Basic Convergence Results for Particle Filtering Methods: Theory for the Users

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

Division of Automatic Control

E-mail: x33hu@ecemail.uwaterloo.ca, schon@isy.liu.se, ljung@isy.liu.se

21st August 2009

Report no.: LiTH-ISY-R-2914
Submitted to IEEE Transactions on Signal Processing

Address:
Department of Electrical Engineering
Linköpings universitet
SE-581 83 Linköping, Sweden

WWW: http://www.control.isy.liu.se

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET

Abstract

This work extends our recent work on proving that the particle filter converge for unbounded function to a more general case. More specifically, we prove that the particle filter converge for unbounded functions in the sense of L^p-convergence, for an arbitrary $p \geq 2$. Related to this, we also provide proofs for the case when the function we are estimating is bounded. In the process of deriving the main result we also established a new Rosenthal type inequality.

Keywords: Convergence, particle filter, nonlinear filtering, dynamic systems

Basic Convergence Results for Particle Filtering Methods: Theory for the Users

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

2007-07-20

Abstract

This work extends our recent work on proving that the particle filter converge for unbounded function to a more general case. More specifically, we prove that the particle filter converge for unbounded functions in the sense of L^p-convergence, for an arbitrary $p \geq 2$. Related to this, we also provide proofs for the case when the function we are estimating is bounded. In the process of deriving the main result we also established a new Rosenthal type inequality.

1 Introduction

The main purpose of the present work is to extend our previous results on particle filtering convergence [13] for unbounded functions to a more general setting. More specifically, we will here prove L^p-convergence for an arbitrary $p \geq 2$, of the particle filter. Hence, the main idea of the proof is present in [13]. However, to prove the $L^p, p \geq 2$ case requires some nontrivial embellishments, which forms the contribution of the present work. As a first step, we consider only the most basic problem: for any fixed time instance t, under what conditions and for what kind of function ϕ does the particle filtering approximation converges to the optimal filter

$$E[\phi(x_t)|y_1, \ldots, y_t]? \tag{1}$$

Moreover, we also establish two convergence results related to bounded function, which slightly extends the corresponding results in [2] in the sense that we consider a more general particle filtering algorithm.

The main contributions of this work are as follows,

- Convergence proof for the particle filter, regarding unbound functions ϕ (in $E[\phi(x_t)|y_1, \ldots, y_t]$) under more general conditions compared our previous work [13]. See Theorem 4.3.
- Convergence results for bounded function are also proposed, to slightly extend the counterpart of [2]. See Theorem 4.1.
- A Rosenthal type inequality under more loose setting in Lemma 4.1 is established during the theoretical preparation.

In Section 2 we introduce the models and the optimal filters that we are trying approximate and in Sections 3 the particle filter is introduced. However, these sections are intentionally rather brief, since a more detailed background using the same notation is already provided in [13]. The result are then presented in Section 4 and the conclusions are given in Section 5. Hence, readers familiar to the problem, can without problem directly jump to Section 4.

2 Model Setting and Optimal Filter

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 usually describes the evolution of the hidden state of a dynamic system, and the n_y-dimensional process Y denotes the available disturbed observation process of the same system. Roughly speaking, filtering the dynamic system is to estimate the state of the system based on observation data.

The state process X is a Markov process with initial state X_0 obeying distribution $\pi_0(dx_0)$ and probability transition kernel $K(dx_t|x_{t-1})$ such that

$$P(X_t \in A | X_{t-1} = x_{t-1}) = \int_A K(dx_t|x_{t-1}), \quad \forall A \in \mathcal{B}(\mathcal{R}^{n_x}). \tag{2}$$

The observations are conditionally independent of X and have marginal distribution

$$P(Y_t \in B | X_t = x_t) = \int_B \rho(dy_t|x_t), \quad \forall B \in \mathcal{B}(\mathcal{R}^{n_y}). \tag{3}$$

For convenience we assume that $K(dx_t|x_{t-1})$ and $\rho(dy_t|x_t)$ have densities with respect to Lebesgue measure. Hence, we can write

$$P(X_t \in dx_t | X_{t-1} = x_{t-1}) = K(dx_t|x_{t-1}) = K(x_t|x_{t-1})dx_t, \tag{4a}$$
$$P(Y_t \in dy_t | X_t = x_t) = \rho(dy_t|x_t) = \rho(y_t|x_t)dy_t. \tag{4b}$$

A frequently used model in practice is as follows using the notations above.

Example 2.1 *The state and observation of the model are described by*

$$x_t = f(x_{t-1}) + v_t, \tag{5a}$$
$$y_t = h(x_t) + e_t, \tag{5b}$$

where transformations $f : \mathcal{R}^{n_x} \times \mathcal{N} \to \mathcal{R}^{n_x}$ and $h : \mathcal{R}^{n_x} \times \mathcal{N} \to \mathcal{R}^{n_y}$, and v_t and e_t are process and observation noises with corresponding dimensions. The probability density functions for v_t and e_t are denoted by $p_v(\cdot, t)$ and $p_e(\cdot, t)$, respectively. For model (5) *we now have,*

$$K(x_t|x_{t-1}) = p_v(x_t - f(x_{t-1}), t), \quad \rho(y_t|x_t) = p_e(y_t - h(x_t), t).$$

Simply denote $Z_{k:l} \triangleq (Z_k, Z_{k+1}, \dots, Z_l)$ for two integers $k \le l$. Define the concerned conditional probability distribution of the system by

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

In practice, we typically care mostly about the marginal distribution $\pi_{t|t}(dx_t)$, since the main target is usually to estimate the standard optimal filter $E[X_t|y_{1:t}]$ and its conditional variance. We formulate the ideal form of $\pi_{t|t}(dx_t)$ first. By the total probability formula and Bayes' theorem, respectively, we have a recursion form of the marginal distribution

$$\pi_{t|t-1}(dx_t) = \int_{\mathcal{R}^{n_x}} \pi_{t-1|t-1}(dx_{t-1})K(dx_t|x_{t-1}) \triangleq b_t(\pi_{t-1|t-1}), \tag{6a}$$

$$\pi_{t|t}(dx_t) = \frac{\rho(y_t|x_t)\pi_{t|t-1}(dx_t)}{\int_{\mathcal{R}^{n_x}} \rho(y_t|x_t)\pi_{t|t-1}(dx_t)} \triangleq a_t(\pi_{t|t-1}), \tag{6b}$$

where a_t and b_t are transformations between probability measures on $\mathcal{R}^{n_x}$.

For convenience to represent the optimal filter, let us introduce some more notations. Given a measure ν, a function ϕ, and a Markov transition kernel K, denote

$$(\nu, \phi) \triangleq \int \phi(x)\nu(dx).$$

Hence,

$$E[\phi(X_t)|y_{1:t}] = (\pi_{t|t}, \phi).$$

Using this notation, by (6), for any function $\phi: R^{n_x} \to R$, we have a recursive form of the optimal filter $E[\phi(X_t)|y_{1:t}]$ according to

$$(\pi_{t|t-1}, \phi) = (\pi_{t-1|t-1}, K\phi), \tag{7a}$$

$$(\pi_{t|t}, \phi) = \frac{(\pi_{t|t-1}, \phi\rho)}{(\pi_{t|t-1}, \rho)}. \tag{7b}$$

Clearly, by (7), see also Lemma 2.1 of [7], we have

$$E[\phi(X_t)|y_{1:t}] = (\pi_{t|t}, \phi) = \frac{\int \cdots \int \pi_0(x_0)K_1\rho_1 \cdots K_t\rho_t\phi(x_t)dx_{0:t}}{\int \cdots \int \pi_0(x_0)K_1\rho_1 \cdots K_t\rho_t dx_{0:t}}, \tag{8}$$

where $K_s \triangleq K(x_s|x_{s-1})$, $\rho_s \triangleq \rho(y_s|x_s)$, $s = 1, \ldots, t$; $dx_{0:t} \triangleq dx_0 \cdots dx_t$; and with integral area all $\mathcal{R}^{n_x}$ omitted.

Technically, it is difficult to have an explicit solution for the optimal filter $E[\phi(X_t)|y_{1:t}]$ by (8) in general setting. Hence, numerical methods, such as the particle filter are introduced to approximate the optimal filter.

