

A General Convergence Result for Particle Filtering

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

Abstract—The particle filter has become an important tool in solving nonlinear filtering problems for dynamic systems. This correspondence extends our recent work, where we proved that the particle filter converges for unbounded functions, using L^4 -convergence. More specifically, the present contribution is that we prove that the particle filter converge for unbounded functions in the sense of L^p -convergence, for an arbitrary $p \geq 2$.

I. INTRODUCTION

The main purpose of the present work is to extend our previous results on particle filtering convergence for unbounded functions [1], where we, for simplicity, only proved L^4 -convergence. Here, we will prove L^p -convergence for an arbitrary $p \geq 2$, of the particle filter. This requires some nontrivial embellishments, which form the contribution of the present work, including the introduction and use of a new Rosenthal-type inequality [2].

The particle filter provides a solution to the nonlinear filtering problem, which amounts to, recursively in time computing an estimate of the state in a dynamic system,

$$\begin{aligned} x_{t+1} &= f_t(x_t, v_t), \\ y_t &= h_t(x_t, e_t). \end{aligned} \quad (1a) \quad (1b)$$

Here, x_t denotes the state, y_t denotes the measurement, v_t and e_t denote the stochastic process and measurement noise, respectively. Most estimation algorithms aim at computing an approximation of the conditional expectation

$$E(\phi(x_t)|y_{1:t}) = \int \phi(x_t)p(x_t|y_{1:t})dx_t, \quad (2)$$

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. The particle filter computes an approximation to (2) by forming an approximation of the filtering distribution according to

$$\hat{p}_N(x_t|y_{1:t}) = \sum_{i=1}^N w_t^i \delta_{x_t^i}(dx_t), \quad (3)$$

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 .

The first complete particle filter was introduced by Gordon *et al.* in 1993 [3]. Since then the particle filter has become an important tool in solving complicated estimation problems. For more information about the particle filter we refer to the text books [4]–[6] and the survey papers [6]–[10]. When it comes to convergence results for the particle filter the book [11] contains a lot of useful results. Furthermore, the excellent survey papers [12], [13] are very informative.

The outline of the paper is as follows. In Section II we briefly introduce the models, the optimal filters that we are trying to approximate and the particle filter. However, these sections are intentionally rather brief, since a more detailed background using the same notation is

X-L. Hu is with the School of Electrical Engineering and Computer Science, The University of Newcastle, Newcastle NSW 2308, Australia, e-mail: xiaoli.hu@newcastle.edu.au, xlhu@amss.ac.cn, Phone: +61 2 49215921

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

already provided in [1] and the related technical report [20]. The main result is then presented and proved in Section III and the conclusions are given in Section IV. There is also an appendix containing the necessary auxiliary lemmas.

II. BACKGROUND

In order to understand the general convergence result proved in the present work we will here briefly explain the background when it comes to models and optimal filters in Section II-A and the particle filter in Section II-B.

A. Models and Optimal Filters

In order to develop the theory below we need to represent the nonlinear system (1) in a way that facilitates the use of the relevant theoretical tools. We are concerned with two real vector-valued stochastic processes $X = \{X_t\}_{t=1}^N$ and $Y = \{Y_t\}_{t=1}^N$, which are defined on a probability space. The n_x -dimensional process X describes the evolution of the hidden state and it is a Markov process with initial state X_0 and an initial distribution $\pi_0(dx_0)$. Furthermore, a Markov transition kernel $K(dx_{t+1}|x_t)$ is used to model the state evolution over time according to

$$P(X_{t+1} \in A | X_t = x_t) = \int_A K(dx_{t+1}|x_t), \quad (4)$$

for all $A \in \mathcal{B}(\mathbb{R}^{n_x})$, where $\mathcal{B}(\mathbb{R}^{n_x})$ denotes the Borel σ -algebra on \mathbb{R}^{n_x} . The n_y -dimensional process Y describes the available measurements, which are assumed conditionally independent given the states and

$$P(Y_t \in B | X_t = x_t) = \int_B \rho(dy_t|x_t), \quad \forall B \in \mathcal{B}(\mathbb{R}^{n_y}). \quad (5)$$

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

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

$$P(Y_t \in dy_t | X_t = x_t) = \rho(y_t|x_t)dy_t. \quad (6b)$$

Since we are trying to approximate (2) we are indirectly interested in finding approximations of the filtering distribution, i.e., the distribution of the state conditioned on the measurements $\pi_{t|t}(dx_t)$ which is ideally given by

$$\pi_{t|t-1}(dx_t) = \int_{\mathbb{R}^{n_x}} \pi_{t-1|t-1}(dx_{t-1})K(dx_t|x_{t-1}), \quad (7a)$$

$$\pi_{t|t}(dx_t) = \frac{\rho(y_t|x_t)\pi_{t|t-1}(dx_t)}{\int_{\mathbb{R}^{n_x}} \rho(y_t|x_t)\pi_{t|t-1}(dx_t)}. \quad (7b)$$

In the interest of a more compact notation, let us introduce the following. Given a measure ν , a function ϕ , and a Markov transition kernel K , denote

$$(\nu, \phi) \triangleq \int \phi(x)\nu(dx), \quad K\phi(x) = \int K(dz|x)\phi(z). \quad (8)$$

This implies that $E(\phi(x_t)|y_{1:t}) = (\pi_{t|t}, \phi)$. From (7) we now have the following recursive form for the optimal filter $E(\phi(x_t)|y_{1:t})$,

$$(\pi_{t|t-1}, \phi) = (\pi_{t-1|t-1}, K\phi), \quad (9a)$$

$$(\pi_{t|t}, \phi) = \frac{(\pi_{t|t-1}, \phi\rho)}{(\pi_{t|t-1}, \rho)}. \quad (9b)$$

B. Particle Filters

The particle filter we are concerned with in this work is given in detail in Algorithm 1 below.

