A Basic Convergence Result for Particle Filtering

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

N.B.: When citing this work, cite the original article.

Original publication:
http://dx.doi.org/10.1109/TSP.2007.911295.

Postprint available free at: 
Linköping University E-Press: http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-11748
A Basic Convergence Result for Particle Filtering

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

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.

Index Terms—Convergence of numerical methods, nonlinear estimation, particle filter, state estimation.

I. INTRODUCTION

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

\begin{align}
  x_{t+1} &= f_t(x_t, v_t) \tag{1a} \\
  y_t &= h_t(x_t, e_t) \tag{1b}
\end{align}

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):

\begin{align}
  x_{t+1} &= f_t(x_t) + v_t \tag{2a} \\
  y_t &= h_t(x_t) + e_t \tag{2b}
\end{align}

with \( v_t \) and \( e_t \) independent and identically distributed (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

\begin{equation}
  \mathbb{E}(\phi(x_t) | y_{1:t}) \tag{3}
\end{equation}

where \( y_{1:t} \triangleq (y_1, \ldots, y_t) \) and \( \phi : \mathbb{R}^{n_{yz}} \rightarrow \mathbb{R} \) 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 \( \phi(x_t) = x_t \), \( i = 1, \ldots, n_x \), where \( x_t = (x_t[1], \ldots, x_t[n_x])^T \).

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 (PFs) 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

\begin{equation}
  \hat{p}_N(x_t | y_{1:t}) = \sum_{i=1}^N w_t^i \delta_{x_t^i}(x_t) \tag{4}
\end{equation}

where each particle \( x_t^i \) has a weight \( w_t^i \) associated to it, and \( \delta_{x_t^i}(\cdot) \) denotes the delta-Dirac mass located in \( x_t \). 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 1940s [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] and [17]), central limit theorems (see also [18]) and large deviations (see also [19] and [20]). The previous results prove convergence of probability measures and only treat bounded functions \( \phi \), effectively excluding the most commonly

1053-587X/$25.00 © 2008 IEEE
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

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

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

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

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

$$P(X_{t+1} \in dx_{t+1}|X_t = x_t) = K(dx_{t+1}|x_t) = K(x_{t+1}|x_t)dx_{t+1}$$

(8a)

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

(8b)

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

1) **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:

$$K(x_{t+1}|x_t) = p_{v_t}(x_{t+1} - f_t(x_t))$$

(9a)

$$\rho(y_t|x_t) = p_{e_t}(y_t - h_t(x_t)).$$

(9b)

**B. Conceptual Solution**

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

$$\pi_{tt-1}(dx_t) = \int_{\mathbb{R}^{n_y}} \pi_{tt-1}(y_{t-1},dx_t)K(dx_t|y_{t-1})$$

$$\triangleq b_t(\pi_{tt-1})$$

(10a)

$$\pi_{tt}(dx_t) = \int_{\mathbb{R}^{n_y}} \rho(dy_t|x_t)\pi_{tt-1}(dx_t)$$

$$\triangleq a_t(\pi_{tt-1})$$

(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

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

(11)
Hence, $E(\phi(x_t)|y_{1:t}) = (\pi_{1:t}, \phi)$. Using this notation, by (10), for any function $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$, we have the following recursive form for the optimal filter $E(\phi(x_t)|y_{1:t})$:

$$
(\pi_{t-1}, \phi) = (\pi_{t-1}, K \phi)
$$

(12a)

and

$$
(\pi_{t}, \phi) = \left(\frac{\pi_{t-1} \rho_p}{\rho_p}\right).
$$

(12b)

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

$$
E(\phi(x_t)|y_{1:t}) = (\pi_{1:t}, \phi) = \frac{1}{\rho_p} \int \cdots \int \pi_0(x_0) K_1 \rho_1 \cdots K_t \rho_t \phi(x_t) dx_0 dt
$$

(13)

where $K_s \triangleq K(x_s|x_{s-1})$, $\rho_s \triangleq \rho(y_s|x_s)$, $s = 1, \ldots, t$, $dx_0 dt \triangleq \{dx_0, \ldots, dx_t\}$, 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]–[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

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

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

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

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

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:

$$
\hat{\pi}_{t-1|t-1}^N(dx_{t-1}) \triangleq \frac{1}{N} \sum_{i=1}^N \delta_{x_{t-1}^i}(dx_{t-1}).
$$

(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

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

(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 $\pi_{t|t}(dx_t)$:

$$
\hat{\pi}_{t|t}^N(dx_{t}) \triangleq \frac{\rho(y_t|x_{t})\hat{\pi}_{t|t-1}^N(dx_{t})}{\int_{\mathbb{R}^n} \rho(y_t|x_{t})\hat{\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)}.
$$

