho(y_t|\tilde{x}_t^i) \delta_{\tilde{x}_t^i}(dx_t)}{\sum_{i=1}^N \rho(y_t|\tilde{x}_t^i)}. \end{aligned} \quad (16)$$

In practice (16) is usually written using the so called normalized importance weights  $\tilde{w}_t^i$ , defined as

$$\tilde{\pi}_{t|t}^N(dx_t) = \sum_{i=1}^N \tilde{w}_t^i \delta_{\tilde{x}_t^i}(dx_t), \quad \tilde{w}_t^i \triangleq \frac{\rho(y_t|\tilde{x}_t^i)}{\sum_{i=1}^N \rho(y_t|\tilde{x}_t^i)}. \quad (17)$$

Intuitively, these weights contain information about how probable the corresponding particles are. Finally, the important resampling step is performed. Here, a new set of equally weighted particles is generated using the information in the normalized importance weights. This will reduce the problem of having a high dependence on a few particles with large weights. With sample  $x_t^i$  obeying  $\tilde{\pi}_{t|t}^N(dx_t)$  the resample step will provide an equally weighted empirical distribution

$$\pi_{t|t}^N(dx_t) = \frac{1}{N} \sum_{i=1}^N \delta_{x_t^i}(dx_t) \quad (18)$$

to approximate  $\pi_{t|t}(dx_t)$ . This completes one pass of the particle filter as it is given in Algorithm 1.

#### B. Extended Setting

We will now introduce an extended algorithm, which is used in the theoretical analysis that follows. The extension is that the prediction step (step 2 in Algorithm 1) is replaced with the following

$$\tilde{x}_t^i \sim \sum_{j=1}^N \alpha_j^i K(dx_t|x_{t-1}^j), \quad (19)$$

where a new set of weights  $\alpha^i$  have been introduced. Note that this case occurs for instance if samples are drawn from a Gaussian-sum approximation as in [27] and when the particle filter is derived using point-wise approximations as in [28].

The weights  $\alpha^i$  are defined according to

$$\alpha^i = (\alpha_1^i, \alpha_2^i, \dots, \alpha_N^i), \quad (20)$$

where

$$\alpha_j^i \geq 0, \quad \sum_{j=1}^N \alpha_j^i = 1, \quad \sum_{i=1}^N \alpha_j^i = 1. \quad (21)$$

Clearly,

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \alpha_j^i K(dx_t | x_{t-1}^j) &= \frac{1}{N} \sum_{j=1}^N \left( \sum_{i=1}^N \alpha_j^i K(dx_t | x_{t-1}^j) \right) \\ &= \frac{1}{N} \sum_{j=1}^N K(dx_t | x_{t-1}^j) = (\pi_{t-1|t-1}^N, K). \end{aligned} \quad (22)$$

Note that if  $\alpha_j^i = 1$  for  $j = i$ , and  $\alpha_j^i = 0$  for  $j \neq i$ , the sampling method introduced in (19) is reduced to the one employed in Algorithm 1. Furthermore, when  $\alpha_j^i = 1/N$  for all  $i$  and  $j$ , (19) turns out to be a convenient form for theoretical treatment. This is exploited by nearly all existing references dealing with theoretical analysis of the particle filter, see for example [14], [16], [15]. An extended particle filtering algorithm is given in Algorithm 2 below.

*Algorithm 2: Extended particle filter*

- 1) Initialize the particles,  $\{x_0^i\}_{i=1}^N \sim \pi_0(dx_0)$ .
- 2) Predict the particles by drawing independent samples according to

$$\tilde{x}_t^i \sim \sum_{j=1}^N \alpha_j^i K(dx_t | x_{t-1}^j), \quad i = 1, \dots, N.$$

- 3) Compute the importance weights  $\{w_t^i\}_{i=1}^N$ ,

$$w_t^i = \rho(y_t | \tilde{x}_t^i), \quad i = 1, \dots, N,$$

and normalize  $\tilde{w}_t^i = w_t^i / \sum_{j=1}^N w_t^j$ .

- 4) Resample,  $x_t^i \sim \tilde{\pi}_{t|t}^N(dx_t)$ ,  $i = 1, \dots, N$ . ( $\tilde{\pi}$  defined in (16).)  $\pi_{t|t}^N(dx_t) = \frac{1}{N} \sum_{i=1}^N \delta_{x_t^i}(dx_t)$ .

In Fig. 1 we provide a schematic illustration of the particle filter given in Algorithm 2. Let us now discuss the trans-

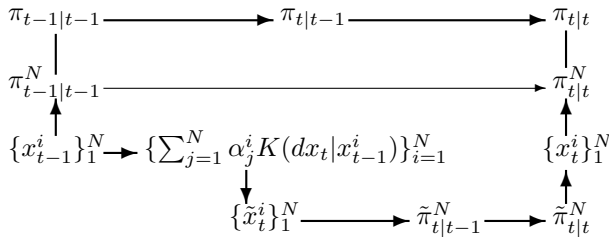


Fig. 1: Illustration of how the particle filter transforms the probability measures. The theoretical transformation (10) is given at the top. The bottom describes what happens during one pass in the particle filter.

formations of the involved probability measures a bit further,

they are

$$\begin{aligned} \pi_{t-1|t-1}^N &\xrightarrow{\text{projection}} \begin{bmatrix} \delta_{x_{t-1}^1} \\ \dots \\ \delta_{x_{t-1}^N} \end{bmatrix} \xrightarrow{b_t} \begin{bmatrix} K(dx_t | x_{t-1}^1) \\ \dots \\ K(dx_t | x_{t-1}^N) \end{bmatrix} \\ &\xrightarrow{\Lambda} \begin{bmatrix} \sum_{j=1}^N \alpha_j^i K(dx_t | x_{t-1}^j) \\ \dots \\ \sum_{j=1}^N \alpha_j^i K(dx_t | x_{t-1}^j) \end{bmatrix}, \end{aligned}$$

where  $\Lambda$  denotes the  $N \times N$  weight matrix  $(\alpha_j^i)_{i,j}$ . Let us, for simplicity, denote the entire transformation above by  $\Lambda b_t$ . Furthermore, we will use  $c^n(\nu)$  to denote the empirical distribution of a sample of size  $n$  from a probability distribution  $\nu$ . Then, we have

$$\tilde{\pi}_{t|t-1}^N = c(N) \circ \Lambda b_t(\pi_{t-1|t-1}^N), \quad (23)$$

where  $c(N) \triangleq \frac{1}{N} [c^1 \dots c^1]$  (Note that  $c^1$  refers to a single sample.) and  $\circ$  denotes composition of transformations in the form of a vector multiplication. Hence, we have

$$\pi_{t|t}^N = c^N \circ a_t \circ c(N) \circ \Lambda b_t(\pi_{t-1|t-1}^N), \quad (24)$$

where  $\circ$  denotes composition of transformations. Therefore,

$$\pi_{t|t}^N = c^N \circ a_t \circ c(N) \circ \Lambda b_t \circ \dots \circ c^N \circ a_1 \circ c(N) \circ \Lambda b_1 \circ c^N(\pi_0).$$

While, in the existing theoretical versions of particle filter algorithm in [14], [16], [15], [13], as stated in [14], the transformation between time  $t-1$  and  $t$  is in a somewhat simpler form,

$$\begin{aligned} \pi_{t|t}^N &= c^N \circ a_t \circ c^N \circ b_t(\pi_{t-1|t-1}^N) \\ &= c^N \circ a_t \circ c^N \circ b_t \circ \dots \circ c^N \circ a_1 \circ c^N \circ b_1 \circ c^N(\pi_0). \end{aligned} \quad (25)$$

The theoretical results and analysis in [29] are based on the following transformation (in our notation):

$$\pi_{t|t}^N = a_t \circ b_t \circ c^N(\pi_{t-1|t-1}^N), \quad (26)$$

rather than (25).