3 Particle Filtering

Roughly speaking, particle filtering methods are numerical algorithms to approximate the conditional distribution $\pi_{t|t}(dx_t)$ by an empirical distribution, constituted by a cloud of particles at each time instant. One important feature of the particle filter is that the integral operator over the empirical distribution turns to be a sum form. Hence, the difficult integral operation is simplified. Since there are two integral operators in (6), a standard practical particle filter usually sample particles two times from time $t-1$ to t for the estimates.

Specifically, at time $t = 0$, N initial particles $\{x_0^i\}_{i=1}^N$ are independently generated to obey the distribution $\pi_0(dx_0)$. Then, we introduce the algorithm in a recursive form. Let us at time $t-1$ assume that we have an approximation of the distribution $\pi_{t-1|t-1}(dx_{t-1})$ constituted by an empirical distribution

$$\pi^N_{t-1|t-1}(dx_{t-1}) \triangleq \frac{1}{N}\sum_{i=1}^N \delta_{x^i_{t-1}}(dx_t),$$

where $\delta_x(dx_t)$ denotes a delta-Dirac mass located in x.

In order to include the two slightly different kinds of particle filtering methods typically introduced by [10] in practise and by [4] for theoretical analysis respectively, we introduce weights for densities to sample particles. Denote

$$\alpha^i = (\alpha^i_1, \alpha^i_2, \dots, \alpha^i_N), \quad \alpha^i_j \geq 0, \quad \sum_{j=1}^N \alpha^i_j = 1, \quad \sum_{i=1}^N \alpha^i_j = 1.$$

Sample $\tilde{x}^i_t$ obeying $\sum_{j=1}^N \alpha^i_j K(dx_t|x^j_{t-1})$. Clearly,

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N \alpha^i_j K(dx_t|x^j_{t-1}) &= \frac{1}{N}\sum_{j=1}^N\left(\sum_{i=1}^N \alpha^i_j K(dx_t|x^j_{t-1})\right)\\
&= \frac{1}{N}\sum_{j=1}^N K(dx_t|x^j_{t-1})\\
&= (\pi^N_{t-1|t-1}, K). \qquad (9)
\end{aligned}$$

When $\alpha^i_j = 1$ for $j = i$, and $\alpha^i_j = 0$ for $j \neq i$, the sampling method reduces to a traditional way, as introduced by [10], see also [9, 18]. When $\alpha^i_j = 1/N$ for all i and j, it turns out to be a convenient form for theoretical treatment, as introduced by nearly all existing theoretical analysis references, for example [2, 4, 7, 8]. The empirical distribution of $\{\tilde{x}^i_t\}_{i=1}^N$

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

constitutes an estimate of $\pi_{t|t-1}$. When this estimate is substituted into (6b), we have an approximation for $\pi_{t|t}$

$$\tilde{\pi}^N_{t|t}(dx_t) \triangleq \frac{\rho(y_t|x_t)\tilde{\pi}^N_{t|t-1}(dx_t)}{\int_{R^{n_x}} \rho(y_t|x_t)\tilde{\pi}^N_{t|t-1}(dx_t)} = \frac{\sum_{i=1}^N \rho(y_t|\tilde{x}^i_t)\delta_{\tilde{x}^i_t}(dx_t)}{\sum_{i=1}^N \rho(y_t|\tilde{x}^i_t)}.$$

In practice, it is usually written using importance weights,

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

A very important step in the particle filter is the resampling step, which generates new equally weighted particles for the next step. So high dependence on a few particles with large weights is excluded. Specifically, sample x^i_t obeying $\tilde{\pi}^N_{t|t}(dx_t)$, then we get an equally weighted empirical distribution

$$\pi^N_{t|t}(dx_t) = \frac{1}{N}\sum_{i=1}^N \delta_{x^i_t}(dx_t)$$

to approximate $\pi_{t|t}$.

Let us point out the transformations of probabilities in the particle filtering algorithm. Recall the generation of $\tilde{x}_t^i$ first. We have the following transformations between probability measures immediately:

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

where Λ is an $N \times N$ matrix $(\alpha_j^i)_{i,j}$. Denote the whole transformation above as Λb_t for simplicity. We further denote by $c^n(\nu)$ the emperical distribution of a sample of size n from a probability distribution ν. Then, we have

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

where $c(N) \triangleq \frac{1}{N}[c^1 \ \ldots \ c^1]$ and $\bar{\circ}$ denotes composition of transformations in a vector multiplying form. Hence, in the general version of particle filtering algorithm, we have

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

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

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

While, in the existing theoretical version of particle filter in [2, 4, 7, 8], as stated in [2], the transformation between time $t-1$ and t is somewhat in a simple form:

$$\pi_{t|t}^N = c^N \circ a_t \circ c^N \circ b_t(\pi_{t-1|t-1}^N). \tag{10}$$

Hence,

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

The theoretical results and analysis in [15] are based on the following transformation (in our notation):

$$\pi_{t|t}^N = a_t \circ b_t \circ c^N(\pi_{t-1|t-1}^N), \tag{11}$$

which is the first formula in page 1999 at the begining of Section 4 in [15], rather than (10). Thus, the theoretical results do not include the standard particle filter in the popular theoretical setting, as in [2, 4, 7, 8]. As pointed at the beginning of this section, a standard particle filter sample particles two times from time $t-1$ to t to simplify the two integral operators in (6).

The whole procedure of particle filtering can be illustrated as in Figure 1. While the transformations of probability measures are showed in Figure 2.

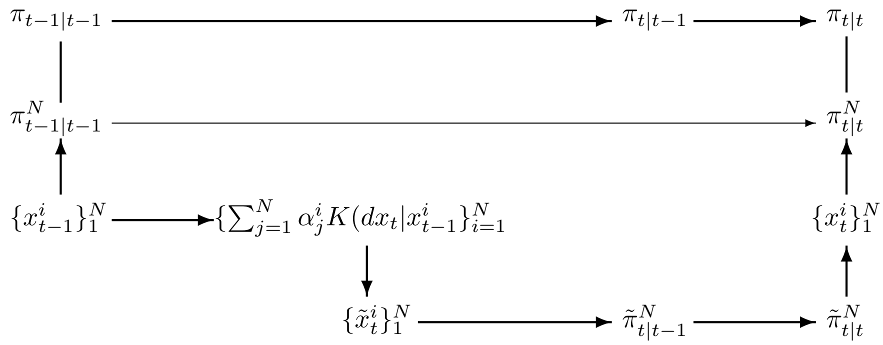

Figure 1. Illustration of the entire particle filtering algorithm.

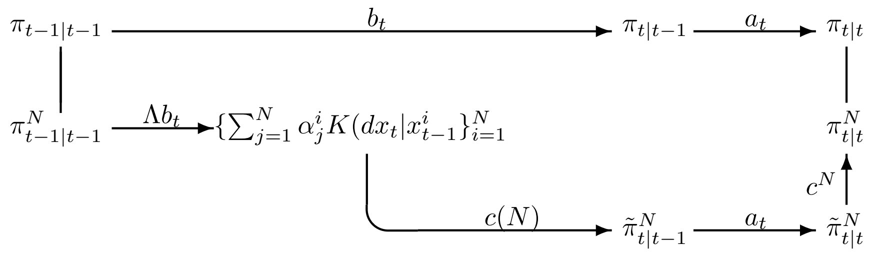

Figure 2. Transformation of probability measures in the particle filter.

Let us write the traditional form of the algorithm mentioned above in brief.

(0) $x^i_0 \sim \pi_0(dx_0)$, $i = 1, \ldots, N$.

(1) $\tilde{x}^i_t \sim \sum_{j=1}^N \alpha^i_j K(dx_t|x^j_{t-1})$, $i = 1, \ldots, N$.

(2) $\tilde{\pi}^N_{t|t}(dx_t) = \sum_{i=1}^N w^i_t \delta_{\tilde{x}^i_t}(dx_t)$, $w^i_t = \frac{\rho(y_t|\tilde{x}^i_t)}{\sum_{i=1}^N \rho(y_t|\tilde{x}^i_t)}$.

(3) $x^i_t \sim \tilde{\pi}^N_{t|t}(dx_t)$, $i = 1, \ldots, N$. $\pi^N_{t|t}(dx_t) = \frac{1}{N}\sum_{i=1}^N \delta_{x^i_t}(dx_t)$.

However, in order to avoid the well-known degeneracy of particle weight (see [2, 16]) and some difficulties of theoretical analysis for considering convergences to the optimal filter, we modify the particle filter above a little.