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 samples,

$$\bar{x}_t^i \sim \sum_{j=1}^N \alpha_j^i K(dx_t | x_{t-1}^j), \quad i = 1, \dots, N.$$
- 3) If $\frac{1}{N} \sum_{i=1}^N \rho(y_t | \bar{x}_t^i) \geq \gamma_t$, proceed to step 4 otherwise return to step 2.
- 4) Rename $\bar{x}_t^i = \tilde{x}_t^i$, compute $w_t^i = \rho(y_t | \tilde{x}_t^i)$ and normalize $\tilde{w}_t^i = w_t^i / \sum_{j=1}^N w_t^j$ for $i = 1, \dots, N$.
- 5) Resample, $x_t^i \sim \tilde{\pi}_{t|t}^N(dx_t) = \sum_{i=1}^N \tilde{w}_t^i \delta_{\tilde{x}_t^i}(dx_t)$, $i = 1, \dots, N$.
- 6) Set $t := t + 1$ and repeat from step 2.

The particle filtering algorithm given above is different from the standard particle filter in two ways. The first difference is that we have, in step (2), introduced the weights α_j^i , satisfying

$$\alpha_j^i \geq 0, \quad \sum_{j=1}^N \alpha_j^i = 1, \quad \sum_{i=1}^N \alpha_j^i = 1. \quad (10)$$

These weights allows us to represent two slightly different particle filters at once. More specifically, when $\alpha_j^i = 1$ for $j = i$, and $\alpha_j^i = 0$ for $j \neq i$, the sampling method is reduced to the original particle filter introduced by [3], see also e.g., [6], [14]. On the other hand, when $\alpha_j^i = 1/N$ for all i and j , it turns out to be a convenient form for theoretical treatment, as used by nearly all existing theoretical analysis, see e.g., [11]–[13], [15]. Let us also point out a useful formula for future use. In step (2), when sampling \bar{x}_t^i from the distribution $\sum_{j=1}^N \alpha_j^i K(dx_t | x_{t-1}^j)$, we have

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \alpha_j^i K(dx_t | x_{t-1}^j) &= \frac{1}{N} \sum_{j=1}^N \left(\sum_{i=1}^N \alpha_j^i K(dx_t | x_{t-1}^j) \right) \\ &= \frac{1}{N} \sum_{j=1}^N K(dx_t | x_{t-1}^j) = (\pi_{t-1|t-1}^N, K). \end{aligned} \quad (11)$$

The second difference worth commenting is that we in step (3) require that the sampled particles $\{\bar{x}_t^i\}_{i=1}^N$ satisfies

$$\frac{1}{N} \sum_{i=1}^N \rho(y_t | \bar{x}_t^i) \geq \gamma_t > 0, \quad (12)$$

where the real number γ_t is selected by experience. If the above inequality holds, the algorithm proceeds to the next step, whereas if it does not hold, we regenerate $\{\bar{x}_t^i\}_{i=1}^N$ again until (12) is satisfied. After renaming $\{\bar{x}_t^i\}_{i=1}^N$ by $\{\tilde{x}_t^i\}_{i=1}^N$, the requirement is

$$(\tilde{\pi}_{t|t-1}^N, \rho) = \frac{1}{N} \sum_{i=1}^N \rho(y_t | \tilde{x}_t^i) \geq \gamma_t > 0. \quad (13)$$

The requirement is used in the proof of the main results of this paper. Furthermore, from the more practical side, it helps in reducing the risk of filter divergence.

III. GENERAL CONVERGENCE RESULT

In this section we consider convergence of the particle filter, Algorithm 1, to the optimal filter

$$E(\phi(x_t) | y_{1:t}) \quad (14)$$

in the case where ϕ is an unbounded function. It is also worth noting that all the stochastic quantifiers below (like E and “w.p. 1”) are with respect to the random variables related to the particles. Below we list the conditions that we need in order to establish the convergence result.

H0. For given $y_{1:s}$, $s = 1, 2, \dots, t$, $(\pi_{s|s-1}, \rho) > 0$, and the constant γ_s used in the algorithm satisfies $0 < \gamma_s < (\pi_{s|s-1}, \rho)$, $s = 1, 2, \dots, t$.

H1. $\rho(y_s | x_s) < \infty$; $K(x_s | x_{s-1}) < \infty$ for given $y_{1:s}$, $s = 1, 2, \dots, t$.

H2. For some $p > 1$, the function $\phi(\cdot)$ satisfies $\sup_{x_s} |\phi(x_s)|^p \rho(y_s | x_s) < C(y_{1:s})$ for given $y_{1:s}$, $s = 1, \dots, t$.

Let us denote the set of functions ϕ satisfying H2 by $L_t^p(\rho)$. Denote the maximum norm $\|\phi(x)\| = \max_x |\phi(x)|$ for any bounded function of $x = (x_1, \dots, x_t)$ with respect to fixed y_1, \dots, y_t . For example, we have $\|\rho\| < \infty$ and $\|K\| < \infty$ by H1, and $\|\phi^p \rho\| < \infty$ by H2.

Remark 3.1: Based on (9b) we see that $(\pi_{s|s-1}, \rho) > 0$ in H0 is a basic requirement for the optimal filter $E(\phi(x_t) | y_{1:t})$ to exist.