#### IV. A MODIFIED PARTICLE FILTER

The particle filter algorithm has to be modified in order to perform the convergence results which follows in the subsequent sections. This modification is described in Section IV-A and its implications are illustrated in Section IV-B.

##### A. Algorithm Modification

From the optimal filter recursion (12b) it is clear that we have to require that

$$(\pi_{t|t-1}, \rho) > 0, \quad (27)$$

in order for the optimal filter to exist. In the approximation to (12b) we have used (15) to approximate  $\pi_{t|t-1}(dx_t)$ , implying that the following is used in the particle filter algorithm

$$\begin{aligned} (\pi_{t|t-1}, \rho) &\approx (\tilde{\pi}_{t|t-1}^N, \rho) = \int \rho(y_t | x_t) \frac{1}{N} \sum_{i=1}^N \delta_{\tilde{x}_t^i}(dx_t) \\ &= \frac{1}{N} \sum_{i=1}^N \rho(y_t | \tilde{x}_t^i). \end{aligned} \quad (28)$$

This is implemented in step 3 of Algorithm 1 and 2, i.e., in the importance weight computation. In order to make sure that (27) is fulfilled the algorithm has to be modified. The modification takes the following form, in sampling for  $\{\tilde{x}_t^i\}_{i=1}^N$  in step 2 of Algorithm 1 and 2, it is required that the following inequality is satisfied

$$(\tilde{\pi}_{t|t-1}^N, \rho) = \frac{1}{N} \sum_{i=1}^N \rho(y_t | \tilde{x}_t^i) \geq \gamma_t > 0. \quad (29)$$

Now, clearly, the threshold  $\gamma_t$  must be chosen so that the inequality may be satisfied for sufficiently large  $N$ , i.e., so that the true conditional expectation is larger than  $\gamma_t$ . Since this value is typically unknown, it may mean that the problem dependent constant  $\gamma_t$  has to be selected by trial and error and experience. If the inequality (29) holds, the algorithm proceeds as proposed, whereas if it does not hold, a new set of particles  $\{\tilde{x}_t^i\}_{i=1}^N$  is generated and (29) is checked again and so on. The modified algorithm is given in Algorithm 3 below.

*Algorithm 3: A modified particle filter*

- 1) Initialize the particles,  $\{x_0^i\}_{i=1}^N \sim \pi_0(dx_0)$ .
- 2) Predict the particles by drawing independent samples according to

$$\tilde{x}_t^i \sim \sum_{j=1}^N \alpha_j^i K(dx_t | x_{t-1}^j), \quad i = 1, \dots, N.$$

- 3) If  $\frac{1}{N} \sum_{i=1}^N \rho(y_t | \tilde{x}_t^i) \geq \gamma_t$ , proceed to step 4 otherwise return to step 2.
- 4) Rename  $\tilde{x}_t^i = \tilde{x}_t^i, i = 1, \dots, N$  and compute the importance weights  $\{w_t^i\}_{i=1}^N$ ,

$$w_t^i = \rho(y_t | \tilde{x}_t^i), \quad i = 1, \dots, N,$$

and normalize  $\tilde{w}_t^i = w_t^i / \sum_{j=1}^N w_t^j$ .

- 5) Resample,  $x_t^i \sim \tilde{\pi}_{t|t}^N(dx_t) = \sum_{i=1}^N \tilde{w}_t^i \delta_{x_t^i}(dx_t), i = 1, \dots, N$ .
- 6) Set  $t := t + 1$  and repeat from step 2.

For each time step, the conditional distribution is

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

The reason for renaming in step 4 is that distribution of the particles changes by the test in step 3,  $\tilde{x}$  which have passed the test have a different distribution from  $\tilde{x}$ . It is interesting to note that this modification, motivated by (12b), makes sense in its own right. Indeed, it has previously, more or less ad hoc, been used as an indicator for divergence in the particle filter and to obtain a more robust algorithm. Furthermore, this modification is related to the well known degeneracy of the particle weights, see e.g., [14], [17] for insightful discussions on this topic.

Clearly, the choice of  $\gamma_t$  may be non-trivial. If it is chosen too large (larger than the true conditional expectation), steps 2-3 may be an infinite loop. However, it will be proved in

Theorem 6.1 in Section VI that such an infinite loop will not occur if  $\gamma_t$  is chosen small enough. It may have to involve some trial and error to tune in such a choice.

It is worth noting that originally given  $\{x_{t-1}^i\}_{i=1}^N$  the joint density of  $\{\tilde{x}_t^i\}_{i=1}^N$  is

$$P[\tilde{x}_t^i = s_i, i = 1, \dots, N] = \prod_{i=1}^N \sum_{j=1}^N \alpha_j^i K(s_i | x_{t-1}^j) \triangleq \Pi_{\alpha_1, \dots, \alpha_N}^N. \quad (30)$$

Yet, after the modification it is changed to be

$$\bar{\Pi}_{\alpha_1, \dots, \alpha_N}^N = \frac{\Pi_{\alpha_1, \dots, \alpha_N}^N I_{[\frac{1}{N} \sum_{i=1}^N \rho(y_t | s_i) \geq \gamma_t]}}{\int \dots \int \Pi_{\alpha_1, \dots, \alpha_N}^N I_{[\frac{1}{N} \sum_{i=1}^N \rho(y_t | s_i) \geq \gamma_t]} ds_{1:N}}, \quad (31)$$

where the record  $y_t$  is also given.

### B. Numerical Illustration

In order to illustrate the impact of the algorithm modification (29), we study the following nonlinear time-varying system,

$$x_{t+1} = \frac{x_t}{2} + \frac{25x_t}{1+x_t^2} + 8 \cos(1.2t) + v_t, \quad (32a)$$

$$y_t = \frac{x_t^2}{20} + e_t, \quad (32b)$$

where  $v_t \sim \mathcal{N}(0, 10), e_t \sim \mathcal{N}(0, 1)$ , the initial state  $x_0 \sim \mathcal{N}(0, 5)$  and  $\gamma_t = 10^{-4}$ . In the experiment we used 250 time instants and 500 simulations, all using the same measurement sequence. We used the modified particle filter given in Algorithm 3 in order to compute an approximation of the estimate  $\hat{x}_t = E(x_t | y_{1:t})$ . In accordance with both Theorem 6.1 and intuition the quality of the estimate improves with the number of particles  $N$  used in the approximation. The algorithm modification (29) is only active when a small amount of particles is used. That this is indeed the case is evident from Fig. 2, where the average number of interventions due to violations of (29) are given as a function of the number of particles used in the filter.

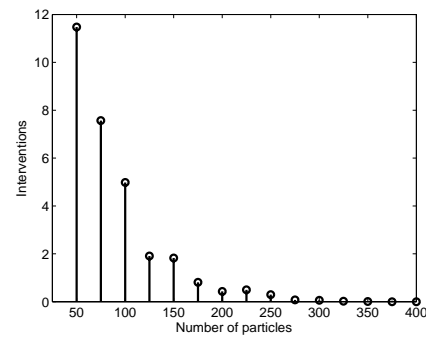


Fig. 2: Illustration of the impact of the algorithm modification (29) introduced in Algorithm 3. The figure shows the number of times (29) was violated and the particles had to be regenerated, as a function of the number of particles used. This is the average result from 500 simulations.