When we sample $\{\tilde{x}^i_t\}^N_1$ in the step (1) of the algorithm above, we check if

$$(\tilde{\pi}^N_{t|t-1}, \rho) = \sum_{i=1}^N \rho(y_t|\tilde{x}^i_t) \geq \gamma_t > 0, \tag{12}$$

where the real number γ_t is selected by experience, say $\gamma_t = \gamma(\pi_{t|t-1}, \rho)$ if $(\pi_{t|t-1}, \rho) > 0$ is known and $0 < \gamma < 1$. If the inequality holds, the algorithm proceeds as proposed, whereas if (12) does not hold, we regenerate $\{\tilde{x}^i_t\}^N_1$ again until (12) is satisfied. That is, we change step (1) of the algorithm into the following form:

(1′) $\tilde{x}^i_t \sim \sum_{j=1}^N \alpha^i_j K(dx_t|x^j_{t-1})$, $i = 1, \ldots, N$, with (12) satisfied.

The modified algorithm proceeds as: (0)(1′)(2)(3), and the following theoretical analyses are all based on this version. With help of Lemma 4.4 and (45) in the proof of Theorem 4.3, we conclude the following:

Proposition 3.1 *The modified algorithm will not run into an infinite loop for sufficiently large N under the conditions of Theorem 4.3.*

Proof. We get formula (45) in the second step of the proof of Theorem 4.3. Based on this formula, we first calculate the following probability:

$$\begin{aligned} P[(\tilde{\pi}^N_{t|t-1}, \rho) < \gamma_t] &= P[(\tilde{\pi}^N_{t|t-1}, \rho) - (\pi_{t|t-1}, \rho) < \gamma_t - (\pi_{t|t-1}, \rho)] \\ &\leq P[|(\tilde{\pi}^N_{t|t-1}, \rho) - (\pi_{t|t-1}, \rho)| > |\gamma - 1|(\pi_{t|t-1}, \rho)] \\ &\leq \frac{1}{(1-\gamma)^p (\pi_{t|t-1}, \rho)^p} E|(\tilde{\pi}^N_{t|t-1}, \rho) - (\pi_{t|t-1}, \rho)|^p \\ &\leq \frac{\tilde{C}_{t|t-1}}{(1-\gamma)^p (\pi_{t|t-1}, \rho)^p} \cdot \frac{\|\rho\|^p_{t-1,p}}{N^{p-p/r}} \xrightarrow[N\to\infty]{} 0. \end{aligned} \tag{13}$$

We use (45) with ϕ replaced by ρ in the last step of (13). Hence, $P[(\tilde{\pi}^N_{t|t-1}, \rho) < \gamma_t] < 1$ for sufficiently large N. In view of Lemma 4.4, the modified step $(1')$ is impossible to run into infinite loop. This proves the assertion.

□

By (13), $P[(\tilde{\pi}^N_{t|t-1}, \rho) \geq \gamma_t] \xrightarrow[N\to\infty]{} 1$, which means the lower bound for $(\tilde{\pi}_{t|t-1}, \rho)$ is almost always satisfied, provided that N is sufficiently large. See [13] for a numerical experiment, showing the relation between the sample times and N.

It is worth noting that originally given $\{x^i_{t-1}, i = 1, \ldots, N\}$ the joint density of $\tilde{x}^i_t$, $i = 1, \ldots, N$ is

$$P\left[\tilde{x}^i_t = s_i, i = 1, \ldots, N\right] = \prod_{i=1}^N \sum_{j=1}^N \alpha^i_j K(s_i | x^j_{t-1}) \triangleq \Pi^N_{\alpha_1,\ldots,\alpha_N}. \tag{14}$$

Yet, after the modification it is changed to be

$$\bar{\Pi}^N_{\alpha_1,\ldots,\alpha_N} = \frac{\Pi^N_{\alpha_1,\ldots,\alpha_N} I_{[\frac{1}{N}\sum_{i=1}^N \rho(y_t|s_i) \geq \gamma_t]}}{\int \cdots \int \Pi^N_{\alpha_1,\ldots,\alpha_N} I_{[\frac{1}{N}\sum_{i=1}^N \rho(y_t|s_i) \geq \gamma_t]} ds_{1:N}}, \tag{15}$$

where the record y_t is given. A related theoretical preliminary regarding this fact has been proposed in Lemma 4.5.

4 Convergence to Optimal Filters

In this section we consider under what conditions the particle filtering approximation converges to the optimal filters (8), with respect to bounded and unbounded function $\phi(\cdot)$ respectively, when the number of the particles N tends to infinity. All the following convergence results are based on the assumption that the observation process is fixed to a given observation record $Y_s = y_s$, $s = 1, \ldots, t$, which is a general theoretical setting for the existing convergence results, see, for instance, [2, 4, 7, 8]. Thus, the expectation operators in the Theorem 4.1, Theorem 4.3, and their proofs are in the sense of $E[\cdot|Y_{1:s} = y_{1:s}]$, $s = 1, \ldots, t$. Hence, the constants there may depend on $y_{1:t}$.

4.1 Auxiliary Lemmas

In order to establish some of the convergence results, the following powerful Rosenthal type inequality is needed. This inequality hold in the sense of almost sure, since it is in the form of a conditional expectation. However, in the interest of readability, we omit the notation of almost sure in the following lemma and its proof.

Lemma 4.1 *Let $p > 0$, $1 \leq r \leq 2$, and let $\{\xi_i,\ i = 1, \ldots, n\}$ be conditionally independent random variables, given a 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*

$$E\left[\left|\sum_{i=1}^n \xi_i\right|^p |\mathcal{G}\right] \leq C(p)\left[\sum_{i=1}^n E[|\xi_i|^p|\mathcal{G}] + \left(\sum_{i=1}^n E[|\xi_i|^r|\mathcal{G}]\right)^{p/r}\right]. \quad (16)$$

Remark 4.1 *When $r = 2$, (16) was first introduced in [17] for the special case of independent random variables, and then extend to martingale difference sequences in [1]. The best constants $C(p)$ for both cases can be found in [14] and [12], respectively. For a brief proof of the independent case we refer to the Appendix C of [11]. However, all the references mentioned require that $r = 2$, and so the order of integrability should be no less than 2. This restriction has been relaxed to $r \in [1, 2]$ in Lemma 4.1, and so the order need only not less than 1 here.*

Remark 4.2 *For $0 < p \leq 2$ and $r = 2$, by the classic convexity inequality, (16) assumes a simpler form (see also Appendix C of [11])*

$$E\left[\left|\sum_{i=1}^n \xi_i\right|^p |\mathcal{G}\right] \leq \left(E\left[\left|\sum_{i=1}^n \xi_i\right|^2 |\mathcal{G}\right]\right)^{p/2} = \left(\sum_{i=1}^n E\left[\xi_i^2|\mathcal{G}\right]\right)^{p/2}. \quad (17)$$

Proof. Here, we only consider the case of $1 < r < 2$, since the proof for $r = 2$ is nearly the same as Appendix C of [11], and $r = 1$ is a trivial case with $C(p) = 1$ and the first term in right hand side is omitted. We first prove a basic inequality, and then prove (16).

Let $\{\eta_i, i = 1, \ldots, n\}$ be a sequence of independent random variables such that $E\eta_i \leq 0$, $P[\eta_i \leq M] = 1$, $0 < M < \infty$, and denote $\sigma_r(\eta) = \sum_{i=1}^n E[|\eta_i|^r|\mathcal{G}]$, for any $\lambda \geq \lambda(M) \triangleq (e^2 - 1)\sigma_r(\eta)/M^{r-1} > 0$, we prove the following Bennett-type inequality

$$P\left[\sum_{i=1}^n \eta_i > \lambda|\mathcal{G}\right] \leq \exp\left(-\frac{\sigma_r(\eta)}{M^r}\theta\left(\frac{\lambda M^{r-1}}{\sigma_r(\eta)}\right)\right), \quad (18)$$

where $\theta(x) = (1 + x)\log(1 + x) - x$.