Remark 3.2: By the conditions $(\pi_{s|s-1}, \rho) > 0$ and $\sup_{x_s} |\phi(x_s)|^p \rho(y_s | x_s) < \infty$, we have

$$(\pi_{s|s}, |\phi|^p) = \frac{(\pi_{s|s-1}, \rho |\phi|^p)}{(\pi_{s|s-1}, \rho)} < \infty. \quad (15)$$

Theorem 3.1: If H0-H2 hold, then for any $\phi \in L_t^p(\rho)$ and $p \geq 2$, $1 \leq r \leq 2$, and sufficiently large N , there exists a constant $C_{t|t}$ independent of N such that

$$E \left| (\pi_{t|t}^N, \phi) - (\pi_{t|t}, \phi) \right|^p \leq C_{t|t} \frac{\|\phi\|_{t,p}^p}{N^{p-p/r}}, \quad (16)$$

where $\|\phi\|_{t,p} \triangleq \max \left\{ 1, (\pi_{s|s}, |\phi|^p)^{1/p}, s = 0, 1, \dots, t \right\}$.

Proof. The proof is carried out using an induction framework, similar to the one introduced in [12] and further used in [1].

1: Initialization Let $\{x_0^i\}_{i=1}^N$ be independent random variables from the distribution $\pi_0(dx_0)$. Then, with the use of Lemmas A.1, A.2 and A.3 (note that A here implies that the lemmas are to be found in the Appendix) we obtain

$$\begin{aligned} E \left| (\pi_0^N, \phi) - (\pi_0, \phi) \right|^p &= \frac{1}{N^p} E \left| \sum_{i=1}^N (\phi(x_0^i) - E[\phi(x_0^i)]) \right|^p \\ &\leq \frac{C(p)}{N^p} \left[\sum_{i=1}^N E |\phi(x_0^i) - E[\phi(x_0^i)]|^p \right. \\ &\quad \left. + \left[\sum_{i=1}^N E |\phi(x_0^i) - E[\phi(x_0^i)]|^r \right]^{p/r} \right] \\ &\leq 2^p C(p) \left[\frac{E |\phi(x_0^i)|^p}{N^{p-1}} + \frac{E^{p/r} |\phi(x_0^i)|^r}{N^{p(1-1/r)}} \right] \\ &\leq 2^{p+1} C(p) \frac{E |\phi(x_0^i)|^p}{N^{p(1-1/r)}} \triangleq C_{0|0} \frac{\|\phi\|_{0,p}^p}{N^{p(1-1/r)}}. \end{aligned} \quad (17)$$

Note that in the last two inequalities i refers to an arbitrary $i = 1, \dots, N$. Similarly,

$$\begin{aligned} E \left| (\pi_0^N, |\phi|^p) - (\pi_0, |\phi|^p) \right| &\leq \frac{1}{N} E \left| \sum_{i=1}^N (|\phi(x_0^i)|^p - E |\phi(x_0^i)|^p) \right| \\ &\leq 2E |\phi(x_0^i)|^p. \end{aligned} \quad (18)$$

Hence,

$$E \left| (\pi_0^N, |\phi|^p) \right| \leq 3E |\phi(x_0^i)|^p \triangleq M_{0|0} \|\phi\|_{0,p}^p. \quad (19)$$

2: Prediction Based on (17) and (19), we assume that for $t-1$ and $\forall \phi \in L_t^p(\rho)$

$$E \left| (\pi_{t-1|t-1}^N, \phi) - (\pi_{t-1|t-1}, \phi) \right|^p \leq C_{t-1|t-1} \frac{\|\phi\|_{t-1,p}^p}{N^{p(1-1/r)}} \quad (20)$$

and

$$E \left| (\pi_{t-1|t-1}^N, |\phi|^p) \right| \leq M_{t-1|t-1} \|\phi\|_{t-1,p}^p \quad (21)$$

hold for sufficiently large N , where $C_{t-1|t-1} > 0$ and $M_{t-1|t-1} > 0$. In this step we analyze $E \left| (\tilde{\pi}_{t-1|t-1}^N, \phi) - (\pi_{t-1|t-1}, \phi) \right|^p$ and $E \left| (\tilde{\pi}_{t-1|t-1}^N, |\phi|^p) \right|$.

Proposition 3.1 given below shows that the modified algorithm will not run into an infinite loop. Let \mathcal{F}_{t-1} denote the σ -algebra generated by $\{x_{i-1}^i\}_{i=1}^N$. Notice that

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

where

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

and $\pi_{t-1|t-1}^{N,\alpha_i} = \sum_{j=1}^N \alpha_j^i \delta_{x_{t-1}^j}$. Below we will consider the three terms Π_1 , Π_2 and Π_3 separately, but first we point out some basic facts which are needed in the analysis. Let $\{x_{t-1}^i\}_{i=1}^N$ and y_t be given, then we know from Algorithm 1 that \tilde{x}_t^i obeys $(\pi_{t-1|t-1}^{N,\alpha_i}, K)$, $i = 1, \dots, N$,

$$E[\phi(\tilde{x}_t^i) | \mathcal{F}_{t-1}] = \sum_{j=1}^N \alpha_j^i K \phi(x_{t-1}^j) = (\pi_{t-1|t-1}^{N,\alpha_i}, K\phi). \quad (22)$$

Based on (22) and (11), we have

$$\begin{aligned} E \left(\frac{1}{N} \sum_{i=1}^N \rho(y_t | \tilde{x}_t^i) | \mathcal{F}_{t-1} \right) &= \frac{1}{N} \sum_{i=1}^N (\pi_{t-1|t-1}^{N,\alpha_i}, K\rho) \\ &= (\pi_{t-1|t-1}^N, K\rho). \end{aligned} \quad (23)$$

Note that $\{\tilde{x}_t^i, i = 1, \dots, N\}$ are particles generated without any modification and $\{x_t^i, i = 1, \dots, N\}$ the modified particles by (12). The term Π_2 denotes the difference between these two series of particles. Lemma A.5 can now be used to analyze the terms Π_1 and Π_2 introduced above, since (40) of Proposition 3.1,

$$P \left[\frac{1}{N} \sum_{i=1}^N \rho(y_t | \tilde{x}_t^i) < \gamma_t \right] < \epsilon_t < 1 \quad (24)$$

holds for sufficiently large N .