(16)

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

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

(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 $\hat{\pi}_{t|t}^N(dx_{t})$ the resample step will provide an equally weighted empirical distribution

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

(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:

\[
\tilde{x}_t^i \sim \sum_{j=1}^{N} \alpha_j^i K(dx_t|\tilde{x}_{t-1}^i) \tag{19}
\]

where 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

\[
\alpha^i = (\alpha_1^i, \alpha_2^i, \ldots, \alpha_N^i) \tag{20}
\]

where

\[
\alpha_j^i \geq 0, \quad \sum_{j=1}^{N} \alpha_j^i = 1, \quad \sum_{i=1}^{N} \alpha_j^i = 1. \tag{21}
\]

Clearly

\[
\frac{1}{N} \sum_{i=1}^{N} \alpha_j^i K(dx_t|\tilde{x}_{t-1}^i) = \frac{1}{N} \sum_{i=1}^{N} \left( \sum_{j=1}^{N} \alpha_j^i K(dx_t|\tilde{x}_{t-1}^i) \right) = \frac{1}{N} \sum_{i=1}^{N} K(dx_t|\tilde{x}_{t-1}^i) = (\pi_{t-1|t-1}, K). \tag{22}
\]

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]. 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

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

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

\[
w_t^i = \rho(y_t|\tilde{x}_t^i), \quad i = 1, \ldots, N,
\]

and normalize \( \tilde{w}_t^i = w_t^i / \sum_{j=1}^{N} w_t^j \).
4. Resample, \( x_t^i \sim \tilde{\pi}_t(dx_t), \quad i = 1, \ldots, N \) (\( \tilde{\pi} \) defined in (16)) \( \pi_t(dx_t) = (1/N) \sum_{i=1}^{N} \delta_{x_t^i}(dx_t) \).

\[
\pi_{t-1|t-1} \rightarrow \pi_{t|t-1} \rightarrow \pi_{t|t}
\]

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.

In Fig. 1 we provide a schematic illustration of the particle filter given in Algorithm 2. Let us now discuss the transformations of the involved probability measures a bit further, they are

\[
\pi_{t-1|t-1} \rightarrow \left[ \begin{array}{c} \delta_{x_{t-1}^1} \\ \vdots \\ \delta_{x_{t-1}^N} \end{array} \right] \rightarrow \left[ \begin{array}{c} \delta_{x_{t}^1} \\ \vdots \\ \delta_{x_{t}^N} \end{array} \right] \rightarrow \left[ \begin{array}{c} \sum_{j=1}^{N} \alpha_j^1 K(dx_t|\tilde{x}_t^1) \\ \sum_{j=1}^{N} \alpha_j^2 K(dx_t|\tilde{x}_t^2) \\ \vdots \\ \sum_{j=1}^{N} \alpha_j^N K(dx_t|\tilde{x}_t^N) \end{array} \right]
\]

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

\[
\pi_{t|t-1} = c(\nu) \sigma \Lambda \nu_b(\pi_{t-1|t-1})
\]

where \( c(\nu) \triangleq (1/N) [c_1 \ldots c_1] \) (Note that \( c_1 \) refers to a single sample.) and \( \sigma \) denotes composition of transformations in the form of a vector multiplication. Hence, we have

\[
\pi_{t|t} = c^N \circ \mu_t \circ c(\nu) \sigma \Lambda \nu_b(\pi_{t-1|t-1})
\]

where \( \circ \) denotes composition of transformations. Therefore

\[
\pi_{t|t} = c^N \circ \mu_t \circ c(\nu) \sigma \Lambda \nu_b \circ \cdots \circ c^N \circ c^{N-1} \circ \mu_1 \circ c(\nu) \sigma \Lambda \nu_b \circ c^N(\pi_0).
\]

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

\[
\pi_{t|t} = c^N \circ \mu_t \circ c^N \circ b_t(\pi_{t-1|t-1})
\]

\[
= c^N \circ \mu_t \circ c^N \circ b_t \circ \cdots \circ c^N \circ c^1 \circ c^1 \circ b_1 \circ c^N(\pi_0).
\]

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

\[
\pi_{t|t} = c_t \circ b_t \circ c^N(\pi_{t-1|t-1})
\]

rather than (25).
IV. 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

\[(\pi_{t|t-1}, \rho) > 0\]  \hspace{1cm} (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:

\[(\pi_{t|t-1}, \rho) \approx (\pi_{t|t-1}^N, \rho) = \int \rho(y_t|x_t) \frac{1}{N} \sum_{i=1}^{N} \delta_{\tilde{x}_t^i}(dx_t)\]

\[= \frac{1}{N} \sum_{i=1}^{N} \rho(y_t|\tilde{x}_t^i), \hspace{1cm} (28)\]

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:

\[(\pi_{t|t-1}^N, \rho) = \frac{1}{N} \sum_{i=1}^{N} \rho(y_t|\tilde{x}_t^i) \geq \gamma_t > 0. \hspace{1cm} (29)\]

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

\[\tilde{x}_t^i \sim \sum_{j=1}^{N} \alpha_{t}^j K(dx_t|x_t^i), \quad i = 1, \ldots, N.\]

3) If \((1/N) \sum_{i=1}^{N} \rho(y_t|x_t^i) \geq \gamma_t\), proceed to step 4 otherwise return to step 2.