## V. THE BASIC CONVERGENCE RESULT

The filtered state estimate is

$$\hat{x}_t = \mathbb{E}(x_t | Y_{1:t}). \quad (33)$$

This is the mean of the conditional distribution

$$\pi_{t|t}(dx_t) = P(X_t \in dx_t | Y_{1:t} = y_{1:t}). \quad (34)$$

The modified particle filter, given in Algorithm 3, provides an estimate of these two quantities based on  $N$  particles which we denote by

$$\hat{x}_t^N \quad (35)$$

and

$$\pi_{t|t}^N(dx_t). \quad (36)$$

For given  $y_{1:t}$ ,  $\hat{x}_t$  is a given vector, and  $\pi_{t|t}(dx_t)$  is a given function. However,  $\hat{x}_t^N$  and  $\pi_{t|t}^N(dx_t)$  are random, since they depend on the randomly generated particles. Clearly, a crucial question is how these random variables behave as  $N$  increases. We will throughout the remainder of this paper consider this question for a given  $t$  and given observed outputs  $y_{1:t}$ . *Hence all stochastic quantifiers below (like  $\mathbb{E}$  and “w.p.1”) will be with respect to the random variables related to the particles.*

This problem has been well studied in the literature. The excellent survey [14] gives several results of the kind

$$(\pi_{t|t}^N, \phi) = \int \phi(x_t) \pi_{t|t}^N(dx_t) \rightarrow \mathbb{E}(\phi(x_t) | y_{1:t}) \text{ as } N \rightarrow \infty, \quad (37)$$

for functions of the posterior distribution. The notation introduced in (11) has been used in the first equality in (37). Note that the  $i$ -th component of the estimate  $\hat{x}_t^N$  is obtained for  $\phi(x) = x[i]$  where  $x = (x[1], \dots, x[n_x])^T$ ,  $i = 1, \dots, n_x$ . However, apparently all known results on convergence and other properties of (37) assume  $\phi$  to be a bounded function. Therefore convergence of the particle filter state estimate itself cannot be handled by these results.

In this and the following sections we develop results that are valid also for a class of unbounded functions  $\phi$ .

The basic result is a bound on the 4-th moment of the estimated conditional mean

$$\mathbb{E} \left\| \int \phi(x_t) \pi_{t|t}^N(dx_t) - \int \phi(x_t) \pi_{t|t}(dx_t) \right\|^4 \leq \frac{C_\phi}{N^2}. \quad (38)$$

Here  $C_\phi$  is a constant that depends on the function  $\phi$ , which will be defined later. (Of course, it also depends on the fixed variables  $t$  and  $y_{1:t}$ . There is no guarantee that the bound will be uniform in these variables.)

From the Glivenko-Cantelli Lemma [30] we have

$$\int \phi(x_t) \pi_{t|t}^N(dx_t) \rightarrow \int \phi(x_t) \pi_{t|t}(dx_t) \text{ w.p.1 as } N \rightarrow \infty. \quad (39)$$

In particular, under certain conditions applying this result to the cases  $\phi(x) = x[i]$  where  $x = (x[1], \dots, x[n_x])^T$ ,  $i = 1, \dots, n_x$ , we obtain

$$\hat{x}_t^N \rightarrow \hat{x}_t \text{ w.p.1 as } N \rightarrow \infty.$$

So the particle filter state estimate will converge to the true estimate as the number of particles tends to infinity (for given  $t$  and for any given sequence  $y_{1:t}$ ), subject to certain conditions (see the discussions of the defined conditions below).

## VI. THE MAIN RESULT

To formally prove the results of the previous section we need to assume certain conditions for the filtering problem and the function  $\phi$  in (37). The first one is to assure that Bayes' formula (10b) (or (12b)) is well defined, so that the numerator is guaranteed to be nonzero:

$$(\pi_{t|t-1}, \rho) = \int_{R^{n_x}} \rho(y_t | x_t) \pi_{t|t-1}(dx_t) > 0$$

Since  $\rho(y_t | x_t)$  is the conditional density of  $y_t$  given the state  $x_t$  and  $\pi_{t|t-1}(dx_t)$  is the conditional density of  $x_t$  given  $y_{1:t-1}$  this expression is the *conditional density of  $y_t$  given previous outputs*  $p(y_t | y_{1:t-1})$ . To assume that this conditional density is nonzero is no major restriction, since the condition is to be imposed on the observed sequence of  $y_t$ .

**H0.** For given  $y_{1:s}$ ,  $s = 1, \dots, t$ ,  $(\pi_{s|s-1}, \rho) > 0$ ; and the constant  $\gamma_s$  used in the modified algorithm satisfies

$$0 < \gamma_s < (\pi_{s|s-1}, \rho), \quad s = 1, \dots, t.$$

We also need to assume that the conditional densities  $K$  and  $\rho$  are bounded. Hence, the first condition on the densities of the system is

**H1.**  $\rho(y_s | x_s) < \infty$ ;  $K(x_s | x_{s-1}) < \infty$  for given  $y_{1:s}$ ,  $s = 1, \dots, t$ .

To prove results for a general function  $\phi(x)$  in (37) we also need some mild restrictions on how fast it may increase with  $x$ . This is expressed using the conditional observation density  $\rho$ :

**H2.** The function  $\phi(\cdot)$  satisfies  $\sup_{x_s} |\phi(x_s)|^4 \rho(y_s | x_s) < C(y_{1:s})$  for given  $y_{1:s}$ ,  $s = 1, \dots, t$ .

Note that  $C(y_{1:s})$  in H2 is a finite constant that may depend on  $y_{1:s}$ .

The essence of condition H2 is that the conditional observation density (for given  $y_s$ ) decreases faster than the  $\phi$  function increases. Since typical distributions decay exponentially or have bounded support, this is not a strong restriction for  $\phi$ .

Note that H1 and H2 imply that the conditional fourth moment of  $\phi$  is bounded.

$$\begin{aligned} \int |\phi(x)|^4 \pi_{s|s}(dx) &= \frac{\int |\phi(x)|^4 \rho(y_s | x) \pi_{s|s-1}(dx)}{(\pi_{s|s-1}, \rho)} \\ &\leq \frac{C(y_{1:s}) \int \pi_{s|s-1}(dx)}{(\pi_{s|s-1}, \rho)} < \infty \end{aligned}$$

The following examples provide two typical one dimensional noises, i.e.,  $n_x = n_y = 1$ , satisfying condition H2.

*Example 6.1:*  $p_e(z, s) = O(\exp(-|z|^\nu))$  as  $z \rightarrow \infty$  with  $\nu > 0$ ; and  $\liminf_{|x| \rightarrow \infty} \frac{|h(x, s)|}{|x|^{\nu_1}} > 0$  with  $\nu_1 > 0$ ,  $s = 1, \dots, t$ . It is now easy to verify that H2 holds for any function  $\phi$  satisfying  $\phi(z) = O(|z|^q)$  as  $z \rightarrow \infty$ , where  $q \geq 0$ .

*Example 6.2:*  $p_e(z, s) = \frac{1}{b-a} I_{[a,b]}$  with  $a < 0 < b$ ; and function  $h(x, s) \triangleq h_s$  satisfying that the set  $h_s^{-1}([y-b, y-a])$  is bounded for any given  $y_s$ ,  $s = 1, \dots, t$ . It is now easy to verify that H2 holds for any function  $\phi$ .

Before we give the main result, let us introduce the following notation. The class of functions  $\phi$  satisfying H2 will be denoted by

$$L_t^4(\rho), \quad (40)$$

where  $\rho$  satisfies H1.

*Theorem 6.1:* Suppose that H0, H1 and H2 hold and consider the modified version of the particle filter algorithm (Algorithm 3). Then the following holds:

(i) For sufficiently large  $N$ , the algorithm will not run into an infinite loop in steps 2-3.

(ii) For any  $\phi \in L_t^4(\rho)$ , there exists a constant  $C_{t|t}$ , independent of  $N$  such that

$$\mathbb{E} \left| (\pi_{t|t}^N, \phi) - (\pi_{t|t}, \phi) \right|^4 \leq C_{t|t} \frac{\|\phi\|_{t,4}^4}{N^2}, \quad (41)$$

where  $\|\phi\|_{t,4} \triangleq \max \{1, (\pi_{s|s}, |\phi|^4)^{1/4}, s = 0, 1, \dots, t\}$  and  $\pi_{s|s}^N$  is generated by the algorithm.

By the Borel-Cantelli lemma, e.g., [30], we have a corollary as follow.

*Corollary 6.1:* If H1 and H2 hold, then for any  $\phi \in L_t^4(\rho)$ ,

$$\lim_{N \rightarrow \infty} (\pi_{t|t}^N, \phi) = (\pi_{t|t}, \phi), \quad \text{almost surely.} \quad (42)$$

## VII. PROOF

In this section we will give the proof for the main result, given above in Theorem 6.1. However, before starting the proof we list some lemmas that will be used in the proof.