By Lemmas A.1, A.2, A.5 (conditional case), (22) and (11),

$$\begin{aligned} E(|\Pi_1|^p | \mathcal{F}_{t-1}) &= \frac{1}{N^p} E \left(\left| \sum_{i=1}^N [\phi(\tilde{x}_t^i) - E(\phi(\tilde{x}_t^i) | \mathcal{F}_{t-1})] \right|^p \middle| \mathcal{F}_{t-1} \right) \\ &\leq \frac{2^p C(p)}{N^p} \left[\sum_{i=1}^N E \left(|\phi(\tilde{x}_t^i)|^p | \mathcal{F}_{t-1} \right) + \left(\sum_{i=1}^N E \left(|\phi(\tilde{x}_t^i)|^r | \mathcal{F}_{t-1} \right) \right)^{p/r} \right] \\ &\leq \frac{2^p C(p)}{N^p (1 - \epsilon_t)^{p/r}} \left[\sum_{i=1}^N E \left(|\phi(\tilde{x}_t^i)|^p | \mathcal{F}_{t-1} \right) + \left(\sum_{i=1}^N E \left(|\phi(\tilde{x}_t^i)|^r | \mathcal{F}_{t-1} \right) \right)^{p/r} \right] \end{aligned}$$

$$\begin{aligned} &\leq \frac{2^p C(p)}{N^p (1 - \epsilon_t)^{p/r}} \left[\sum_{i=1}^N \left(\pi_{t-1|t-1}^{N,\alpha_i}, K|\phi|^p \right) + \left(\sum_{i=1}^N \left(\pi_{t-1|t-1}^{N,\alpha_i}, K|\phi|^r \right) \right)^{p/r} \right] \\ &\leq \frac{2^p C(p)}{(1 - \epsilon_t)^{p/r}} \left[\frac{(\pi_{t-1|t-1}^N, K|\phi|^p)}{N^{p-1}} + \frac{(\pi_{t-1|t-1}^N, K|\phi|^r)^{p/r}}{N^{p-p/r}} \right]. \end{aligned}$$

Hence, by Lemma A.3 and (21),

$$E|\Pi_1|^p \leq \frac{2^{p+1} C(p) \|K\|^p M_{t-1|t-1}}{(1 - \epsilon_t)^{p/r}} \cdot \frac{\|\phi\|_{t-1,p}^p}{N^{p-p/r}} \triangleq C_{\Pi_1} \cdot \frac{\|\phi\|_{t-1,p}^p}{N^{p-p/r}}. \quad (25)$$

By (22)-(24), applying Lemma A.5 to $\xi = \frac{1}{N} \sum_{i=1}^N \phi(\tilde{x}_t^i)$ and $\eta = \frac{1}{N} \sum_{i=1}^N \phi(\tilde{x}_t^i)$ with $\epsilon = \frac{C_{\gamma_t} \|\rho\|_{t-1,p}^p}{N^p (1-1/r)} < \epsilon_t < 1$ (by (23) and (38) and the generation of $\{\tilde{x}_t^i\}$ in the algorithm), we have

$$\begin{aligned} |\Pi_2|^p &= \left| \frac{1}{N} \sum_{i=1}^N E \left(\phi(\tilde{x}_t^i) | \mathcal{F}_{t-1} \right) - \frac{1}{N} \sum_{i=1}^N E \left(\phi(\tilde{x}_t^i) | \mathcal{F}_{t-1} \right) \right|^p \\ &\leq \frac{2^p}{(1 - \epsilon)^p} \epsilon^{p-1} \cdot E \left[\left| \frac{1}{N} \sum_{i=1}^N \phi(\tilde{x}_t^i) \right|^p \middle| \mathcal{F}_{t-1} \right] \\ &\leq \frac{2^p}{(1 - \epsilon)^p} \epsilon^{p-1} \cdot \frac{1}{N} \sum_{i=1}^N E \left[|\phi(\tilde{x}_t^i)|^p | \mathcal{F}_{t-1} \right] \\ &\leq \frac{2^p}{(1 - \epsilon)^p} \epsilon^{p-1} \cdot \frac{1}{N} \sum_{i=1}^N (\pi_{t-1|t-1}^{N,\alpha_i}, K|\phi|^p) \\ &\leq \frac{2^p}{(1 - \epsilon_t)^p} \left(\frac{C_{\gamma_t} \|\rho\|_{t-1,p}^p}{N^p (1-1/r)} \right)^{p-1} \cdot \frac{1}{N} \sum_{i=1}^N (\pi_{t-1|t-1}^{N,\alpha_i}, K|\phi|^p) \\ &\leq C'_{\Pi_2} \cdot \frac{(\pi_{t-1|t-1}^N, K|\phi|^p)}{N^{p-p/r}}, \end{aligned}$$

where

$$C'_{\Pi_2} = \frac{2^p (C_{\gamma_t} \|\rho\|_{t-1,p}^p)^{p-1}}{(1 - \epsilon_t)^p}.$$