Define function $\psi(x) = (e^x - 1 - x)/|x|^r$ for $x \neq 0$, and $\psi(0) = \lim_{x\to 0} \psi(x)$. Clearly, $\psi(x)$ is a positive and non-decreasing function on the interval $[0, \infty)$, while it is still positive and has just one maximum, denoted by x_0, on the interval $(-\infty, 0]$. Clearly, x_0 satisfy $\psi'(x) = 0$, which is equivalent to

$$-x_0 e^{x_0} + x_0 + re^{x_0} - r - rx_0 = 0.$$

Hence,

$$\psi(x_0) = \frac{e^{x_0} - 1 - x_0}{(-x_0)^r} = \frac{1 - e^{x_0}}{r(-x_0)^{r-1}} < \frac{\min\{1, -x_0\}}{r(-x_0)^{r-1}} < 1.$$

Define $x_0^+ > 0$ which satisfiers $\psi(x_0^+) = \psi(x_0)$. Notice that

$$\psi(x_0^+) < 1 < \frac{e^2 - 1 - 2}{4} < \frac{e^2 - 1 - 2}{2^r} = \psi(2),$$

we have $0 < x_0^+ < 2$ by the monotonicity of ψ on $[0, \infty)$. Thus, for any $x_1 < x_2$ and $x_2 \geq x_0^+$, we have $\psi(x_1) < \psi(x_2)$.

Clearly, for any $t > 0$, using the Markov inequality and conditional independence we have

$$\begin{aligned} P\left[\sum_{i=1}^n \eta_i > \lambda|\mathcal{G}\right] &\leq \exp(-\lambda t) E\left[\exp\left(\sum_{i=1}^n t\eta_i\right)|\mathcal{G}\right] \\ &= \exp\left(-\lambda t + \sum_{i=1}^n \log E[e^{t\eta_i}|\mathcal{G}]\right). \end{aligned} \tag{19}$$

Notice that $E[\eta_i|\mathcal{G}] \leq 0$, $\log(1+x) \leq x$ for $x > -1$, and the property of function ψ, for $tM \geq 1$ we have

$$\begin{aligned} \log E[e^{t\eta_i}|\mathcal{G}] &= \log E[e^{t\eta_i} - 1 - t\eta_i + 1 + t\eta_i|\mathcal{G}] \\ &\leq \log(E[e^{t\eta_i} - 1 - t\eta_i|\mathcal{G}] + 1) \\ &= \log(1 + E[|t\eta_i|^r \psi(t\eta_i)|\mathcal{G}]) \\ &\leq E[|\eta_i|^r t^r \psi(t\eta_i)|\mathcal{G}] \\ &\leq \psi(tM) t^r E[|\eta_i|^r|\mathcal{G}]. \end{aligned}$$

Hence, (19) turns to be

$$\begin{aligned} P\left[\sum_{i=1}^n \eta_i > \lambda|\mathcal{G}\right] &\leq \exp\left(-[\lambda t - t^r \sigma_r(\eta)\psi(tM)]\right) \\ &= \exp\left(-[\lambda t - \sigma_r(\eta)(e^{tM} - 1 - tM)/M^r]\right). \end{aligned}$$

The optimal selection of $tM \geq 2$ is

$$t = \frac{1}{M} \log\left(1 + \frac{\lambda M^{r-1}}{\sigma_r(\eta)}\right),$$

which yields (18) and requires that $\lambda \geq (e^2 - 1)\sigma_r(\eta)/M^{r-1}$.

Now we are in a position to prove (16). For simplicity, we use the function $x \log x(1 + x) - x$, which is smaller than $\theta(x)$, in the inequality (18). Let us define an upper bounded function first. For $M > 0$, define $\eta_i = \xi_i I_{[|\xi_i| \leq M]}$. Thus $E[\eta_i|\mathcal{G}] \leq E[\xi_i|\mathcal{G}] = 0$, $\eta_i \leq M$, and

$$\sigma_r(\eta) \triangleq \sum_{i=1}^n E[|\eta_i|^r|\mathcal{G}] \leq \sum_{i=1}^n E[|\xi_i|^r|\mathcal{G}] \triangleq \sigma_r.$$

Putting $M = \lambda/\kappa$, $\kappa \geq 1$. By (18), for

$$\lambda \geq \lambda_0 \triangleq [(e^2 - 1)\kappa^{r-1}\sigma_r]^{1/r} \geq [(e^2 - 1)\kappa^{r-1}\sigma_r(\eta)]^{1/r},$$

we have

$$P\left[\sum_{i=1}^{n}\eta_i > \lambda|\mathcal{G}\right] \le \exp\left\{-\kappa\left[\log\left(1+\frac{\lambda^r}{\kappa^{r-1}\sigma_r}\right)-1\right]\right\}.$$

Hence, for $\lambda \ge \lambda_0$, we have

$$\begin{aligned}P\left[\sum_{i=1}^{n}\xi_i > \lambda|\mathcal{G}\right] &= P\left[\sum_{i=1}^{n}\xi_i > \lambda, \xi_i < M, i=1,\dots,n|\mathcal{G}\right] + P\left[\sum_{i=1}^{n}\xi_i > \lambda,\ \max_{1\le i\le n}\xi_i \ge M|\mathcal{G}\right]\\ &\le P\left[\sum_{i=1}^{n}\eta_i > \lambda|\mathcal{G}\right] + P\left[\max_{1\le i\le n}\xi_i \ge M|\mathcal{G}\right]\\ &\le \exp\left\{-\kappa\left[\log\left(1+\frac{\lambda^r}{\kappa^{r-1}\sigma_r}\right)-1\right]\right\} + \sum_{i=1}^{n}P\left[\xi_i \ge M|\mathcal{G}\right].\end{aligned} \tag{20}$$

Similarly, we can obtain an inequality in the same form as (20) for $\sum_{i=1}^{n}(-\xi_i)$. Therefore,

$$P\left[\left|\sum_{i=1}^{n}\xi_i\right| > \lambda|\mathcal{G}\right] \le 2\exp\left\{-\kappa\left[\log\left(1+\frac{\lambda^r}{\kappa^{r-1}\sigma_r}\right)-1\right]\right\} + \sum_{i=1}^{n}P\left[\kappa|\xi_i| \ge \lambda|\mathcal{G}\right]. \tag{21}$$

Now, using (21), we have

$$\begin{aligned}E\left|\sum_{i=1}^{n}\xi_i\right|^p &= E\left(\left|\sum_{i=1}^{n}\xi_i\right| I_{[|\sum_{i=1}^{n}\xi_i|<\lambda_0]}\right)^p + E\left(\left|\sum_{i=1}^{n}\xi_i\right| I_{[|\sum_{i=1}^{n}\xi_i|\ge\lambda_0]}\right)^p\\ &< \lambda_0^p + \int_{\lambda_0}^{\infty} pt^{p-1}P\left[\left|\sum_{i=1}^{n}\xi_i\right| > t|\mathcal{G}\right]dt\\ &\le \lambda_0^p + 2p\int_{\lambda_0}^{\infty} t^{p-1}\exp\left\{-\kappa\left[\log\left(1+\frac{t^r}{\kappa^{r-1}\sigma_r}\right)-1\right]\right\}dt\\ &\quad + \sum_{i=1}^{n}\int_{\lambda_0}^{\infty} pt^{p-1}P\left[\kappa|\xi_i| \ge t|\mathcal{G}\right]dt\\ &\le (\kappa^{r-1}\sigma_r)^{p/r}\left[(e^2-1)^{1/r} + 2pe^{\kappa}\int_{(e^2-1)^{1/r}}^{\infty} s^{p-1}(1+s^r)^{-\kappa}ds\right] + \sum_{i=1}^{n}E|\kappa\xi_i|^p,\end{aligned}$$

where the variable substitution $t = (\kappa^{r-1}\sigma_r)^{1/r}s$ has been used. For the convergence of the integral on right hand side, we select $\kappa > \max\{1, p/r\}$. Then the proof of the lemma is completed with

$$C(p) = \max\left\{\kappa^{p(r-1)/r}\left[(e^2-1)^{1/r} + 2pe^{\kappa}\int_{(e^2-1)^{1/r}}^{\infty} s^{p-1}(1+s^r)^{-\kappa}ds\right], \kappa^p\right\}.$$

$\square$

Below we provide two Lemmas which will become useful in the following.

Lemma 4.2 *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,

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

which derives the desired inequality.

□

Lemma 4.3 *If $0 < r_1 \leq r_2$ and $E|\xi|^{r_2} < \infty$, then $E^{1/r_1}|\xi|^{r_1} \leq E^{1/r_2}|\xi|^{r_2}$.*

Proof. Simply by Hölder's inequality: $E\left[|\xi|^{r_1} \cdot 1\right] \leq E^{r_1/r_2}\left[(|\xi|^{r_1})^{r_2/r_1}\right]$. Then the lemma follows.