### A. Auxiliary Lemmas

It is clear that the inequalities in Lemmas 7.1 and 7.4 hold almost surely, since they are in the form of a conditional expectation. For the sake of brevity we omit the notation for almost sure in the following lemmas and their proof. Furthermore, it is also easy to see that Lemmas 7.2 and 7.3 also hold if conditional expectation is used.

*Lemma 7.1:* Let  $\{\xi_i, i = 1, \dots, n\}$  be conditionally independent random variables given  $\sigma$ -algebra  $\mathcal{G}$  such that  $\mathbb{E}(\xi_i|\mathcal{G}) = 0$ ,  $\mathbb{E}(|\xi_i|^4|\mathcal{G}) < \infty$ . Then

$$\mathbb{E} \left( \left| \sum_{i=1}^n \xi_i \right|^4 \middle| \mathcal{G} \right) \leq \sum_{i=1}^n \mathbb{E}(|\xi_i|^4|\mathcal{G}) + \left( \sum_{i=1}^n \mathbb{E}(|\xi_i|^2|\mathcal{G}) \right)^2. \quad (43)$$

*Proof:* Notice that

$$\mathbb{E} \left( \left| \sum_{i=1}^n \xi_i \right|^4 \middle| \mathcal{G} \right) = \sum_{i=1}^n \mathbb{E}(|\xi_i|^4|\mathcal{G}) + \sum_{i,j,i \neq j} \mathbb{E}(|\xi_i|^2|\mathcal{G}) \cdot \mathbb{E}(|\xi_j|^2|\mathcal{G}),$$

the assertion follows.  $\blacksquare$

*Lemma 7.2:* If  $\mathbb{E}|\xi|^p < \infty$ , then  $\mathbb{E}|\xi - \mathbb{E}\xi|^p \leq 2^p \mathbb{E}|\xi|^p$ , for any  $p \geq 1$ .

*Proof:* By Jensen's inequality (e.g., [30]), for  $p \geq 1$ ,  $(\mathbb{E}|\xi|)^p \leq \mathbb{E}|\xi|^p$ . Hence,  $\mathbb{E}|\xi| \leq (\mathbb{E}|\xi|^p)^{1/p}$ . Then by Minkowski's inequality (e.g., [30]),

$$(\mathbb{E}|\xi - \mathbb{E}\xi|^p)^{1/p} \leq (\mathbb{E}|\xi|^p)^{1/p} + |\mathbb{E}\xi| \leq 2(\mathbb{E}|\xi|^p)^{1/p}, \quad (44)$$

which derives the desired inequality.  $\blacksquare$

*Lemma 7.3:* If  $1 \leq r_1 \leq r_2$  and  $\mathbb{E}|\xi|^{r_2} < \infty$ , then  $\mathbb{E}^{1/r_1}|\xi|^{r_1} \leq \mathbb{E}^{1/r_2}|\xi|^{r_2}$ .

*Proof:* Simply by Hölder's inequality (e.g., [30]):  $\mathbb{E}(|\xi|^{r_1} \cdot 1) \leq \mathbb{E}^{r_1/r_2}(|\xi|^{r_1})^{r_2/r_1}$ . Then the assertion follows.  $\blacksquare$

Based on Lemmas 7.1 and 7.3, we have

*Lemma 7.4:* Let  $\{\xi_i, i = 1, \dots, n\}$  be conditionally independent random variables given  $\sigma$ -algebra  $\mathcal{G}$  such that  $\mathbb{E}(\xi_i|\mathcal{G}) = 0$ ,  $\mathbb{E}(|\xi_i|^4|\mathcal{G}) < \infty$ . Then

$$\mathbb{E} \left( \left| \frac{1}{n} \sum_{i=1}^n \xi_i \right|^4 \middle| \mathcal{G} \right) \leq \frac{2 \max_{1 \leq i \leq n} \mathbb{E}(|\xi_i|^4|\mathcal{G})}{n^2}. \quad (45)$$

*Lemma 7.5:* Let the probability density function for the random variable  $\eta$  be  $p(x)$  and let the probability density function for the random variable  $\xi$  be

$$\frac{p(x)I_A}{\int p(y)I_A dy},$$

where  $I_A$  is the indicator function for a set  $A$ , such that

$$P[\eta \in \Omega - A] \leq \epsilon < 1. \quad (46)$$

Let  $\psi$  be a measurable function satisfying  $\mathbb{E}\psi^2(\eta) < \infty$ . Then, we have

$$|\mathbb{E}\psi(\xi) - \mathbb{E}\psi(\eta)| \leq \frac{2\sqrt{\mathbb{E}\psi^2(\eta)}}{1-\epsilon}\sqrt{\epsilon}. \quad (47)$$

In the case  $\mathbb{E}|\psi(\eta)| < \infty$ ,

$$\mathbb{E}|\psi(\xi)| \leq \frac{\mathbb{E}|\psi(\eta)|}{1-\epsilon}. \quad (48)$$

**Proof.** Clearly, since the density of  $\xi$  is

$$\frac{p(t)I_A}{\int p(y)I_A dy},$$

it is easy to show (48) as follows

$$\begin{aligned} \mathbb{E}|\psi(\xi)| &= \left| \frac{\int \psi(t)p(t)I_A dt}{\int p(y)I_A dy} \right| \leq \frac{1}{1-\epsilon} \int |\psi(t)p(t)I_A| dt \\ &\leq \frac{1}{1-\epsilon} \int |\psi(t)p(t)| dt = \frac{\mathbb{E}|\psi(\eta)|}{1-\epsilon} \end{aligned}$$

While

$$\begin{aligned} |\mathbb{E}\psi(\xi) - \mathbb{E}\psi(\eta)| &= \left| \frac{\int \psi(t)p(t)I_A dt}{\int p(y)I_A dy} - \int \psi(t)p(t) dt \right| \\ &\leq \frac{1}{1-\epsilon} \left| \int \psi(t)p(t)I_A dt - \int \psi(t)p(t) dt \cdot (1-\epsilon) \right| \\ &\leq \frac{1}{1-\epsilon} \left[ \int |\psi(t)p(t)I_{\Omega-A}| dt + \int |\psi(t)p(t) dt \cdot \epsilon \right] \\ &\leq \frac{1}{1-\epsilon} \left[ \sqrt{\int |\psi(t)|^2 p(t) dt} \cdot \sqrt{\int p(t)I_{\Omega-A} dt} + \mathbb{E}|\psi(\eta)| \cdot \epsilon \right] \\ &\leq \frac{1}{1-\epsilon} \left[ \sqrt{\mathbb{E}\psi^2(\eta)} \cdot \sqrt{\epsilon} + \mathbb{E}|\psi(\eta)| \cdot \epsilon \right] \leq \frac{2\sqrt{\mathbb{E}\psi^2(\eta)}}{1-\epsilon}\sqrt{\epsilon}, \end{aligned}$$

which derives (47).  $\square$

The result of Lemma 7.5 can be extended to cover conditional expectations as well.

## B. Proof of Theorem 6.1

*Proof:* The proof is carried out in the standard induction framework, employed for example in [14].