4) Rename \(\tilde{x}_t^i \equiv \tilde{x}_t^j, i = 1, \ldots, N\) and compute the importance weights \(\{u_t^i\}_{i=1}^{N}\):

\[u_t^i = \rho(y_t|x_t^i), \quad i = 1, \ldots, N,\]

and normalize \(\tilde{u}_t = u_t^i / \sum_{j=1}^{N} u_t^j\).

5) Resample, \(x_t^i \sim \pi_{t|t}(dx_t) = \sum_{i=1}^{N} \tilde{u}_t^i \delta_{x_t^i}(dx_t), i = 1, \ldots, N.\)

6) Set \(t := t + 1\) and repeat from step 2.

For each time step, the filtering distribution is

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

The reason for renaming in step 4 is that the 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] and [17] for insightful discussions on this topic.

Clearly, the choice of \(\gamma_t\) may be nontrivial. If it is chosen too large (larger than the true conditional expectation), steps 2 and 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 \(\{x_t^i\}_{i=1}^{N}\) the joint density of \(\{x_t^i\}_{i=1}^{N}\) is

\[P[\tilde{x}_t^i = s_i, i = 1, \ldots, N] = \prod_{i=1}^{N} \sum_{j=1}^{N} \alpha_{t}^j K(s_i|x_t^i-1) = \Pi_{\alpha_{t}, \ldots, \alpha_{N}}^{N}. \hspace{1cm} (30)\]

Yet, after the modification it is changed to be

\[\Pi_{\alpha_{t}, \ldots, \alpha_{N}}^{N} I = \int \cdots \int \Pi_{\alpha_{t}, \ldots, \alpha_{N}}^{N} I \frac{\rho(y_t|s_i) \geq \gamma_t}{[1/(N) \sum_{i=1}^{N} \rho(y_t|s_i) \geq \gamma_t]} ds_{1:N}. \hspace{1cm} (31)\]

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:

\[x_{t+1} = \frac{x_t}{2} + \frac{2 \sqrt{x_t}}{1 + x_t^2} + 8 \cos(1.2t) + u_t \hspace{1cm} (32a)\]

\[y_t = \frac{x_t^2}{2a} + e_t \hspace{1cm} (32b)\]
where \( v_t \sim \mathcal{N}(0,10) \), \( \epsilon_t \sim \mathcal{N}(0, 1) \), 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 \( \hat{x}_t = \mathbb{E}(x_t | y_{1:t}) \). 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 is 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.

V. THE BASIC CONVERGENCE RESULT

The filtered state estimate is

\[
\hat{x}_t = \mathbb{E}(x_t | y_{1:t}).
\]

This is the mean of the conditional distribution

\[
\pi_{t|t}(dx_t) = P(X_t \in dx_t | y_{1:t} = y_{1:t}).
\]

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

\[
\hat{x}_t^N
\]

and

\[
\pi_{t|t}(dx_t).
\]

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 \( \pi_{t|t}(dx_t) \) 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

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

(37)

for functions of the posterior distribution. The notation introduced in (11) has been used in the first equality in (37). Note that the \( j \)th component of the estimate \( \hat{x}_t^N \) is obtained for \( \phi(x) = x[j] \) where \( x = (x[1], \ldots, x[n_x])^T \), \( i = 1, \ldots, n_x \). 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 fourth moment of the estimated conditional mean

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

(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

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

(39)

In particular, under certain conditions applying this result to the cases \( \phi(x) = x[j] \) where \( x = (x[1], \ldots, x[n_x])^T \), \( i = 1, \ldots, n_x \), 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. 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 $y_t$.

$H_0$: For given $y_{1:s}, s = 1, \ldots, t$, $(\pi_{s:s-1, \rho}) > 0$; and the constant $\gamma_s$ used in the modified algorithm satisfies

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

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 as follows (see H1).

$H_1$: $\rho(y_t|x_s) < \infty; K(x_s|x_{s-1}) < \infty$ for given $y_{1:s}, s = 1, \ldots, t$.

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 $\rho$ (see H2).