□

Lemma 4.4 *Assume that a random variable ξ satisfies $P[\xi < \gamma] < 1$, where γ is a known constant. Independently generate a sample ξ_1 with the same distribution as ξ. If $\xi_1 < \gamma$, then independently generate ξ_2 and check again; otherwise, stop. This procedure cannot run into an infinite loop.*

The proof is quite straightforward. Suppose the converse, i.e., there exist a sequence of i.i.d. random variables $\{\xi_i\}$ such that $\xi_i < \gamma$ for any i. Then,

$$P[\xi_i < \gamma,\ i = 1, 2, \ldots] = \Pi_{i=1}^{\infty} P[\xi < \gamma] = 0,$$

which means the probability is 0.

Lemma 4.5 *Let A is a Borel measurable subset of $\mathcal{R}^m$ and sample random vector ξ obey a probability density $d(t)$ until the relization belong to A, $t \in \mathcal{R}^m$. Suppose that*

$$P[\eta \in\ \Omega - A] \leq \epsilon < 1, \tag{22}$$

where the random vector η obey the density $d(t)$ and ψ is a measurable function satisfying $E\psi^p(\eta) < \infty$, $p > 1$. Then, we have

$$|E\psi(\xi) - E\psi(\eta)| \leq \frac{2E^{1/p}|\psi(\eta)|^p}{1-\epsilon}\epsilon^{\frac{p-1}{p}}. \tag{23}$$

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

$$E|\psi(\xi)| \leq \frac{E|\psi(\eta)|}{1-\epsilon}. \tag{24}$$

Proof. Notice that the density of ξ is

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

It is trivial for (24). While

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

which derives (23).

□

The result of Lemma 4.5 is easily to extend to condtional expectaion case.

4.2 Convergence for Bounded Functions

Let us first consider convergence issues regarding bounded function ϕ in the optimal filter $E[\phi(x_t)|y_{1:t}]$. Although this topic has been studied in many existing references, see, for instance, [2, 4, 7, 8], yet, as stated in Section 3, to the authors' knowledge all existing theoretical convergence results are based on a theoretical setting of particle filter and unable to include the most frequently used form of the particle filter, as proposed in [9, 10, 18]. Moreover, the following Theorem 4.1 and Theorem 4.2 slightly extend the results of [2].

Define the norm $\|f(x)\| \triangleq \max_x |f(x)|$. Denote $B(R^{n_x})$ all bounded functions on R^{n_x}.

H0. $\rho(y_t|x_t)$ is a bounded and positive function for given $y_{1:t}$.

Theorem 4.1 *If H0 holds then, for any $\phi \in B(R^{n_x})$ and $p > 0$, there exists a constant $c_{t|t}$ independent of N such that*

$$
E\left|(\pi_{t|t}^N, \phi) - (\pi_{t|t}, \phi)\right|^p \leq c_{t|t} \frac{\|\phi\|^p}{N^{p/2}}. \tag{25}
$$

Proof. The proof is in the form of a mathematical induction.

1: Initialization

Let $\{x_0^i\}_{i=1}^N$ be independent random variables with the same distribution $\pi_0(dx_0)$.

Then, for $p > 2$ using Lemmas 4.1 with $r = 2$ it is clear that

$$\begin{aligned}
E\left|(\pi_0^N,\phi)-(\pi_0,\phi)\right|^p &= \frac{1}{N^p}E\left|\sum_{i=1}^N(\phi(x_0^i)-E[\phi(x_0^i)])\right|^p \\
&\le \frac{C(p)}{N^p}\left[\sum_{i=1}^N E|\phi(x_0^i)-E[\phi(x_0^i)]|^p+\left[\sum_{i=1}^N E|\phi(x_0^i)-E[\phi(x_0^i)]|^2\right]^{p/2}\right] \\
&\le 2^pC(p)\left[\frac{\|\phi\|^p}{N^{p-1}}+\frac{\|\phi\|^p}{N^{p/2}}\right] \\
&\le 2^{p+1}C(p)\frac{\|\phi\|^p}{N^{p/2}} \triangleq c_{0|0}\frac{\|\phi\|^p}{N^{p/2}}. \qquad (26)
\end{aligned}$$

For $0 < p \le 2$, using (17) we also have an inequality in the same form as (26).

2: Prediction

Based on (26), assume that for $t-1$ and $\forall \phi \in B(R^{n_x})$

$$E\left|(\pi_{t-1|t-1}^N,\phi)-(\pi_{t-1|t-1},\phi)\right|^p \le c_{t-1|t-1}\frac{\|\phi\|^p}{N^{p/2}} \qquad (27)$$

holds. In this step we analyse $E\left|(\tilde{\pi}_{t|t-1}^N,\phi)-(\pi_{t|t-1},\phi)\right|^p$. The fact that

$$|K\phi| = \left|\int K(dx_t|x_{t-1})\phi(x_t)\right| \le \|\phi\|$$

will be frequently used in the rest of this proof.

Notice that

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

where

$$\Pi_1 \triangleq \left[(\tilde{\pi}_{t|t-1}^N,\phi)-\frac{1}{N}\sum_{i=1}^N(\pi_{t-1|t-1}^{N,\alpha_i},K\phi)\right],$$

$$\Pi_2 \triangleq \left[\frac{1}{N}\sum_{i=1}^N(\pi_{t-1|t-1}^{N,\alpha_i},K\phi)-(\pi_{t|t-1},\phi)\right],$$

and $\pi_{t-1|t-1}^{N,\alpha_i}=\sum_{j=1}^N\alpha_j^i\delta_{x_{t-1}^j}$. We will now investigate Π_1 and Π_2 more closely.

Let $\mathcal{F}_{t-1}$ denote the σ-algebra generated by $\{x_{t-1}^i, i=1,\dots,N\}$. From the generation of $\tilde{x}_t^i$, we have,

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

and hence,

$$\Pi_1 = \frac{1}{N}\sum_{i=1}^N(\phi(\tilde{x}_{t-1}^i)-E[\phi(\tilde{x}_{t-1}^i)|\mathcal{F}_{t-1}]).$$

Thus, for $p > 2$ by Lemmas 4.1 with $r = 2$ and (9),

$$\begin{aligned}
E\left[|\Pi_1|^p|\mathcal{F}_{t-1}\right] &= \frac{1}{N^p}E\left[\left|\sum_{i=1}^N(\phi(\tilde{x}_{t-1}^i)-E[\phi(\tilde{x}_{t-1}^i)|\mathcal{F}_{t-1}])\right|^p\middle|\mathcal{F}_{t-1}\right] \\
&\le 2^pC(p)\left[\frac{(\pi_{t-1|t-1}^N,K|\phi|^p)}{N^{p-1}}+\frac{(\pi_{t-1|t-1}^N,K|\phi|^2)^{p/2}}{N^{p/2}}\right].
\end{aligned}$$

For $0 < p \leq 2$, using (17) we have an inequality similar to the one above.

$$E|\Pi_1|^p \leq 2^{p+1} C(p) \frac{\|\phi\|^p}{N^{p/2}}. \tag{28}$$

By (9),

$$\frac{1}{N} \sum_{i=1}^{N} (\pi_{t-1|t-1}^{N,\alpha_i}, K\phi) = (\pi_{t-1|t-1}^{N}, K\phi).$$

Notice the assumption (27),

$$E|\Pi_2|^p \leq c_{t-1|t-1} \frac{\|\phi\|^p}{N^{p/2}}. \tag{29}$$

Then, by Minkowski's inequality, (27), (28) and (29),

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

That is

$$E \left| (\tilde{\pi}_{t|t-1}^N, \phi) - (\pi_{t|t-1}, \phi) \right|^p \leq \tilde{c}_{t|t-1} \frac{\|\phi\|^p}{N^{p/2}}. \tag{30}$$

3: Update In this step we go one step further to analyse $E \left| (\tilde{\pi}_{t|t}^N, \phi) - (\pi_{t|t}, \phi) \right|^p$ based on (30). Clearly,

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

where

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

Note that ϕ, ρ are bounded functions and that ρ is a positive function. Then we have,