### 1: Initialization

Let  $\{x_0^i\}_{i=1}^N$  be independent random variables with the same distribution  $\pi_0(dx_0)$ . Then, using Lemmas 7.4 and 7.2, it is clear that

$$\begin{aligned} \mathbb{E} |(\pi_0^N, \phi) - (\pi_0, \phi)|^4 &= \frac{1}{N^4} \mathbb{E} \left| \sum_{i=1}^N (\phi(x_0^i) - \mathbb{E}(\phi(x_0^i))) \right|^4 \\ &\leq \frac{2}{N^2} \mathbb{E} |\phi(x_0^i) - \mathbb{E}(\phi(x_0^i))|^4 \\ &\leq \frac{32}{N^2} \mathbb{E} |\phi(x_0^i)|^4 \leq \frac{32}{N^2} \|\phi\|_{0,4}^4 \\ &\triangleq C_{0|0} \frac{\|\phi\|_{0,4}^4}{N^2}. \end{aligned} \quad (49)$$

Similarly,

$$\begin{aligned} \mathbb{E} |(\pi_0^N, |\phi|^4) - (\pi_0, |\phi|^4)| \\ \leq \frac{1}{N} \mathbb{E} \left| \sum_{i=1}^N (|\phi(x_0^i)|^4 - \mathbb{E} |\phi(x_0^i)|^4) \right| \\ \leq 2 \mathbb{E} |\phi(x_0^i)|^4. \end{aligned}$$

Note that  $x_0^i$  have the same distribution for all  $i$ , so the expected values do not depend on  $i$ . Hence,

$$\mathbb{E} |(\pi_0^N, |\phi|^4) - (\pi_0, |\phi|^4)| \leq 3 \mathbb{E} |\phi(x_0^i)|^4 \triangleq M_{0|0} \|\phi\|_{0,4}^4. \quad (50)$$

### 2: Prediction

Based on (49) and (50), we assume that for  $t-1$  and  $\forall \phi \in L_t^4(\rho)$

$$\mathbb{E} |(\pi_{t-1|t-1}^N, \phi) - (\pi_{t-1|t-1}, \phi)|^4 \leq C_{t-1|t-1} \frac{\|\phi\|_{t-1,4}^4}{N^2} \quad (51)$$

and

$$\mathbb{E} |(\pi_{t-1|t-1}^N, |\phi|^4) - (\pi_{t-1|t-1}, |\phi|^4)| \leq M_{t-1|t-1} \|\phi\|_{t-1,4}^4 \quad (52)$$

holds, where  $C_{t-1|t-1} > 0$  and  $M_{t-1|t-1} > 0$ . We analyse  $\mathbb{E} |(\tilde{\pi}_{t|t-1}^N, \phi) - (\pi_{t|t-1}, \phi)|^4$  and  $\mathbb{E} |(\tilde{\pi}_{t|t-1}^N, |\phi|^4) - (\pi_{t|t-1}, |\phi|^4)|$  in this step.

Let  $\mathcal{F}_{t-1}$  denote the  $\sigma$ -algebra generated by  $\{x_{t-1}^i, i = 1, \dots, N\}$ . Notice that

$$(\tilde{\pi}_{t|t-1}^N, \phi) - (\pi_{t|t-1}, \phi) \triangleq \Pi_1 + \Pi_2 + \Pi_3,$$

where

$$\begin{aligned} \Pi_1 &\triangleq (\tilde{\pi}_{t|t-1}^N, \phi) - \frac{1}{N} \sum_{i=1}^N \mathbb{E} [\phi(\tilde{x}_t^i) | \mathcal{F}_{t-1}], \\ \Pi_2 &\triangleq \frac{1}{N} \sum_{i=1}^N \mathbb{E} [\phi(\tilde{x}_t^i) | \mathcal{F}_{t-1}] - \frac{1}{N} \sum_{i=1}^N (\pi_{t-1|t-1}^{N, \alpha_i}, K\phi), \\ \Pi_3 &\triangleq \frac{1}{N} \sum_{i=1}^N (\pi_{t-1|t-1}^{N, \alpha_i}, K\phi) - (\pi_{t|t-1}, \phi), \end{aligned}$$

and  $\pi_{t-1|t-1}^{N, \alpha_i} = \sum_{j=1}^N \alpha_j^i \delta_{x_{t-1}^j} (dx_{t-1})$ . We consider the three terms  $\Pi_1$ ,  $\Pi_2$  and  $\Pi_3$  separately in the following.

Let  $\tilde{x}_t^i$  be drawn from the distribution  $(\pi_{t-1|t-1}^{N, \alpha_i}, K)$  as in step 2 of the algorithm. Then we have

$$\mathbb{E} [\phi(\tilde{x}_{t-1}^i) | \mathcal{F}_{t-1}] = (\pi_{t-1|t-1}^{N, \alpha_i}, K\phi). \quad (53)$$

Recall that the distribution of  $\tilde{x}_t^i$  differs from the distribution of  $\tilde{x}_{t-1}^i$ , which has passed the test in step 3 of the algorithm and is thus conditioned on the event

$$A_t = \{(\pi_{t-1|t-1}^N, K\rho) \geq \gamma_t\} \quad (54)$$

Now, let us check the probability of this event. In view of (53) and (22)

$$\mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^N \rho(y_s | \tilde{x}_s^i) | \mathcal{F}_{t-1} \right] = (\pi_{t-1|t-1}^N, K\rho).$$

Thus,

$$P \left[ \frac{1}{N} \sum_{i=1}^N \rho(y_t | \tilde{x}_t^i) < \gamma_t | \mathcal{F}_{t-1} \right] = P \left[ (\pi_{t-1|t-1}^N, K\rho) < \gamma_t \right]. \quad (55)$$

By (51), we have

$$\begin{aligned} P \left[ (\pi_{t-1|t-1}^N, K\rho) < \gamma_t \right] &= P \left[ (\pi_{t-1|t-1}^N, K\rho) - (\pi_{t-1|t-1}, K\rho) \right. \\ &\quad \left. < \gamma_t - (\pi_{t-1|t-1}, K\rho) \right] \\ &\leq P \left[ |(\pi_{t-1|t-1}^N, K\rho) - (\pi_{t-1|t-1}, K\rho)| > |\gamma_t - (\pi_{t-1|t-1}, K\rho)| \right] \\ &\leq \frac{\mathbb{E} |(\pi_{t-1|t-1}^N, K\rho) - (\pi_{t-1|t-1}, K\rho)|^4}{|\gamma_t - (\pi_{t-1|t-1}, K\rho)|^4} \\ &\leq \frac{C_{t-1|t-1} \|K\|^4}{|\gamma_t - (\pi_{t-1|t-1}, \rho)|^4} \cdot \frac{\|\rho\|_{t-1,4}^4}{N^2} \triangleq C_{\gamma_t} \cdot \frac{\|\rho\|_{t-1,4}^4}{N^2}. \end{aligned} \quad (56)$$

Here we used condition H0. Consequently, for sufficiently large  $N$  we have

$$P(A_t) > 1 - \epsilon_t; \quad 0 < \epsilon_t < 1$$

We can now handle the difference between  $\tilde{x}_t^i$  and  $\tilde{x}_{t-1}^i$  using Lemma 7.5, and by Lemmas 7.1, 7.2, (53) and (22), we obtain

$$\begin{aligned} \mathbb{E} [|\Pi_1|^4 | \mathcal{F}_{t-1}] &= \frac{1}{N^4} \mathbb{E} \left[ \left| \sum_{i=1}^N [\phi(\tilde{x}_t^i) - \mathbb{E}(\phi(\tilde{x}_t^i) | \mathcal{F}_{t-1})] \right|^4 | \mathcal{F}_{t-1} \right] \\ &\leq \frac{2^4}{N^4} \left[ \sum_{i=1}^N \mathbb{E} [|\phi(\tilde{x}_t^i)|^4 | \mathcal{F}_{t-1}] + \left( \sum_{i=1}^N \mathbb{E} [|\phi(\tilde{x}_t^i)|^2 | \mathcal{F}_{t-1}] \right)^2 \right] \\ &\leq \frac{2^4}{N^4(1-\epsilon_t)^2} \left[ \sum_{i=1}^N \mathbb{E} [|\phi(\tilde{x}_t^i)|^4 | \mathcal{F}_{t-1}] + \left( \sum_{i=1}^N \mathbb{E} [|\phi(\tilde{x}_t^i)|^2 | \mathcal{F}_{t-1}] \right)^2 \right] \\ &\leq \frac{2^4}{N^4(1-\epsilon_t)^2} \left[ \sum_{i=1}^N (\pi_{t-1|t-1}^{N, \alpha_i}, K|\phi|^4) + \left( \sum_{i=1}^N (\pi_{t-1|t-1}^{N, \alpha_i}, K|\phi|^2) \right)^2 \right] \\ &\leq \frac{2^4}{(1-\epsilon_t)^2} \left[ \frac{(\pi_{t-1|t-1}^N, K|\phi|^4)}{N^3} + \frac{(\pi_{t-1|t-1}^N, K|\phi|^2)^2}{N^2} \right]. \end{aligned}$$