Here, Lemma A.5 is applied in the second line and in the third line we use Jensen's Inequality. Hence, by (21) and the above formula

$$E|\Pi_2|^p \leq C_{\Pi_2} \cdot \frac{\|\phi\|_{t-1,p}^p}{N^{p-p/r}}, \quad (26)$$

where $C_{\Pi_2} = C'_{\Pi_2} M_{t-1|t-1} \|K\|$. By (11) and (20),

$$E|\Pi_3|^p \leq C_{t-1|t-1} \|K\|^p \cdot \frac{\|\phi\|_{t-1,p}^p}{N^{p-p/r}} \triangleq C_{\Pi_3} \cdot \frac{\|\phi\|_{t-1,p}^p}{N^{p-p/r}}. \quad (27)$$

Then, using Minkowski's inequality, (25), (26) and (27), we have

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

$$E \left| (\tilde{\pi}_{t-1|t-1}^N, \phi) - (\pi_{t-1|t-1}, \phi) \right|^p \leq \tilde{C}_{t|t-1} \frac{\|\phi\|_{t-1,p}^p}{N^{p-p/r}}. \quad (28)$$

Let us now derive the fact that

$$E \left| (\tilde{\pi}_{t-1|t-1}^N, |\phi|^p) - (\pi_{t-1|t-1}, |\phi|^p) \right| \leq \tilde{M}_{t-1} \|\phi\|_{t-1,p}^p. \quad (29)$$

where

$$\tilde{M}_{t-1} \triangleq \left(\frac{4 - \epsilon_t}{1 - \epsilon_t} + 2 \right) \|K\|^p M_{t-1|t-1} \|\phi\|_{t-1,p}^p$$

using a separation similar to the one above. By Lemma A.5 and (21),

$$\begin{aligned}
 & E \left(\left(E \left[\left(\tilde{\pi}_{t|t-1}^N, |\phi|^p \right) - \frac{1}{N} \sum_{i=1}^N E \left(|\phi(\tilde{x}_t^i)|^p | \mathcal{F}_{t-1} \right) \right] \right) \right) \\
 &= \frac{1}{N} E \left(\left(E \left[\sum_{i=1}^N \left[|\phi(\tilde{x}_t^i)|^p - E(|\phi(\tilde{x}_t^i)|^p | \mathcal{F}_{t-1}) \right] \right] \right) \right) \\
 &\leq \frac{2}{N} E \left(\sum_{i=1}^N E(|\phi(\tilde{x}_t^i)|^p | \mathcal{F}_{t-1}) \right) \\
 &\leq \frac{2}{N(1-\epsilon_t)} E \left(\sum_{i=1}^N E(|\phi(\tilde{x}_t^i)|^p | \mathcal{F}_{t-1}) \right) \\
 &\leq \frac{2}{1-\epsilon_t} E(\pi_{t-1|t-1}^N, K|\phi|^p) \leq \frac{2}{1-\epsilon_t} \|K\|^p M_{t-1|t-1} \|\phi\|_{t-1,p}^p.
 \end{aligned} \tag{30}$$

By (22), (11), Lemma A.5 and (21),

$$\begin{aligned}
 & E \left| \frac{1}{N} \sum_{i=1}^N E \left[|\phi(\tilde{x}_t^i)|^p | \mathcal{F}_{t-1} \right] - \frac{1}{N} \sum_{i=1}^N E \left(|\phi(\tilde{x}_t^i)|^p | \mathcal{F}_{t-1} \right) \right| \\
 &= E \left| \frac{1}{N} \sum_{i=1}^N \left(E \left(|\phi(\tilde{x}_t^i)|^p | \mathcal{F}_{t-1} \right) - E \left(|\phi(\tilde{x}_t^i)|^p | \mathcal{F}_{t-1} \right) \right) \right| \\
 &\leq \frac{1}{N} \sum_{i=1}^N E \left(E \left(|\phi(\tilde{x}_t^i)|^p | \mathcal{F}_{t-1} \right) + E \left(|\phi(\tilde{x}_t^i)|^p | \mathcal{F}_{t-1} \right) \right) \\
 &\leq \left(\frac{1}{1-\epsilon_t} + 1 \right) \cdot \frac{1}{N} \sum_{i=1}^N E(\pi_{t-1|t-1}^{N,\alpha_i}, K|\phi|^p) \\
 &= \frac{2-\epsilon_t}{1-\epsilon_t} \cdot E(\pi_{t-1|t-1}^N, K|\phi|^p) \\
 &\leq \frac{2-\epsilon_t}{1-\epsilon_t} \cdot \|K\|^p M_{t-1|t-1} \|\phi\|_{t-1,p}^p.
 \end{aligned} \tag{31}$$

By (21) and noticing (23), we have

$$\begin{aligned}
 & E \left| \frac{1}{N} \sum_{i=1}^N (\pi_{t-1|t-1}^{N,\alpha_i}, K|\phi|^p) - (\pi_{t|t-1}, |\phi|^p) \right| \\
 &\leq \|K\|^p (M_{t-1|t-1} + 1) \|\phi\|_{t-1,p}^p.
 \end{aligned} \tag{32}$$

Then, by (30) (31) and (32), we have now proved (29).