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

By Minkowski's inequality and (30),

$$E^{1/p} \left| (\tilde{\pi}_{t|t}^N, \phi) - (\pi_{t|t}, \phi) \right|^p \leq E^{1/p} |\tilde{\Pi}_1|^p + E^{1/p} |\tilde{\Pi}_1|^p \leq \frac{2\|\rho\| \tilde{c}_{t|t-1}^{1/p}}{(\pi_{t|t-1}, \rho)} \cdot \frac{\|\phi\|}{N^{1/2}},$$

which implies,

$$E\left|(\tilde{\pi}^N_{t|t},\phi)-(\pi_{t|t},\phi)\right|^p \le \frac{2^p\|\rho\|^p\tilde{c}_{t|t-1}}{(\pi_{t|t-1},\rho)^p}\cdot\frac{\|\phi\|}{N^{p/2}} \triangleq \tilde{c}_{t|t}\frac{\|\phi\|^p}{N^{p/2}}. \tag{31}$$

4: Resampling Finally, we analyse $E\left|(\pi^N_{t|t},\phi)-(\pi_{t|t},\phi)\right|^p$ based on (31). Let us start by noticing that

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

where

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

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

$$E[\phi(x^i_t)|\mathcal{G}_t]=(\tilde{\pi}^N_{t|t},\phi),$$

and then

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

Now, for $p>2$ by Lemmas 4.1 with $r=2$, we have

$$\begin{aligned} E\left[|\bar{\Pi}_1|^p|\mathcal{G}_t\right] &= \frac{1}{N^p}E\left[\left|\sum_{i=1}^N(\phi(x^i_t)-E[\phi(x^i_t)|\mathcal{G}_t])\right|^p\Big|\mathcal{G}_t\right] \\ &\le 2^pC(p)\left[\frac{1}{N^{p-1}}E\left[|\phi(x^i_t)|^p|\mathcal{G}_t\right]+\frac{1}{N^{p/2}}E^{p/2}\left[|\phi(x^i_t)|^2|\mathcal{G}_t\right]\right]. \end{aligned}$$

For $0<p\le 2$, using (17) we have an inequality similar to the one above. Hence,

$$E|\bar{\Pi}_1|^p \le 2^{p+1}C(p)\frac{\|\phi\|^p}{N^{p/2}}. \tag{32}$$

Then, by Minkowski's inequality, (31) and (32),

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

That is,

$$E\left|(\pi^N_{t|t},\phi)-(\pi_{t|t},\phi)\right|^p \le c_{t|t}\frac{\|\phi\|^p}{N^{p/2}},$$

which completes the proof of Theorem 4.1.

□

Remark 4.3 *One can also use a Marcinkiewicz-Zygmund type inequality (see Lemma 7.3.3 of [8]) to prove the result of Theorem 4.1 for* $p\ge 1$.

For $p > 2$ in Theorem 4.1, by Borel-Cantelli Lemma we have a weak convergence result as follow.

Theorem 4.2 *If H0 holds, then for any fixed t, $\pi^N_{t|t}$ converges weakly to $\pi_{t|t}$ almost surely, i.e., for any bounded continuous function ϕ on R^{n_x},*

$$\lim_{N\to\infty} (\pi^N_{t|t}, \phi) = (\pi_{t|t}, \phi)$$

almost surely.

Remark 4.4 *For the algorithm (0)(1′)(2)(3), Theorems 4.1 and 4.2 hold for the simplified version of condition H0:*

H0′. *$\rho(y_t|x_t)$ is a bounded function for given $y_{1:t}$ such that $(\pi_{s|s-1}, \rho) > 0$, $s = 1, 2, \ldots, t$.*

4.3 Convergence for Unbounded Functions

In this section we consider convergences to the optimal filter $E[\phi(x_t)|y_{1:t}]$ in the case where ϕ is an unbounded function, based on the modified version of particle filter proposed in Section 3.

Below we list conditions that we need for further considerations of convergences with respect to unbounded function ϕ.

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

$$0 < \gamma_s < (\pi_{s|s-1}, \rho), \;\; s = 1, 2, \ldots, t,$$

equivalently, $\gamma_s = \gamma(\pi_{s|s-1}, \rho)$ with $0 < \gamma < 1$, $s = 1, 2, \ldots, t$.

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

H2. For some $p > 1$, function $\phi(\cdot)$ satisfy $|\phi(x_s)|^p \rho(y_s|x_s) < \infty$ for given $y_{1:s}$, $s = 1, \ldots, t$.

Remark 4.5 *In view of (7b), clearly, $(\pi_{s|s-1}, \rho) > 0$ in H0 is a basic requirement of the Bayesian philosophy, under which the optimal filter $E[\phi(x_t)|y_{1:t}]$, as showed in (8), can exist.*

Remark 4.6 *By the conditions $(\pi_{s|s-1}, \rho) > 0$ and $|\phi(x_s)|^p \rho(y_s|x_s) < \infty$, we have*

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

Remark 4.7 *We list two typical one dimensional noises, i.e., $n_x = n_y = 1$, and analyze the corresponding unbounded functions satisfying condition H2 as follows:*

(i) $p_w(z, s) = O(\exp(-|z|^\nu))$ as $z \to \infty$ with $\nu > 0$; and $\liminf_{|x|\to\infty} \frac{|h(x,s)|}{|x|^{\nu_1}} > 0$ with $\nu_1 > 0$, $s = 1, \ldots, t$. Then it is easy to check that H2 holds for any function ϕ satisfying $\phi(z) = O(|z|^q)$ as $z \to \infty$, where $q \geq 0$. Hence, Theorem 4.3 holds for the underlying model with any finite $p > 1$.

(ii) $p_w(z, s) = \frac{1}{b-a} I_{[a,b]}$ with $a < 0 < b$; and function $h(x, s) \triangleq h_s$ satisfying that the set $h_s^{-1}([y - a, y - b])$ is bounded for any given y, $s = 1, \ldots, t$. Then it is easy to check that H2 holds for any function ϕ. Hence, Theorem 4.3 holds for the underlying model with any finite $p > 1$.

In the multidimensional cases we need only view the absolute value as certain norms in (i) and (ii), and with all variables being corresponding vectors. Then same results still hold.

Denote the set of functions ϕ satisfying H2 by $L_t^p(\rho)$.

Theorem 4.3 *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*

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

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

Proof. The proof is carried out using a framework similar to the one used in proving Theorem 4.1.

1: Initialization

Let $\{x_0^i\}_{i=1}^N$ be independent random variables with the same distribution $\pi_0(dx_0)$. Then, with the use of Lemmas 4.1, 4.2, 4.3 it is clear that

$$\begin{aligned}
E\left|(\pi_0^N, \phi) - (\pi_0, \phi)\right|^p &= \frac{1}{N^p} E\left|\sum_{i=1}^N (\phi(x_0^i) - E[\phi(x_0^i)])\right|^p \\
&\leq \frac{C(p)}{N^p}\left[\sum_{i=1}^N E|\phi(x_0^i) - E[\phi(x_0^i)]|^p + \left[\sum_{i=1}^N E|\phi(x_0^i) - E[\phi(x_0^i)]|^r\right]^{p/r}\right] \\
&\leq 2^p C(p)\left[\frac{E|\phi(x_0^i)|^p}{N^{p-1}} + \frac{E^{p/r}|\phi(x_0^i)|^r}{N^{p(1-1/r)}}\right] \\
&\leq 2^{p+1} C(p) \frac{E|\phi(x_0^i)|^p}{N^{p(1-1/r)}} \triangleq C_{0|0} \frac{\|\phi\|_{0,p}^p}{N^{p(1-1/r)}}.
\end{aligned} \tag{34}$$

Similarly,

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

Hence,

$$E\left|(\pi_0^N, |\phi|^p)\right| \leq 3E|\phi(x_0^i)|^p \triangleq M_{0|0}\|\phi\|_{0,p}^p. \tag{35}$$

2: Prediction

Based on (34) and (35), we assume that for $t-1$ and $\forall \phi \in L_t^p(\rho)$

$$E\left|(\pi_{t-1|t-1}^N, \phi) - (\pi_{t-1|t-1}, \phi)\right|^p \leq C_{t-1|t-1} \frac{\|\phi\|_{t-1,p}^p}{N^{p(1-1/r)}} \tag{36}$$

and

$$E\left|(\pi_{t-1|t-1}^N, |\phi|^p)\right| \leq M_{t-1|t-1}\|\phi\|_{t-1,p}^p \tag{37}$$

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

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

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

where

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

and $\pi_{t-1|t-1}^{N,\alpha_i} = \sum_{j=1}^N \alpha_j^i \delta_{x_{t-1}^j}$. We consider the three terms Π_1, Π_2 and Π_3 separately in the following.