Hence, by Lemma 7.3 and (52),

$$\mathbb{E} |\Pi_1|^4 \leq \frac{2^5 \|K\|^4 M_{t-1|t-1}}{(1-\epsilon_t)^2} \cdot \frac{\|\phi\|_{t-1,4}^4}{N^2} \triangleq C_{\Pi_1} \cdot \frac{\|\phi\|_{t-1,4}^4}{N^2}. \quad (57)$$



By (53), Lemma 7.5 and (22),

$$\begin{aligned}
 |\Pi_2|^4 &= \left| \frac{1}{N} \sum_{i=1}^N E[\phi(\tilde{x}_t^i) | \mathcal{F}_{t-1}] - \frac{1}{N} \sum_{i=1}^N E[\phi(\tilde{x}_t^i) | \mathcal{F}_{t-1}] \right|^4 \\
 &= \left| \frac{1}{N} \sum_{i=1}^N (E[\phi(\tilde{x}_t^i) | \mathcal{F}_{t-1}] - E[\phi(\tilde{x}_t^i) | \mathcal{F}_{t-1}]) \right|^4 \\
 &\leq \frac{2^4 C_{\gamma_t}^2 \|\rho\|_{t-1,4}^8}{(1-\epsilon_t)^4 N^4} \cdot \frac{1}{N} \sum_{i=1}^N (\pi_{t-1|\tau-1}^{N, \alpha_i}, K \phi^4) \\
 &= \frac{2^4 C_{\gamma_t}^2 \|\rho\|_{t-1,4}^8}{(1-\epsilon_t)^4 N^4} \cdot (\pi_{t-1|\tau-1}^N, K \phi^4) \\
 &\triangleq C_{\Pi_2} \cdot \frac{(\pi_{t-1|\tau-1}^N, K \phi^4)}{N^4}.
 \end{aligned}$$

Hence,

$$E|\Pi_2|^4 \leq C_{\Pi_2} \cdot \frac{\|K\| \cdot \|\phi\|_{t-1,4}^4}{N^4} \leq C_{\Pi_2} \|K\| \cdot \frac{\|\phi\|_{t-1,4}^4}{N^2}. \quad (58)$$

This proves the first part of Theorem 6.1, i.e., that the algorithm will not run into an infinite loop in steps 2 – 3.

By (22) and (51),

$$E|\Pi_3|^4 \leq C_{t-1|\tau-1} \|K\|^4 \cdot \frac{\|\phi\|_{t-1,4}^4}{N^2} \triangleq C_{\Pi_3} \cdot \frac{\|\phi\|_{t-1,4}^4}{N^2}. \quad (59)$$

Then, using Minkowski's inequality, (57), (58) and (59), we have

$$\begin{aligned}
 E^{1/4} \left| (\tilde{\pi}_{t|\tau-1}^N, \phi) - (\pi_{t|\tau-1}, \phi) \right|^4 &\leq E^{1/4} |\Pi_1|^4 + E^{1/4} |\Pi_2|^4 \\
 &+ E^{1/4} |\Pi_3|^4 \leq \left( C_{\Pi_1}^{1/4} + [C_{\Pi_2} \|K\|]^{1/4} + C_{\Pi_3}^{1/4} \right) \frac{\|\phi\|_{t-1,4}}{N^{1/2}} \\
 &\triangleq \tilde{C}_{t|\tau-1}^{1/4} \frac{\|\phi\|_{t-1,4}}{N^{1/2}}.
 \end{aligned}$$

That is

$$E \left| (\tilde{\pi}_{t|\tau-1}^N, \phi) - (\pi_{t|\tau-1}, \phi) \right|^4 \leq \tilde{C}_{t|\tau-1} \frac{\|\phi\|_{t-1,4}^4}{N^2}. \quad (60)$$

By Lemma 7.2 and (52)

$$\begin{aligned}
 E \left( E \left( (\tilde{\pi}_{t|\tau-1}^N, |\phi|^4) - \frac{1}{N} \sum_{i=1}^N (\pi_{t-1|\tau-1}^{N, \alpha_i}, K |\phi|^4) | \mathcal{F}_{t-1} \right) \right) \\
 = \frac{1}{N} E \left( E \left( \sum_{i=1}^N (|\phi(\tilde{x}_{t-1}^i)|^4 - E(|\phi(\tilde{x}_{t-1}^i)|^4 | \mathcal{F}_{t-1})) \right) \right) \\
 \leq 2 E(\pi_{t-1|\tau-1}^N, K |\phi|^4) \leq 2 \|K\|^4 M_{t-1|\tau-1} \|\phi\|_{t-1,4}^4.
 \end{aligned}$$

Then, using a similar separation mentioned above, by (52) we have

$$\begin{aligned}
 E \left| (\tilde{\pi}_{t|\tau-1}^N, |\phi|^4) - (\pi_{t|\tau-1}, |\phi|^4) \right| \\
 \leq \|K\|^4 (3M_{t-1|\tau-1} + 1) \|\phi\|_{t-1,4}^4 \triangleq \tilde{M}_{t|\tau-1} \|\phi\|_{t-1,4}^4. \quad (61)
 \end{aligned}$$

### 3: Update

In this step we go one step further to analyse  $E \left| (\tilde{\pi}_{t|\tau}^N, \phi) - (\pi_{t|\tau}, \phi) \right|^4$  and  $E(\tilde{\pi}_{t|\tau}^N, |\phi|^4)$  based on (60) and (61). Clearly,

$$(\tilde{\pi}_{t|\tau}^N, \phi) - (\pi_{t|\tau}, \phi) = \frac{(\tilde{\pi}_{t|\tau-1}^N, \rho \phi)}{(\tilde{\pi}_{t|\tau-1}^N, \rho)} - \frac{(\pi_{t|\tau-1}, \rho \phi)}{(\pi_{t|\tau-1}, \rho)} = \tilde{\Pi}_1 + \tilde{\Pi}_2,$$

where

$$\begin{aligned}
 \tilde{\Pi}_1 &\triangleq \frac{(\tilde{\pi}_{t|\tau-1}^N, \rho \phi)}{(\tilde{\pi}_{t|\tau-1}^N, \rho)} - \frac{(\tilde{\pi}_{t|\tau-1}^N, \rho \phi)}{(\pi_{t|\tau-1}, \rho)}, \\
 \tilde{\Pi}_2 &\triangleq \frac{(\tilde{\pi}_{t|\tau-1}^N, \rho \phi)}{(\pi_{t|\tau-1}, \rho)} - \frac{(\pi_{t|\tau-1}, \rho \phi)}{(\pi_{t|\tau-1}, \rho)}.
 \end{aligned}$$

By condition H1 and the modified version of the algorithm we have,

$$\begin{aligned}
 |\tilde{\Pi}_1| &= \left| \frac{(\tilde{\pi}_{t|\tau-1}^N, \rho \phi)}{(\tilde{\pi}_{t|\tau-1}^N, \rho)} \cdot \frac{[(\pi_{t|\tau-1}, \rho) - (\tilde{\pi}_{t|\tau-1}^N, \rho)]}{(\pi_{t|\tau-1}, \rho)} \right| \\
 &\leq \frac{\|\rho \phi\|}{\gamma_t (\pi_{t|\tau-1}, \rho)} \left| (\pi_{t|\tau-1}, \rho) - (\tilde{\pi}_{t|\tau-1}^N, \rho) \right|. \quad (62)
 \end{aligned}$$