$H_2$: The function $\phi(\cdot)$ satisfies sup$_x |\phi(x)|^4 \rho(y_s|x_s) < C(y_{1:s})$ for given $y_{1:s}, s = 1, \ldots, t$.

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.

$$\int |\phi(x)|^4 \pi_{s:s-1}(dx) \leq C(y_{1:s}) \int |\pi_{s:s-1}(dx) < \infty.$$

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

Example 6.1: $p_k(z,s) = \text{O}(\text{exp}[-|z|^\nu])$ as $z \to \infty$ with $\nu > 0$; and $\lim_{|z| \to \infty} |b|/|z|^{\nu+1} > 0$ with $\nu > 0, s = 1, \ldots, t$. It is now easy to verify that H2 holds for any function $\phi$ satisfying $|\phi(z)| = \text{O}(|z|^\nu)$ as $z \to \infty$, where $\nu \geq 0$.

Example 6.2: $p_k(z,s) = (1/(b-a))I_{[a,b]}$ with $a < 0 < b$; and function $h(x,s) = h_{s}$ satisfying that the set $\Gamma_{s} = \{|y-b,y-1|\}$ is bounded for any given $y_s, s = 1, \ldots, t$. 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

$$L_4^4(\rho)$$

where $\rho$ satisfies H1.

1) 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:

$\text{i)}$ for sufficiently large $N$, the algorithm will not run into an infinite loop in steps 2–3;

$\text{ii})$ for any $\phi \in L_4^4(\rho)$, there exists a constant $C_{\text{Ht}}$, independent of $N$ such that

$$E \left| (\pi_{N, \text{Ht}}^N, \phi) - (\pi_{\text{Ht}}, \phi) \right|^4 \leq C_{\text{Ht}} \left| \phi \right|^4 N^{4/2}$$

where $\left| \phi \right|^4 \Delta = \max \left\{ 1, (\pi_{s:s-1, \rho})^{1/4}, s = 0, 1, \ldots, t \right\}$ and $\pi_{s:s}^N$ is generated by the algorithm.

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

2) Corollary 6.1: If $H_1$ and $H_2$ hold, then for any $\phi \in L_4^4(\rho)$,

$$\lim_{N \to \infty} (\pi_{N, \text{Ht}}^N, \phi) = (\pi_{\text{Ht}}, \phi),$$

almost surely. (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 $\{\xi_t, i = 1, \ldots, n\}$ be conditionally independent random variables given $\sigma$-algebra $G$ such that $E(\xi_t|G) = 0$, $E(\xi_t^4|G) < \infty$. Then

$$E \left( \sum_{i=1}^n \xi_i^4 \right) \leq \sum_{i=1}^n E(\xi_i^4|G) + \left( \sum_{i=1}^n E(\xi_i^2|G) \right)^2,$$

Proof: Notice that

$$\sum_{i=1}^n E(\xi_i^4|G) \leq \sum_{i=1}^n E(\xi_i^2|G)$$

the assertion follows.

Lemma 7.2: If $E(\xi^p) < \infty$, then $E(\xi - E(\xi|G))^p \leq 2^p E(\xi|G)^p$, for any $p \geq 1$.

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

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

which derives the desired inequality.

Lemma 7.3: If $1 \leq \gamma_1 \leq \gamma_2$ and $E(\xi^p) < \infty$, then $E(\xi^{2/p}) \leq E(\xi^{1/p})^2$. Then the assertion follows.

Proof: Simply by Hölder’s inequality (e.g., [30]):

$E(\xi^{*p}) \leq \gamma_1^{1/p} \left( (\xi^p)^{1/p} \right).$

Then the assertion follows.

Based on Lemmas 7.1 and 7.3, we have Lemma 7.4.

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

$$E \left( \frac{1}{n} \sum_{i=1}^n \xi_i^4 \right) \leq \frac{2 \max_1 \leq \leq \leq n E(\xi_i^4|G)}{n^2}.$$
Lemma 7.5: Let the probability density function for the random variable $\eta$ be $p(\eta)$ and let the probability density function for the random variable $\xi$ be

$$
\frac{p(\xi)p_A}{\int p(y)I_A dy}
$$

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

$$P[\eta \in \Omega - A] \leq \varepsilon < 1.
\label{eq:46}
$$

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

$$
|E\psi(\xi) - E\psi(\eta)| \leq \frac{2\sqrt{E|\psi|^2(\eta)}}{1-\varepsilon}\sqrt{\varepsilon},
\label{eq:47}
$$