For given $\{x_{t-1}^i, i=1,\ldots,N\}$ and y_t, sample $\bar{x}_t^i$ obeying $(\pi_{t-1|t-1}^{N,\alpha_i}, K)$, $i=1,\ldots,N$. Naturally,

$$E[\phi(\bar{x}_t^i)|\mathcal{F}_{t-1}] = (\pi_{t-1|t-1}^{N,\alpha_i}, K\phi). \tag{38}$$

This means that $\{\bar{x}_t^i,\ i=1,\ldots,N\}$ are particles normally generated without any modification. Clearly, the term Π_2 denotes the difference between the two series of particles. In order to use Lemma 4.5, we analyze a probability first.

In view of (38) and (9), we have

$$E\left[\frac{1}{N}\sum_{i=1}^N \rho(y_t|\bar{x}_t^i)\Big|\mathcal{F}_{t-1}\right] = (\pi_{t-1|t-1}^N, K\rho).$$

Thus,

$$P\left[\frac{1}{N}\sum_{i=1}^N \rho(y_t|\bar{x}_t^i) < \gamma_t\Big|\mathcal{F}_{t-1}\right] = P\left[(\pi_{t-1|t-1}^N, K\rho) < \gamma_t\right]. \tag{39}$$

By (36), we have

$$\begin{aligned}
P\left[(\pi_{t-1|t-1}^N, K\rho) < \gamma_t\right] &= P\left[(\pi_{t-1|t-1}^N, K\rho) - (\pi_{t-1|t-1}, K\rho) < \gamma_t - (\pi_{t-1|t-1}, K\rho)\right] \\
&\le P\left[|(\pi_{t-1|t-1}^N, K\rho) - (\pi_{t-1|t-1}, K\rho)| > |\gamma_t - (\pi_{t-1|t-1}, K\rho)|\right] \\
&\le \frac{E|(\pi_{t-1|t-1}^N, K\rho) - (\pi_{t-1|t-1}, K\rho)|^p}{|\gamma_t - (\pi_{t-1|t-1}, K\rho)|^p} \\
&\le \frac{C_{t-1|t-1}\|K\|^p}{|\gamma_t - (\pi_{t-1|t-1}, K\rho)|^p} \cdot \frac{\|\rho\|_{t-1,p}^p}{N^{p(1-1/r)}} \triangleq C_{\gamma_t} \cdot \frac{\|\rho\|_{t-1,p}^p}{N^{p(1-1/r)}}.
\end{aligned} \tag{40}$$

Obviously, the probability in (40) tends to 0 as $N \to \infty$. Thus, for given $\epsilon_t \in (0,1)$ and sufficiently large N, we have

$$P\left[\frac{1}{N}\sum_{i=1}^N \rho(y_t|\bar{x}_t^i) < \gamma_t\Big|\mathcal{F}_{t-1}\right] < \epsilon_t < 1. \tag{41}$$

By Lemmas 4.1, 4.2, 4.5 (conditional case), (38) and (9),

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

Hence, by Lemma 4.3 and (37),

$$
E|\Pi_1|^p \le \frac{2^{p+1}\|K\|^p M_{t-1|t-1}}{(1-\epsilon_t)^{p/r}} \cdot \frac{\|\phi\|_{t-1,p}^p}{N^{p-p/r}} \triangleq C_{\Pi_1} \cdot \frac{\|\phi\|_{t-1,p}^p}{N^{p-p/r}}. \tag{42}
$$

By (38), Lemma 4.5 and (9),

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

Hence,

$$
E|\Pi_2|^p \le C_{\Pi_2}\|K\| \cdot \frac{\|\phi\|_{t-1,p}^p}{N^{p-p/r}}. \tag{43}
$$

By (9) and (36),

$$
E|\Pi_3|^p \le C_{t-1|t-1}\|K\|^p \cdot \frac{\|\phi\|_{t-1,p}^p}{N^{p-p/r}} \triangleq C_{\Pi_3} \cdot \frac{\|\phi\|_{t-1,p}^p}{N^{p-p/r}}. \tag{44}
$$

Then, using Minkowski's inequality, (42), (43) and (44), we have

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

That is

$$E\left|(\tilde{\pi}^N_{t|t-1},\phi)-(\pi_{t|t-1},\phi)\right|^p \le \tilde{C}_{t|t-1}\frac{\|\phi\|^p_{t-1,p}}{N^{p-p/r}}. \tag{45}$$

Based on (45), we know from Proposition 3.1 that the modified algorithm will not run into a infinite loop.

By Lemma 4.2 and (37)

$$\begin{aligned}
&E\left(E\left|(\tilde{\pi}^N_{t|t-1},|\phi|^p)-\frac{1}{N}\sum_{i=1}^N E\left[|\phi(\tilde{x}^i_t)|^p|\mathcal{F}_{t-1}\right]\,\Big|\mathcal{F}_{t-1}\right|\right)\\
&=\frac{1}{N}E\left(E\left|\sum_{i=1}^N[|\phi(\tilde{x}^i_t)|^p-E(|\phi(\tilde{x}^i_t)|^p|\mathcal{F}_{t-1})]\right|\right)\\
&\le\frac{1}{(1-\epsilon_t)N}E\left(E\left[\sum_{i=1}^N[|\phi(\bar{x}^i_t)|^p+E(|\phi(\bar{x}^i_t)|^p|\mathcal{F}_{t-1})]\right]\right)\\
&\le\frac{2}{1-\epsilon_t}E(\pi^N_{t-1|t-1},K|\phi|^p)\le\frac{2}{1-\epsilon_t}\|K\|^pM_{t-1|t-1}\|\phi\|^p_{t-1,p}.
\end{aligned}\tag{46}$$

By (38), Lemma 4.5 and (9),

$$\begin{aligned}
&\left|\frac{1}{N}\sum_{i=1}^N E\left[|\phi(\tilde{x}^i_t)|^p|\mathcal{F}_{t-1}\right]-\frac{1}{N}\sum_{i=1}^N E\left[|\phi(\bar{x}^i_t)|^p|\mathcal{F}_{t-1}\right]\right|\\
&=\left|\frac{1}{N}\sum_{i=1}^N\left(E\left[|\phi(\tilde{x}^i_t)|^p|\mathcal{F}_{t-1}\right]-E\left[|\phi(\bar{x}^i_t)|^p|\mathcal{F}_{t-1}\right]\right)\right|\\
&\le\frac{1}{N}\sum_{i=1}^N\left(E\left[|\phi(\tilde{x}^i_t)|^p|\mathcal{F}_{t-1}\right]+E\left[|\phi(\bar{x}^i_t)|^p|\mathcal{F}_{t-1}\right]\right)\\
&\le\left(\frac{1}{1-\epsilon_t}+1\right)\cdot\frac{1}{N}\sum_{i=1}^N(\pi^{N,\alpha_i}_{t-1|t-1},K|\phi|^p)\\
&=\frac{2-\epsilon_t}{1-\epsilon_t}\cdot(\pi^N_{t-1|t-1},K|\phi|^p)\\
&\le\frac{2-\epsilon_t}{1-\epsilon_t}\cdot\|K\|^pM_{t-1|t-1}\|\phi\|^p_{t-1,p}.
\end{aligned}\tag{47}$$

By (37),

$$\left|\frac{1}{N}\sum_{i=1}^N(\pi^{N,\alpha_i}_{t-1|t-1},K|\phi|^p)-(\pi_{t|t-1},|\phi|^p)\right|\le 2\|K\|^pM_{t-1|t-1}\|\phi\|^p_{t-1,p}. \tag{48}$$

Then, by (46) (47) (48), we have

$$E\left|(\tilde{\pi}^N_{t|t-1},|\phi|^p)-(\pi_{t|t-1},|\phi|^p)\right| \le \left(\frac{4-\epsilon_t}{1-\epsilon_t}+2\right)\|K\|^p M_{t-1|t-1}\|\phi\|^p_{t-1,p} \triangleq \tilde{M}_{t|t-1}\|\phi\|^p_{t-1,p}. \tag{49}$$