Here,  $\gamma_t$  is the threshold used in step 3 of the modified filter (Algorithm 3). Thus, by Minkowski's inequality, (60) and (62),

$$\begin{aligned}
 E^{1/4} \left| (\tilde{\pi}_{t|\tau}^N, \phi) - (\pi_{t|\tau}, \phi) \right|^4 \\
 \leq E^{1/4} |\tilde{\Pi}_1|^4 + E^{1/4} |\tilde{\Pi}_2|^4 \\
 \leq \frac{\tilde{C}_{t|\tau-1}^{1/4} \|\rho\| (\|\rho \phi\| + \gamma_t)}{\gamma_t (\pi_{t|\tau-1}, \rho)} \cdot \frac{\|\phi\|_{t-1,4}}{N^{1/2}} \\
 \triangleq \tilde{C}_{t|\tau}^{1/4} \frac{\|\phi\|_{t-1,4}}{N^{1/2}},
 \end{aligned}$$

which implies

$$E \left| (\tilde{\pi}_{t|\tau}^N, \phi) - (\pi_{t|\tau}, \phi) \right|^4 \leq \tilde{C}_{t|\tau} \frac{\|\phi\|_{t-1,4}^4}{N^2}. \quad (63)$$

Using a similar separation mentioned above, by (61),

$$\begin{aligned}
 E \left| (\tilde{\pi}_{t|\tau}^N, |\phi|^4) - (\pi_{t|\tau}, |\phi|^4) \right| \\
 \leq E \left| \frac{(\tilde{\pi}_{t|\tau-1}^N, \rho |\phi|^4)}{(\pi_{t|\tau-1}^N, \rho)} - \frac{(\tilde{\pi}_{t|\tau-1}^N, \rho |\phi|^4)}{(\pi_{t|\tau-1}, \rho)} \right| \\
 + E \left| \frac{(\tilde{\pi}_{t|\tau-1}^N, \rho |\phi|^4)}{(\pi_{t|\tau-1}, \rho)} - \frac{(\pi_{t|\tau-1}, \rho |\phi|^4)}{(\pi_{t|\tau-1}, \rho)} \right| \\
 \leq \frac{\|\rho \phi^4\| \cdot 2\|\rho\|}{\gamma_t (\pi_{t|\tau-1}, \rho)} + \frac{\tilde{M}_{t|\tau-1} \max\{\|\rho\|, 1\}}{(\pi_{t|\tau-1}, \rho)} \|\phi\|_{t-1,4}^4.
 \end{aligned}$$

Observe that  $\|\phi\|_{s,4} \geq 1$  is increasing with respect to  $s$ . We have

$$\begin{aligned}
 E \left| (\tilde{\pi}_{t|\tau}^N, |\phi|^4) \right| \\
 \leq \frac{\|\rho \phi^4\| \cdot 2\|\rho\|}{\gamma_t (\pi_{t|\tau-1}, \rho)} + \frac{\tilde{M}_{t|\tau-1} \max\{\|\rho\|, 1\}}{(\pi_{t|\tau-1}, \rho)} \|\phi\|_{t-1,4}^4 + (\pi_{t|\tau}, |\phi|^4), \\
 \leq 3 \max \left\{ \frac{\|\rho \phi^4\| \cdot 2\|\rho\|}{\gamma_t (\pi_{t|\tau-1}, \rho)}, \frac{\tilde{M}_{t|\tau-1} \max\{\|\rho\|, 1\}}{(\pi_{t|\tau-1}, \rho)}, 1 \right\} \cdot \|\phi\|_{t,4}^4 \\
 \triangleq \tilde{M}_{t|\tau} \|\phi\|_{t,4}^4. \quad (64)
 \end{aligned}$$

#### 4: Resampling

Finally, we analyse  $\mathbb{E} \left| (\pi_{t|t}^N, \phi) - (\pi_{t|t}, \phi) \right|^4$  and  $\mathbb{E}(\pi_{t|t}^N, |\phi|^4)$  based on (63) and (64). It is now easy to see that

$$(\pi_{t|t}^N, \phi) - (\pi_{t|t}, \phi) = \bar{\Pi}_1 + \bar{\Pi}_2,$$

where

$$\bar{\Pi}_1 \triangleq (\pi_{t|t}^N, \phi) - (\tilde{\pi}_{t|t}^N, \phi), \quad \bar{\Pi}_2 \triangleq (\tilde{\pi}_{t|t}^N, \phi) - (\pi_{t|t}, \phi).$$

Let  $\mathcal{G}_t$  denote the  $\sigma$ -algebra generated by  $\{\tilde{x}_t^i, i = 1, \dots, N\}$ . From the generation of  $x_t^i$ , we have,

$$\mathbb{E}(\phi(x_t^i) | \mathcal{G}_t) = (\tilde{\pi}_{t|t}^N, \phi),$$

and then

$$\bar{\Pi}_1 = \frac{1}{N} \sum_{i=1}^N (\phi(x_t^i) - \mathbb{E}(\phi(x_t^i) | \mathcal{G}_t)).$$

Then, by Lemmas 7.4, 7.2,

$$\begin{aligned} \mathbb{E}(|\bar{\Pi}_1|^4 | \mathcal{G}_t) &= \frac{1}{N^4} \mathbb{E} \left( \left| \sum_{i=1}^N (\phi(x_t^i) - \mathbb{E}(\phi(x_t^i) | \mathcal{G}_t)) \right|^4 \middle| \mathcal{G}_t \right) \\ &\leq 2^5 \frac{\mathbb{E}(|\phi(x_t^1)|^4 | \mathcal{G}_t)}{N^2} = 2^5 \frac{(\tilde{\pi}_{t|t}^N, |\phi|^4)}{N^2}. \end{aligned}$$

Thus, by (64),

$$\mathbb{E}|\bar{\Pi}_1|^4 \leq 2^5 \tilde{M}_{t|t} \frac{\|\phi\|_{t,4}^4}{N^2}. \quad (65)$$

Using Minkowski's inequality, (63) and (65) we have

$$\begin{aligned} \mathbb{E}^{1/4} \left| (\pi_{t|t}^N, \phi) - (\pi_{t|t}, \phi) \right|^4 &\leq \mathbb{E}^{1/4} |\bar{\Pi}_1|^4 + \mathbb{E}^{1/4} |\bar{\Pi}_2|^4 \\ &\leq \left( [2^5 \tilde{M}_{t|t}]^{1/4} + \tilde{C}_{t|t}^{1/4} \right) \frac{\|\phi\|_{t,4}}{N^{1/2}} \\ &\triangleq C_{t|t}^{1/4} \frac{\|\phi\|_{t,4}}{N^{1/2}}. \end{aligned}$$

That is

$$\mathbb{E} \left| (\pi_{t|t}^N, \phi) - (\pi_{t|t}, \phi) \right|^4 \leq C_{t|t} \frac{\|\phi\|_{t,4}^4}{N^2}. \quad (66)$$

Using a separation similar to the one mentioned above, by (64), we have,

$$\begin{aligned} &\mathbb{E} \left| (\pi_{t|t}^N, |\phi|^4) - (\pi_{t|t}, |\phi|^4) \right| \\ &\leq \mathbb{E} \left| (\pi_{t|t}^N, |\phi|^4) - (\tilde{\pi}_{t|t}^N, |\phi|^4) \right| + \mathbb{E} \left| (\tilde{\pi}_{t|t}^N, |\phi|^4) - (\pi_{t|t}, |\phi|^4) \right| \\ &\leq [2\tilde{M}_{t|t} + (\tilde{M}_{t|t} + 1)] \|\phi\|_{t,4}^4 \\ &\leq (3\tilde{M}_{t|t} + 1) \|\phi\|_{t,4}^4. \end{aligned}$$

Hence,

$$\mathbb{E} \left| (\pi_{t|t}^N, |\phi|^4) \right| \leq (3\tilde{M}_{t|t} + 2) \|\phi\|_{t,4}^4 \triangleq M_{t|t} \|\phi\|_{t,4}^4. \quad (67)$$

Therefore, the proof of Theorem 6.1 is completed, since (51) and (52) are successfully replaced by (66) and (67). ■

#### VIII. CONCLUSION

The basic contribution of this paper has been the extension of the existing convergence results to unbounded functions  $\phi$ , which has allowed statements on the filter estimate (conditional expectation) itself. We have had to introduce a slight modification of the particle filter (Algorithm 3) in order to complete the proof. This modification leads to an improved result in practise, which was illustrated by a simple simulation. The simulation study also showed that the effect of the modification decreases with an increased number of particles, all in accordance to theory.