In the case $E|\psi(\eta)| < \infty$

$$
E|\psi(\xi)| \leq \frac{E|\psi(\eta)|}{1-\varepsilon}.
\label{eq:48}
$$

Proof: Clearly, since the density of $\xi$ is

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

it is easy to show (48) as follows:

$$
E|\psi(\xi)| = \left| \int \psi(t)p(t)I_A dt \int p(y)I_A dy \right| \leq \frac{1}{1-\varepsilon} \int |\psi(t)p(t)dt = \frac{E|\psi(\eta)|}{1-\varepsilon}
$$

while

$$
|E\psi(\xi) - E\psi(\eta)| = \left| \int \psi(t)p(t)I_A dt \int p(y)I_A dy |- \int \psi(t)p(t) dt \right|
\leq \frac{1}{1-\varepsilon} \left| \int |\psi(t)p(t)I_A dt + \int |\psi(t)|p(t) dt \cdot \varepsilon \right|
\leq \frac{1}{1-\varepsilon} \left[ \int |\psi(t)|^2 p(t) dt \cdot \varepsilon + \int |\psi(t)|p(t) dt \cdot \varepsilon \right]
\leq \frac{1}{1-\varepsilon} \left[ \sqrt{E|\psi|^2(\eta)} \cdot \sqrt{\varepsilon} + E|\psi(\eta)| \cdot \varepsilon \right]
\leq \frac{2\sqrt{E|\psi|^2(\eta)}}{1-\varepsilon}\sqrt{\varepsilon},
$$

which derives (47).

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].

Initialization: Let $\{x_{i0}^N\}_{i=1}^{N}$ be independent random variables with the same distribution $\pi_0(dx_0)$. Then, using Lemmas 7.4 and 7.2, it is clear that

$$
E \left| (\pi_0^N, \phi) - (\pi_0, \phi) \right|^4 \leq \frac{1}{N^4} \sum_{i=1}^{N} \left(E\left(\phi(x_{i0}^N) - E(\phi(x_{i0}))\right)\right)^4
\leq \frac{2}{N^2} \left(E\left(\phi(x_{i0}^N) - E(\phi(x_{i0}))\right)^4 \right)
\leq \frac{32}{N^2} \left(\phi_{|\Omega_{i0}}^4 \right) \leq \frac{32}{N^2} \left(\phi_{|\Omega_{i0}}^4 \right)
= \frac{32C_{\phi_{|\Omega_{i0}}}^4}{N^2}.
\label{eq:49}
$$

Similarly

$$
E \left| (\pi_0^N, |\phi|^4) - (\pi_0, |\phi|^4) \right|
\leq \frac{1}{N} \sum_{i=1}^{N} \left(\phi_{|\Omega_{i0}}^4 - E(\phi_{|\Omega_{i0}}^4)\right)
\leq 2E(\phi_{|\Omega_{i0}}^4).
\label{eq:50}
$$

Note that $x_{i0}$ have the same distribution for all $i$, so the expected values do not depend on $i$. Hence

$$
E \left| (\pi_0^N, |\phi|^4) \right| \leq 3E(\phi_{|\Omega_{i0}}^4) \leq M_{\phi_{|\Omega_{i0}}} |\phi|^4.
\label{eq:51}
$$

Prediction: Based on (49) and (50), we assume that for $t-1$ and $\forall \phi \in L_2(\rho)$

$$
E \left| (\pi_{t-1|t-1}, \phi) - (\pi_{t-1|t-1}, \phi) \right|^4 \leq C_{t-1|t-1} \frac{|\phi|_{L_{\rho}^2}^4}{N^2}
\label{eq:52}
$$

and

$$
E \left| (\pi_{t-1|t-1}, |\phi|^4) \right| \leq M_{t-1|t-1} |\phi|_{L_{\rho}^4}^4
\label{eq:53}
$$

holds, where $C_{t-1|t-1} > 0$ and $M_{t-1|t-1} > 0$. We analyze

$$
E \left| (\hat{\pi}_{t-1|t-1}, \phi) - (\pi_{t-1|t-1}, \phi) \right|^4
\text{ and } E \left| (\hat{\pi}_{t-1|t-1}, |\phi|^4) \right|
$$

in this step.