3: Update In this step we analyse $E \left| (\tilde{\pi}_{t|t}^N, \phi) - (\pi_{t|t}, \phi) \right|^p$ and $E(\tilde{\pi}_{t|t}^N, |\phi|^p)$ based on (28) and (29). First, let us introduce the following separation

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

where

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

By condition H1 we have

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

Thus, by Minkowski's inequality and (28),

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

which implies

$$E \left| (\tilde{\pi}_{t|t}^N, \phi) - (\pi_{t|t}, \phi) \right|^p \leq \tilde{C}_{t|t} \frac{\|\phi\|_{t-1,p}^p}{N^{p-p/r}}. \tag{33}$$

Using a separation similar to the one mentioned above and (29) results in

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

Now, observing that $\|\phi\|_{s,p}$ is increasing with respect to s results in

$$\begin{aligned}
 & E \left| (\tilde{\pi}_{t|t}^N, |\phi|^p) \right| \leq \frac{\tilde{M}_{t|t-1} \|\rho\| (\|\rho\phi\| + \gamma_t)}{\gamma_t(\pi_{t|t-1}, \rho)} \cdot \|\phi\|_{t-1,p}^p + (\pi_{t|t}, |\phi|^p), \\
 &\leq \left(\frac{\tilde{M}_{t|t-1} \|\rho\| (\|\rho\phi\| + \gamma_t)}{\gamma_t(\pi_{t|t-1}, \rho)} + 1 \right) \cdot \|\phi\|_{t,p}^p \triangleq \tilde{M}_{t|t} \|\phi\|_{t,p}^p.
 \end{aligned} \tag{34}$$

5: Resampling Finally, we analyse $E \left| (\pi_{t|t}^N, \phi) - (\pi_{t|t}, \phi) \right|^p$ and $E(\pi_{t|t}^N, |\phi|^p)$ based on (33) and (34). Let us start by noticing 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 σ -algebra generated by $\{\tilde{x}_t^i\}_{i=1}^N$. From the generation of x_t^i , we have, $E(\phi(x_t^i) | \mathcal{G}_t) = (\tilde{\pi}_{t|t}^N, \phi)$, and then

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

Now, using Lemma A.1 and Lemma A.2, we obtain

$$\begin{aligned}
 E(|\bar{\Pi}_1|^p | \mathcal{G}_t) &= \frac{1}{N^p} E_{\mathcal{G}_t} \left| \sum_{i=1}^N (\phi(x_t^i) - E(\phi(x_t^i) | \mathcal{G}_t)) \right|^p \\
 &\leq 2^p C(p) \left[\frac{1}{N^{p-1}} E \left(|\phi(x_t^i)|^p | \mathcal{G}_t \right) \right. \\
 &\quad \left. + \frac{1}{N^{p(1-1/r)}} E^{p/r} \left(|\phi(x_t^i)|^r | \mathcal{G}_t \right) \right].
 \end{aligned}$$

Thus, by Lemma A.3 and (34),

$$E|\bar{\Pi}_1|^p \leq 2^{p+1} C(p) \tilde{M}_{t|t} \frac{\|\phi\|_{t,p}^p}{N^{p(1-1/r)}}. \tag{35}$$

Then by Minkowski's inequality, (33) and (35)

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

That is

$$E \left| (\pi_{t|t}^N, \phi) - (\pi_{t|t}, \phi) \right|^p \leq C_{t|t} \frac{\|\phi\|_{t,p}^p}{N^{p-p/r}}. \tag{36}$$

Using a separation similar to the one introduced above and (34) gives us

$$E \left| (\pi_{t|t}^N, |\phi|^p) - (\pi_{t|t}, |\phi|^p) \right| \leq (\pi_{t|t}^N, |\phi|^p) + (\pi_{t|t}, |\phi|^p) \leq (\tilde{M}_{t|t} + 1) \|\phi\|_{t,p}^p.$$

Hence,

$$E \left| (\pi_{t|t}^N, |\phi|^p) \right| \leq (\tilde{M}_{t|t} + 1) \|\phi\|_{t,p}^p \triangleq M_{t|t} \|\phi\|_{t,p}^p. \quad (37)$$

Therefore, the proof of Theorem 3.1 is completed, since (20) and (21) are successfully replaced by (36) and (37). \square

By the Borel-Cantelli Lemma and Chebyshev's inequality, we also have a convergence result as follow.

Theorem 3.2: In addition to H1 and H2, if $p > 2$, then for any function $\phi \in L_t^p(\rho)$, $\lim_{N \rightarrow \infty} (\pi_{t|t}^N, \phi) = (\pi_{t|t}, \phi)$ almost surely.

The proposition below guarantees that the requirement (12) does not result in an infinite loop in Algorithm 1.

Proposition 3.1: The particle filtering algorithm given in Algorithm 1 will not run into an infinite loop for sufficiently large N under the conditions of Theorem 3.1.

Proof. Based on the starting point (20) in the step 2 of the proof of the main theorem, we have

$$\begin{aligned} & P \left[(\pi_{t-1|t-1}^N, K\rho) < \gamma_t \right] \\ &= P \left[(\pi_{t-1|t-1}^N, K\rho) - (\pi_{t-1|t-1}, K\rho) < \gamma_t - (\pi_{t-1|t-1}, K\rho) \right] \\ &\leq P \left[|(\pi_{t-1|t-1}^N, K\rho) - (\pi_{t-1|t-1}, K\rho)| > |\gamma_t - (\pi_{t-1|t-1}, K\rho)| \right] \\ &\leq \frac{E |(\pi_{t-1|t-1}^N, K\rho) - (\pi_{t-1|t-1}, K\rho)|^p}{|\gamma_t - (\pi_{t-1|t-1}, K\rho)|^p} \\ &\leq \frac{C_{t-1|t-1} \|K\|^p}{|\gamma_t - (\pi_{t-1|t-1}, K\rho)|^p} \cdot \frac{\|\rho\|_{t-1,p}^p}{N^{p(1-1/r)}} \triangleq C_{\gamma_t} \cdot \frac{\|\rho\|_{t-1,p}^p}{N^{p(1-1/r)}}. \quad (38) \end{aligned}$$