Results similar to the one in (38) can be obtained for moments other than four. This more general case of  $L^p$ -convergence for an arbitrary  $p > 1$  is under consideration, using a Rosenthal-type of inequality [22].

#### ACKNOWLEDGEMENT

This work was supported by the strategic research center MOVIII, funded by the Swedish Foundation for Strategic Research, SSF. The authors would also like to thank the anonymous reviewers for their constructive comments on the manuscripts. We also thank Dr. A. Doucet for valuable assistance with references.

#### REFERENCES

- [1] A. H. Jazwinski, *Stochastic processes and filtering theory*, ser. Mathematics in science and engineering. New York, USA: Academic Press, 1970.
- [2] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME, Journal of Basic Engineering*, vol. 82, pp. 35–45, 1960.
- [3] G. L. Smith, S. F. Schmidt, and L. A. McGee, "Application of statistical filter theory to the optimal estimation of position and velocity on board a circumlunar vehicle," NASA, Tech. Rep. TR R-135, 1962.
- [4] S. F. Schmidt, "Application of state-space methods to navigation problems," *Advances in Control Systems*, vol. 3, pp. 293–340, 1966.
- [5] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sum," *Automatica*, vol. 7, pp. 465–479, 1971.
- [6] R. S. Bucy and K. D. Senne, "Digital synthesis on nonlinear filters," *Automatica*, vol. 7, pp. 287–298, 1971.
- [7] N. Bergman, "Recursive Bayesian estimation: Navigation and tracking applications," Dissertations No 579, SE-581 83 Linkping, Sweden, May 1999.
- [8] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004.
- [9] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. New York: John Wiley & Sons, 2001.
- [10] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.
- [11] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," in *IEE Proceedings on Radar and Signal Processing*, vol. 140, 1993, pp. 107–113.
- [12] N. Metropolis and S. Ulam, "The Monte Carlo method," *Journal of the American Statistical Association*, vol. 44, no. 247, pp. 335–341, 1949.
- [13] P. Del Moral, *Feynman-Kac formulae: Genealogical and Interacting Particle Systems with Applications*, ser. Probability and Applications. New York, USA: Springer, 2004.
- [14] D. Crisan and A. Doucet, "A survey of convergence results on particle filtering methods for practitioners," *IEEE Transactions on Signal Processing*, vol. 50, no. 3, pp. 736–746, 2002.
- [15] P. Del Moral and L. Miclo, *Branching and Interacting Particle Systems Approximations of Feynman-Kac Formulae with Applications to Non-Linear Filtering*, ser. Lecture Notes in Mathematics. Berlin, Germany: Springer-Verlag, 2000, vol. 1729, pp. 1–145.

- [16] P. Del Moral, "Non-linear filtering: Interacting particle solution," *Markov processes and related fields*, vol. 2, no. 4, pp. 555–580, 1996.
- [17] F. Legland and N. Oudjane, "Stability and uniform approximation of nonlinear filters using the hilbert metric, and application to particle filters," INRIA, Paris, France, Tech. Rep. RR-4215, 2001.
- [18] P. Del Moral and A. Guionnet, "A central limit theorem for non linear filtering and interacting particle systems," *Annals of Applied Probability*, vol. 9, no. 2, pp. 275–297, 1999.
- [19] D. Crisan and M. Grunwald, "Large deviation comparison of branching algorithms versus resampling algorithms," Statist. lab. Cambridge University, Cambridge, United Kingdom, Tech. Rep. TR1999-9, 1998.
- [20] P. Del Moral and A. Guionnet, "Large deviations for interacting particle systems: Applications to non linear filtering problems," *Stochastic processes and their applications*, vol. 78, pp. 69–95, 1998.
- [21] T. B. Schön, "*Estimation of Nonlinear Dynamic Systems – Theory and Applications*," Dissertations No 998, Department of Electrical Engineering, Linköping University, Sweden, Feb. 2006.
- [22] H. Rosenthal, "On the subspaces of  $l^p$  ( $p > 2$ ) spanned by sequences of independent random variables," *Israel Journal of Mathematics*, vol. 8, no. 3, pp. 273–303, 1970.
- [23] A. Doucet, S. J. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.
- [24] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. New York, USA: Springer Verlag, 2001.
- [25] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, ser. Springer Series in Statistics. New York, USA: Springer, 2001.
- [26] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: particle filters for tracking applications*. London, UK: Artech House, 2004.
- [27] T. B. Schön, D. Törnqvist, and F. Gustafsson, "Fast particle filters for multi-rate sensors," in *15th European Signal Processing Conference (EUSIPCO)*, Poznań, Poland, Sep. 2007, accepted for publication.
- [28] G. Poyiadjis, A. Doucet, and S. S. Singh, "Maximum likelihood parameter estimation in general state-space models using particle methods," in *Proceedings of the American Statistical Association*, Minneapolis, USA, Aug. 2005.
- [29] H. R. Künsch, "Recursive monte carlo filters: algorithms and theoretical analysis," *The Annals of Statistics*, vol. 33, no. 5, pp. 1983–2021, 2005.
- [30] K. L. Chung, *A course in probability theory*, 2nd ed., ser. Probability and mathematical statistics. New York, USA: Academic Press, 1974, vol. 21.



**Lennart Ljung** received his Ph.D. in Automatic Control from Lund Institute of Technology in 1974. Since 1976 he is Professor of the chair of Automatic Control in Linköping, Sweden, and is presently Director of the Strategic Research Center MOVIII. He has held visiting positions at Stanford and MIT and has written several books on System Identification and Estimation. He is an IEEE Fellow, an IFAC Fellow and an IFAC Advisor as well as a member of the Royal Swedish Academy of Sciences (KVA), a member of the Royal Swedish Academy of Engineering Sciences (IVA), an Honorary Member of the Hungarian Academy of Engineering and a Foreign Associate of the US National Academy of Engineering (NAE). He has received honorary doctorates from the Baltic State Technical University in St Petersburg, from Uppsala University, Sweden, from the Technical University of Troyes, France, and from the Catholic University of Leuven, Belgium. In 2002 he received the Quazza Medal from IFAC, in 2003 the Hendrik W. Bode Lecture Prize from the IEEE Control Systems Society and he is the recipient of the IEEE Control Systems Award for 2007.



**Xiao-Li Hu** was born in Hunan, China in 1975. He received the B.S. degree in Mathematics from Hunan Normal University in 1997, and the M.Sc. degree in Applied Mathematics from Kunming University of Science and Technology in 2003, and the Ph.D. degree in the Key Laboratory of Systems and Control, Chinese Academy of Sciences in 2006. He visited the Division of Automatic Control, Department of Electrical Engineering, Linköping University, from Sep. 2006 to June 2007. He is now with the College of Science, China Jiliang University. His current research interests are system identification, filtering, stochastic approximation, least square algorithms and their applications.



**Thomas B. Schön** was born in Sweden in 1977. He received the B.Sc. degree in Business Administration and Economics in Feb. 2001, the M.Sc. degree in Applied Physics and Electrical Engineering in Feb. 2001 and the Ph.D. degree in Automatic Control in Feb. 2006, all from Linköping University, Linköping, Sweden. He has held visiting positions at the University of Cambridge (UK) and the University of Newcastle (Australia). His research interests are mainly within the areas of signal processing, sensor fusion and system identification, with applications to the automotive and aerospace industry. He is currently a Research Associate at Linköping University.