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

$$
(\hat{\pi}_{t-1|t-1}, \phi) = (\hat{\pi}_{t-1|t-1}, \phi) \oplus \Pi_1 + \Pi_2 + \Pi_3
\label{eq:54}
$$

where

$$
\Pi_1 \equiv (\hat{\pi}_{t-1|t-1}, \phi) - \frac{1}{N} \sum_{i=1}^{N} E \left[ (\phi(x_{t-1,i}) \mid \mathcal{F}_{t-1} \right]
\Pi_2 \equiv \frac{1}{N} \sum_{i=1}^{N} E \left[ (\phi(x_{t-1,i}) \mid \mathcal{F}_{t-1} \right] - \frac{1}{N} \sum_{i=1}^{N} (\pi_{t-1|t-1}, K_{t-1})
\Pi_3 \equiv \frac{1}{N} \sum_{i=1}^{N} (\pi_{t-1|t-1}, K_{t-1}K_{t-1} - (\pi_{t-1|t-1}, K_{t-1})
\label{eq:55}
$$

and $\pi_{t-1|t-1} = \sum_{i=1}^{N} \alpha_i \nu_{\pi_{t-1|t-1}}(dx_{t-1})$. We consider the three terms $\Pi_1, \Pi_2$ and $\Pi_3$ separately in the following.
Let $\hat{x}_t$ be drawn from the distribution $(\pi_{t-|t-1}, K)$ as in step 2 of the algorithm. Then we have

$$E[\phi(\hat{x}_t)]_{[F_{t-1}]} = (\pi_{t-|t-1}, K)\rho).$$

(53)

Recall that the distribution of $\hat{x}_t$ differs from the distribution of $\hat{x}_t$, which has passed the test in step 3 of the algorithm and is thus conditioned on the event

$$A_t = \{((\pi_{t-|t-1}, K)\rho) \geq \gamma_t\}.$$

(54)

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

$$E \left[ \frac{1}{N} \sum_{i=1}^{N} \rho(y_i | \hat{x}_t) \right]_{[F_{t-1}]} = (\pi_{t-|t-1}, K)\rho).$$

Thus,

$$P \left[ \frac{1}{N} \sum_{i=1}^{N} \rho(y_i | \hat{x}_t) < \gamma_t \right]_{[F_{t-1}]} = P \left[ (\pi_{t-|t-1}, K)\rho) < \gamma_t \right].$$

(55)

By (51), we have

$$P \left[ (\pi_{t-|t-1}, K)\rho) < \gamma_t \right]$$

$$= P \left[ (\pi_{t-|t-1}, K)\rho) - (\pi_{t-|t-1}, K)\rho) < \gamma_t - (\pi_{t-|t-1}, K)\rho) \right]$$

$$\leq P \left[ (\pi_{t-|t-1}, K)\rho) - (\pi_{t-|t-1}, K)\rho) > \gamma_t - (\pi_{t-|t-1}, K)\rho) \right]$$

$$\leq \frac{E[\phi(\hat{x}_t)]_{[F_{t-1}]} - (\pi_{t-|t-1}, K)\rho)]^4}{\gamma_t - (\pi_{t-|t-1}, K)\rho)^4} \leq C_{\gamma_t} \frac{||\phi||_{1-1/4}^4}{N^2},$$

(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 $\hat{x}_t$ and $\hat{x}_t$ using Lemma 7.5, and by Lemmas 7.1, 7.2, (53) and (22), we obtain

$$E \left[ \Pi_4 \right]_{[F_{t-1}]}$$

$$= \frac{1}{N^4} E \left[ \sum_{i=1}^{N} \phi(\hat{x}_t) - E[\phi(\hat{x}_t)]_{[F_{t-1}]} \right]_{[F_{t-1}]}$$

$$\leq \frac{2^4}{N^4} \left[ \sum_{i=1}^{N} E \left[ \phi(\hat{x}_t)^4 \right]_{[F_{t-1}]} + \left( \sum_{i=1}^{N} E \left[ \phi(\hat{x}_t)^2 \right]_{[F_{t-1}]} \right)^{2} \right]$$

$$\leq \frac{2^4}{N^4(1-\epsilon_t)^2} \left[ \sum_{i=1}^{N} E \left[ \phi(\hat{x}_t)^4 \right]_{[F_{t-1}]} + \left( \sum_{i=1}^{N} E \left[ \phi(\hat{x}_t)^2 \right]_{[F_{t-1}]} \right)^{2} \right].$$

(57)

Hence, by Lemma 7.3 and (52)

$$E \left[ \Pi_1 \right]_{[F_{t-1}]} \leq \frac{2^4 ||K||^4 M_{t-|t-1}|}{N^3} \Delta \leq C_{\Pi_1} \frac{||K||^4}{N^2}.$$