Obviously, the probability in (38) tends to 0 as $N \rightarrow \infty$. We will now prove that

$$E(\pi_{t-1|t-1}^N, K\rho) > \gamma_t, \quad (39)$$

for large enough N . Note that since $0 < \gamma_t < (\pi_{t|t-1}, \rho)$ (condition H0), there exists a γ'_t such that $0 < \gamma_t < \gamma'_t < (\pi_{t|t-1}, \rho)$. Following the same steps as above, we have

$$P[(\pi_{t-1|t-1}^N, K\rho) < \gamma'_t] = O(1/N^{p(1-1/r)}) \rightarrow 0.$$

Then for sufficiently large N , we have

$$P[(\pi_{t-1|t-1}^N, K\rho) < \gamma'_t] < 1 - \frac{\gamma_t}{\gamma'_t}.$$

Thus,

$$P[(\pi_{t-1|t-1}^N, K\rho) \geq \gamma'_t] > \frac{\gamma_t}{\gamma'_t}.$$

For notational simplicity, define $\zeta \triangleq (\pi_{t-1|t-1}^N, K\rho)$ and use $f_\zeta(\cdot)$ to denote the density function of ζ . Let us now prove $E\zeta > \gamma_t$ for (39). Now,

$$\begin{aligned} E\zeta &= \int x f_\zeta(x) dx = \left(\int_{[\zeta \geq \gamma'_t]} + \int_{[\zeta < \gamma'_t]} \right) x f_\zeta(x) dx \\ &\geq \int_{[\zeta \geq \gamma'_t]} x f_\zeta(x) dx \geq \gamma'_t P[\zeta \geq \gamma'_t] > \gamma_t \cdot \frac{\gamma_t}{\gamma'_t} = \gamma_t, \end{aligned}$$

which is (39). Here, we have used the the fact that $\zeta \geq 0$ by noticing that $K\rho \geq 0$.

By a basic fact of Algorithm 1 demonstrated by (23) and the above formula (39) we know that

$$E \left[\frac{1}{N} \sum_{i=1}^N \rho(y_t | \tilde{x}_t^i) \right] = E(\pi_{t-1|t-1}^N, K\rho) > \gamma_t.$$

Therefore, for a given $\epsilon_t \in (0, 1)$ and a sufficiently large N , we have

$$P \left[\frac{1}{N} \sum_{i=1}^N \rho(y_t | \tilde{x}_t^i) < \gamma_t \right] < \epsilon_t < 1. \quad (40)$$

By Lemma A.4 this concludes that for sufficiently large N , with probability 1, the algorithm will not enter an infinite recursion. \square

IV. CONCLUSION

The main contribution of this work is the proof that the particle filter converge for unbounded functions in the sense of L^p -convergence, for $p \geq 2$. Besides this we also provide Lemma A.1, a new Rosenthal type inequality, which is generally applicable.

ACKNOWLEDGEMENT

We would like first to thank Professor James Lam and the anonymous reviewers for their careful reading and valuable comments which significantly improved the quality of the manuscript.

This work was partly supported by the strategic research center MOVIII, funded by the Swedish Foundation for Strategic Research (SSF) and CADICS, a Linneaus Center funded by the Swedish Research Council (VR). The work was also partially supported by the National Natural Science Foundation of China under Grant 60874029.

APPENDIX

In order to establish the convergence result, the following Rosenthal type inequality is needed.

Lemma A.1: Let $p > 0$, $1 \leq r \leq 2$, and let $\{\xi_i, i = 1, \dots, n\}$ be conditionally independent random variables, given σ -algebra \mathcal{G} such that $E(\xi_i | \mathcal{G}) = 0$, $E(|\xi_i|^p | \mathcal{G}) < \infty$ and $E(|\xi_i|^r | \mathcal{G}) < \infty$. Then there exists a constant $C(p)$ that depends only on p such that

$$E \left(\left| \sum_{i=1}^n \xi_i \right|^p \middle| \mathcal{G} \right) \leq C(p) \left[\sum_{i=1}^n E(|\xi_i|^p | \mathcal{G}) + \left(\sum_{i=1}^n E(|\xi_i|^r | \mathcal{G}) \right)^{p/r} \right]. \quad (41)$$

The inequality stated above hold in the almost sure sense, since it is in the form of a conditional expectation. For convenience, we omit the notation of almost sure in the lemma and its proof.

Remark A.1: When $r = 2$, (41) was first introduced in [2] for the special case of independent random variables, and then extend to a martingale difference sequence in [16]. The best constants $C(p)$ for both cases can be found in [17] and [18], respectively. For a brief proof of the independent case we refer to Appendix C in [19]. However, all the references mentioned require that $r = 2$, implying that the order of integrability should be no less than 2. This restriction has been improved to $r \in [1, 2]$ in Lemma A.1.

Remark A.2: For $0 < p \leq 2$ and $r = 2$ we have the following simplified form for (41) (see also Appendix C in [19])

$$E \left(\left| \sum_{i=1}^n \xi_i \right|^p \middle| \mathcal{G} \right) \leq \left(E \left(\left| \sum_{i=1}^n \xi_i \right|^2 \middle| \mathcal{G} \right) \right)^{p/2} = \left(\sum_{i=1}^n E(\xi_i^2 | \mathcal{G}) \right)^{p/2}. \quad (42)$$

Proof. See [20].

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

Proof. By Jensen's inequality, for $p \geq 1$, $(E|\xi|)^p \leq E|\xi|^p$. Hence, $E|\xi| \leq (E|\xi|^p)^{1/p}$. Then by Minkowski's inequality, we have

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

which derives the desired inequality.

Lemma A.3: If $0 < r_1 \leq r_2$ and $E|\xi|^{r_2} < \infty$, then $E^{1/r_1}|\xi|^{r_1} \leq E^{1/r_2}|\xi|^{r_2}$. \square

Proof. The result follows from Hölder's inequality: $E(|\xi|^{r_1} \cdot 1) \leq E^{r_1/r_2} \left((|\xi|^{r_1})^{r_2/r_1} \right)$.

Lemma A.4: Assume that a random variable ξ satisfies $P\{\xi < \gamma\} < 1$, where γ is a constant. Independently generate a sequence of samples $\{\xi_i\}$ with the same distribution as ξ until some $\xi_i \geq \gamma$. Then, this procedure cannot run into an infinite loop.

Proof. Note that

$$P\{\xi_1 < \gamma, \xi_2 < \gamma, \dots, \xi_n < \gamma\} = p^n \rightarrow 0$$

as $n \rightarrow \infty$, where $p = P\{\xi < \gamma\} < 1$. Thus, the process is almost surely finite. See also [20]. □

Lemma A.5: Let A be a Borel measurable subset of \mathcal{R}^m and sample the random vector ξ , obeying a probability density $d(t)$, until the realization belong to A , $t \in \mathcal{R}^m$. Suppose that $P\{\eta \in \Omega - A\} \leq \epsilon < 1$, where the random vector η obey the density $d(t)$ and ψ is a measurable function satisfying $E|\psi(\eta)|^p < \infty$, $p > 1$. Then, we have

$$|E\psi(\xi) - E\psi(\eta)| \leq \frac{2E^{1/p}|\psi(\eta)|^p}{1 - \epsilon} \epsilon^{\frac{p-1}{p}}. \quad (43)$$

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

$$E|\psi(\xi)| \leq \frac{E|\psi(\eta)|}{1 - \epsilon}. \quad (44)$$

Proof. Notice that the density of ξ is

$$\frac{d(t)I_A(t)}{\int d(t)I_A(t)dt},$$

Let us now prove (43),

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

which finishes the derivation of (43). □

The set A is typically defined by an inequality, say $\{f(\eta) > \gamma\}$. The result of Lemma A.5 can be extended to the conditional expectation case. For instance, in the case of (44), the conditional form would be

$$E[|\psi(\xi)| | \mathcal{F}] \leq \frac{E[|\psi(\eta)| | \mathcal{F}]}{1 - \epsilon},$$

where \mathcal{F} is a given σ -algebra and η has corresponding conditional density under the same condition $P\{\eta \in \Omega - A\} \leq \epsilon < 1$.

REFERENCES

[1] X.-L. Hu, T. B. Schön, and L. Ljung, "A basic convergence result for particle filtering," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1337–1348, Apr. 2008.

[2] 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.

[3] N. J. Gordon, D. J. Salmund, 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.

[4] J. V. Candy, *Bayesian Signal Processing: Classical, Unscented And Particle Filtering Methods*, ser. Adaptive And Learning Systems For Signal Processing, Communications And Control Series. Hoboken, NJ, USA: John Wiley & Sons, 2009.

[5] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: particle filters for tracking applications*. London, UK: Artech House, 2004.

[6] 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.

[7] A. Doucet and A. M. Johansen, "A tutorial on particle filtering and smoothing: Fifteen years later," in *Nonlinear Filtering Handbook*, D. Crisan and B. Rozovsky, Eds. Oxford University Press, 2009, to appear.

[8] O. Cappé, S. Godsill, and E. Moulines, "An overview of existing methods and recent advances in sequential Monte Carlo," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 899–924, 2007.

[9] P. M. Djuric, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Miguez, "Particle filtering," *IEEE Signal Processing Magazine*, vol. 20, no. 5, pp. 19–38, Sep. 2003.

[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] P. Del Moral, *Feynman-Kac formulae: Genealogical and Interacting Particle Systems with Applications*, ser. Probability and Applications. New York, USA: Springer, 2004.

[12] 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.

[13] 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.

[14] T. B. Schön, "Estimation of Nonlinear Dynamic Systems – Theory and Applications," Dissertations No 998, Department of Electrical Engineering, Linköping University, Feb. 2006.

[15] P. Del Moral, "Non-linear filtering: Interacting particle solution," *Markov processes and related fields*, vol. 2, no. 4, pp. 555–580, 1996.

[16] D. L. Burkholder, "Distribution function inequalities for martingales," *The Annals of Probability*, vol. 1, no. 1, pp. 19–42, 1973.

[17] W. B. Johnson, G. Schechtman, and J. Zinn, "Best constants in moment inequalities for linear combination of independent and exchangeable random variables," *The Annals of Probability*, vol. 13, no. 1, pp. 234–253, 1985.

[18] P. Hitczenko, "Best constants in martingale version of rosenthal's inequality," *The Annals of Probability*, vol. 18, no. 4, pp. 1656–1668, 1990.

[19] W. Hardle, G. Kerkycharian, D. Picard, and A. Tsybakov, *Wavelet, Approximation and Statistical Applications, Lectures Notes in Statistics 129*. New York, USA: Springer Verlag, 1998.

[20] X.-L. Hu, T. B. Schön, and L. Ljung, "Basic convergence results for particle filtering methods: Theory for the users," Department of Electrical Engineering, Linköping University, Linköping, Sweden, Tech. Rep. LiTH-ISY-R-2914, aug 2009, available from www.control.isy.liu.se/research/reports/2009/2914.pdf.