By (53), Lemma 7.5 and (22)

$$E \left[ \Pi_2 \right]_{[F_{t-1}]} = \frac{1}{N} \sum_{i=1}^{N} E \left[ \phi(\hat{x}_t) \right]_{[F_{t-1}]} - \frac{1}{N} \sum_{i=1}^{N} E \left[ \phi(\hat{x}_t) \right]_{[F_{t-1}]}$$

$$\leq \frac{2^4 C_2 ||\rho||_{1-1/4}^4}{(1-\epsilon_t)^4 N^4} \sum_{i=1}^{N} (\pi_{t-|t-1}, K)\rho)$$

$$\leq \frac{2^4 C_2 ||\rho||_{1-1/4}^4}{(1-\epsilon_t)^4 N^4} (\pi_{t-|t-1}, K)\rho)$$

$$\leq C_{\Pi_2} \frac{(\pi_{t-|t-1}, K)\rho)}{N^4},$$

(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 and 3. By (22) and (51)

$$E \left[ \Pi_3 \right]_{[F_{t-1}]} \leq C_{\Pi_3} \frac{||K||^4}{N^2} \leq C_{\Pi_3} \frac{||\rho||_{1-1/4}^4}{N^2},$$

(59)

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

$$E \left[ \Pi_3 \right]_{[F_{t-1}]} \leq C_{\Pi_3} \frac{||K||^4}{N^2} \leq C_{\Pi_3} \frac{||\rho||_{1-1/4}^4}{N^2},$$

(60)
By Lemma 7.2 and (52)
\[
E \left( E \left( \frac{\hat{\pi}_{\ell|t-1}^N}{N} \mid |\phi|^4 \right) - \frac{1}{N} \sum_{i=1}^{N} (\hat{\pi}_{\ell|t-1}^N |\phi|^4) |F_{t-1} \right) \\
= \frac{1}{N} E \left( E \left( \sum_{i=1}^{N} (\hat{\phi}(\hat{\pi}_{\ell|t-1}^N) - E(\hat{\phi}(\hat{\pi}_{\ell|t-1}^N) |F_{t-1}) \right) |F_{t-1} \right) \\
\leq 2E(\pi_{\ell|t-1}^N |K| |\phi|^4) \leq 2 ||K|^4 M_{t-1} ||\phi||^4_{1,4}.
\]

Then, using a similar separation mentioned above, by (52) we have
\[
E \left( \frac{\hat{\pi}_{\ell|t-1}^N}{N} |\phi|^4 \right) - (\pi_{\ell|t-1}^N |\phi|^4) \\
\leq ||K|^4 (3M_{t-1} + 1) ||\phi||^4_{1,4} \leq \hat{M}_{t-1} ||\phi||^4_{1,4}.
\]

**Update:** In this step we go one step further to analyze
\[
E \left( \frac{\hat{\pi}_{\ell|t}^N |\phi|^4}{N} \right) \quad \text{and} \quad E(\hat{\pi}_{\ell|t}^N |\phi|^4)
\]
based on (60) and (61). Clearly,
\[
(\pi_{\ell|t}^N |\phi|^4) - (\pi_{\ell|t}^N |\phi|^4) = (\pi_{\ell|t}^N |\phi|^4) - (\pi_{\ell|t}^N |\phi|^4) = \hat{\Pi}_1 + \hat{\Pi}_2
\]

where
\[
\hat{\Pi}_1 \triangleq \frac{\hat{\pi}_{\ell|t}^N |\phi|^4}{\pi_{\ell|t}^N |\phi|^4}
\]
and
\[
\hat{\Pi}_2 \triangleq \frac{\hat{\pi}_{\ell|t}^N |\phi|^4}{\pi_{\ell|t}^N |\phi|^4}.
\]

By condition H1 and the modified version of the algorithm we have
\[
\hat{\Pi}_1 = \frac{\hat{\pi}_{\ell|t}^N |\phi|^4}{\pi_{\ell|t}^N |\phi|^4} \cdot \frac{[\pi_{\ell|t}^N |\phi|^4] - (\pi_{\ell|t}^N |\phi|^4)}{\pi_{\ell|t}^N |\phi|^4}
\leq \frac{||\phi||^4}{\gamma_{t} |\pi_{\ell|t}^N |\phi|^4} \cdot \frac{[\pi_{\ell|t}^N |\phi|^4] - (\pi_{\ell|t}^N |\phi|^4)}{\pi_{\ell|t}^N |\phi|^4}.
\]

Here, \( \gamma_t \) is the threshold used in step 3 of the modified filter (Algorithm 3). Thus, by Minkowski’s inequality, (60) and (62),
\[
E(\phi(\hat{x}_t)) |G_t) = (\pi_{\ell|t}^N |\phi|^4).
\]

and then
\[
\hat{\Pi}_1 = \frac{1}{N} \sum_{i=1}^{N} (\phi(\hat{x}_t^i) - E(\phi(\hat{x}_t^i) |G_t)).
\]

Then, by Lemmas 7.4, 7.2,
\[
E \left( \frac{\hat{\pi}_{\ell|t}^N |\phi|^4}{N} \right) = \frac{1}{N^2} E \left( \sum_{i=1}^{N} (\phi(\hat{x}_t^i) - E(\phi(\hat{x}_t^i) |G_t)) \right) |G_t) \\
\leq 2 E(\phi(\hat{x}_t^i) |G_t) \leq 2 \left( \frac{\hat{M}_{t-1} ||\phi||^4_{1,4}}{N^2} \right).
\]

Thus, by (64),
\[
E|\hat{\Pi}_1|_t^4 \leq 2 \hat{M}_{t-1} \frac{||\phi||^4_{1,4}}{N^2}.
\]
Using Minkowski’s inequality, (63) and (65) we have
\[ E^{1/4} \left| \left( \pi \nabla_{\mathbf{q}_t} \phi \right) - \left( \pi \nabla_{\hat{\mathbf{q}}_t} \phi \right) \right|^4 \leq E^{1/4} \left| \nabla_{\mathbf{e}_1} \right|^4 + E^{1/4} \left| \nabla_{\mathbf{e}_2} \right|^4 \
\leq \left( \frac{2^5 \nabla_{\mathbf{q}_t}^{4/4} + \mathcal{C}^{4/4}}{\mathcal{q}_t} \right) \frac{\left\| \phi \right\|_{L^4}^4}{N^{1/2}} \]
that is
\[ E \left[ \left( \pi \nabla_{\mathbf{q}_t} \phi \right)^4 - \left( \pi \nabla_{\hat{\mathbf{q}}_t} \phi \right)^4 \right] \leq \mathcal{C}_4 \frac{\left\| \phi \right\|_{L^4}^4}{N^{3/2}}. \tag{66} \]
Using a separation similar to the one mentioned above, by (64), we have
\[ E \left[ \left( \pi \nabla_{\mathbf{q}_t} \phi \right)^4 - \left( \pi \nabla_{\hat{\mathbf{q}}_t} \phi \right)^4 \right] \leq \left( \frac{2^5 \nabla_{\mathbf{q}_t}^{4/4} + \mathcal{C}^{4/4}}{\mathcal{q}_t} \right) \frac{\left\| \phi \right\|_{L^4}^4}{N^{1/2}} \]
\[ \leq \left[ 2 \nabla_{\mathbf{q}_t} + (3 \nabla_{\hat{\mathbf{q}}_t} + 1) \right] \left\| \phi \right\|_{L^4}^4 \]
\[ \leq (3 \nabla_{\mathbf{q}_t} + 1) \left\| \phi \right\|_{L^4}^4. \]
Hence
\[ E \left[ \left( \pi \nabla_{\mathbf{q}_t} \phi \right)^4 \right] \leq (3 \nabla_{\mathbf{q}_t} + 2) \left\| \phi \right\|_{L^4}^4 \triangleq \mathcal{M}_4 \left\| \phi \right\|_{L^4}^4. \tag{67} \]
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 practice, 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].

ACKNOWLEDGMENT

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


Xiao-Li Hu was born in Hunan, China, in 1975. He received the B.S. degree in mathematics from Hunan Normal University in 1997, the M.Sc. degree in applied mathematics from Kunming University of Science and Technology in 2003, and the Ph.D. degree from 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, Linköping, Sweden, from September 2006 to June 2007. He is currently with the College of Science, China Jiliang University, Hangzhou, China. His current research interests are system identification, filtering, stochastic approximation and least square algorithms and their applications.

Thomas B. Schön (M’07) was born in Sweden in 1977. He received the B.Sc. degree in business administration and economics in 2001, the M.Sc. degree in applied physics and electrical engineering in 2001, and the Ph.D. degree in automatic control in 2006, all from Linköping University, Linköping, Sweden. He has held visiting positions at the University of Cambridge, Cambridge, U.K., and the University of Newcastle, Newcastle, Australia. He is currently a Research Associate at Linköping University, Linköping, Sweden. His research interests are mainly within the areas of signal processing, sensor fusion, and system identification, with applications to the automotive and aerospace industry.

Lennart Ljung (S’74–M’75–SM’83–F’85) received the Ph.D. degree in automatic control from the Lund Institute of Technology in 1974. Since 1976, he has been a Professor of the Chair of Automatic Control in Linköping University, Linköping, Sweden, and is presently Director of the Strategic Research Center MOVIII. He has held visiting positions at Stanford University, Stanford, CA, and the Massachusetts Institute of Technology (MIT), Cambridge. He has written several books on system identification and estimation.

Dr. Ljung is an IFAC Fellow and an IFAC Advisor as well as a member of the Royal Swedish Academy of Sciences (KVA), an Honorary Member of the Hungarian Academy of Engineering, and a Foreign Associate of the U.S. National Academy of Engineering (NAE). He has received honorary doctorates from the Baltic State Technical University, 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 and 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.