3: Update

In this step we go step further to analyse $E\left|(\tilde{\pi}^N_{t|t},\phi)-(\pi_{t|t},\phi)\right|^p$ and $E(\tilde{\pi}^N_{t|t},|\phi|^p)$ based on (45) and (49). Here, we still use the separation $(\tilde{\pi}^N_{t|t},\phi)-(\pi_{t|t},\phi)=\tilde{\Pi}_1+\tilde{\Pi}_2$, which was introduced in the step (3) in the proof of Theorem 4.1. By condition H1 and the modified version of the algorithm we have,

$$|\tilde{\Pi}_1| = \left|\frac{(\tilde{\pi}^N_{t|t-1},\rho\phi)}{(\tilde{\pi}^N_{t|t-1},\rho)}\cdot\frac{[(\pi_{t|t-1},\rho)-(\tilde{\pi}^N_{t|t-1},\rho)]}{(\pi_{t|t-1},\rho)}\right| \le \frac{\|\rho\phi\|}{\gamma_t(\pi_{t|t-1},\rho)}\left|(\pi_{t|t-1},\rho)-(\tilde{\pi}^N_{t|t-1},\rho)\right|.$$

Thus, by Minkowski's inequality and (45),

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

which implies

$$E\left|(\tilde{\pi}^N_{t|t},\phi)-(\pi_{t|t},\phi)\right|^p \le \tilde{C}_{t|t}\frac{\|\phi\|^p_{t-1,p}}{N^{p-p/r}}. \tag{50}$$

Using a separation similar to the one mentioned above, by (49),

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

Observe that $\|\phi\|_{s,p}$ is increasing with respect to s,

$$\begin{aligned}
E\left|(\tilde{\pi}^N_{t|t},|\phi|^p)\right| &\le \frac{\tilde{M}_{t|t-1}\|\rho\|\left(\|\rho\phi^p\|+\gamma_t\right)}{\gamma_t(\pi_{t|t-1},\rho)}\cdot\|\phi\|^p_{t-1,p}+(\pi_{t|t},|\phi|^p), \\
&\le \left(\frac{\tilde{M}_{t|t-1}\|\rho\|\left(\|\rho\phi^p\|+\gamma_t\right)}{\gamma_t(\pi_{t|t-1},\rho)}+1\right)\cdot\|\phi\|^p_{t,p} \triangleq \tilde{M}_{t|t}\|\phi\|^p_{t,p}.
\end{aligned} \tag{51}$$

4: Resampling

Finally, we analyse $E\left|(\pi^N_{t|t},\phi)-(\pi_{t|t},\phi)\right|^p$ and $E(\pi^N_{t|t},|\phi|^p)$ based on (50) and (51).

Again, we use the separation $(\pi^N_{t|t}, \phi) - (\pi_{t|t}, \phi) = \bar{\Pi}_1 + \bar{\Pi}_2$ and the σ-algebra $\mathcal{G}_t$, which was introduced in step (4) in the proof of Theorem 4.1.

Then, by Lemmas 4.1, 4.2,

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

Thus, by Lemma 4.3 and (51),

$$
E|\bar{\Pi}_1|^p \leq 2^{p+1} C(p)\tilde{M}_{t|t}\frac{\|\phi\|^p_{t,p}}{N^{p(1-1/r)}}. \tag{52}
$$

Then by Minkowski's inequality, (50) and (52)

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

That is

$$
E\left|(\pi^N_{t|t}, \phi) - (\pi_{t|t}, \phi)\right|^p \leq C_{t|t}\frac{\|\phi\|^p_{t,p}}{N^{p-p/r}}. \tag{53}
$$

Using a separation similar to the one mentioned above, by (51),

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

Hence,

$$
E\left|(\pi^N_{t|t}, |\phi|^p)\right| \leq (3\tilde{M}_{t|t} + 2)\|\phi\|^p_{t,p} \triangleq M_{t|t}\|\phi\|^p_{t,p}. \tag{54}
$$

Therefore, the proof of Theorem 4.3 is completed, since (36) and (37) are successfully replaced by (53) and (54).

□

Similar to Theorem 4.2, by Borel-Cantelli Lemma, we have a weak convergence result as follow.

Theorem 4.4 *In addition to H1 and H2, if $p > 2$, then for any function $\phi \in L^p_t(\rho)$, $\lim_{N\to\infty}(\pi^N_{t|t}, \phi) = (\pi_{t|t}, \phi)$ almost surely.*

5 Conclusions

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 derived a new Rosenthal type inequality and provided slightly extended convergence results when it comes to bounded functions.

6 Acknowledgements

This work was supported by the strategic research center MOVIII, funded by the Swedish Foundation for Strategic Research, SSF.

Avdelning, Institution Division, Department Division of Automatic Control Department of Electrical Engineering	**Datum** Date 2009-08-21

Språk Language ☐ Svenska/Swedish ☒ Engelska/English ☐ ______________	**Rapporttyp** Report category ☐ Licentiatavhandling ☐ Examensarbete ☐ C-uppsats ☐ D-uppsats ☒ Övrig rapport ☐ ______________	**ISBN** — **ISRN** — **Serietitel och serienummer** Title of series, numbering **ISSN** 1400-3902 LiTH-ISY-R-2914

URL för elektronisk version

http://www.control.isy.liu.se

Titel / Title: Basic Convergence Results for Particle Filtering Methods: Theory for the Users

Författare / Author: Xiao-Li Hu, Thomas B. Schön, Lennart Ljung

Sammanfattning
Abstract

This work extends our recent work on proving that the particle filter converge for unbounded function to a more general case. More specifically, we prove that the particle filter converge for unbounded functions in the sense of L^p-convergence, for an arbitrary $p \geq 2$. Related to this, we also provide proofs for the case when the function we are estimating is bounded. In the process of deriving the main result we also established a new Rosenthal type inequality.

Nyckelord
Keywords: Convergence, particle filter, nonlinear filtering, dynamic systems

References

[1] D. L. Burkholder, Distribution function inequalities for martingales, Ann. Probab., 1, 19-42, 1973.

[2] D. Crisan, A. Doucet, A Survey of Convergence Results on Particle Filtering Methods for Practitioners, IEEE Trans. Signal Processing, vol. 50, no. 3, pp. 736-746, 2002.

[3] D. Crisan and M. Grunwald, Large Deviation Comparison of Branching Algorithms versus Resampling Algorithms: Application to Discrete Time Stochastic Filtering, Statist. Lab., Cambridge University, Cambridge, U.K., Tech. Rep., TR1999-9, 1999.

[4] P. Del Moral, Non-linear Filtering: Interacting Particle Solution, Markov Processes and Related Fields, Volume 2, Number 4, 555–580, 1996.

[5] P. Del Moral and A. Guionnet, Large Deviations for Interacting Particle Systems: Applications to Non-Linear Filtering Problems, Stochastic Processes and their Applications, 78, 69-95, 1998.

[6] P. Del Moral and A. Guionnet, A Central Limit Theorem for Non Linear Filtering using Interacting Particle Systems, Annals of Applied Probability, Vol. 9, No. 2, 275-297, 1999.

[7] P. Del Moral, L. Miclo, Branching and Interacting Particle Systems Approximations of Feynman-Kac Formulae with Applications to Non-Linear Filtering Seminaire de Probabilites XXXIV, Ed. J. Azema and M. Emery and M. Ledoux and M. Yor, Lecture Notes in Mathematics, Springer-Verlag Berlin, Vol. 1729, 1-145, 2000.

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

[9] A. Doucet, S. J. Godsill , and C. Andrieu, On sequential Monte Carlo sampling methods for Bayesian filtering, Statist. Comp., 10:197-208, 2000.

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

[11] W. Hardle, G. Kerkyacharian, D. Picard, A. Tsybakov, Wavelet, Approximation and Statistical Applications, Lectures Notes in Statistics 129, Springer Verlag, New York, 1998.

[12] P. Hitczenko, Best constants in martingale version of Rosenthal's inequality, Ann. Probab., 18, no. 4, 1656-1668, 1990.

[13] X.-L. Hu, T. B. Schön and L. Ljung, A Basic Convergence Result for Particle Filtering, Submitted to IEEE Transactions on Signal Processing, 2007.

[14] W. B. Johnson, G. Schechtman and J. Zinn, Best constants in moment inequalities for linear combination of independent and exchangeable random variables, Ann. Probab., 13, 234-253, 1985.

[15] H. R. Kunsch, Recursive Monte Carlo Filters: Algorithms and Theoretical Analysis, Annals of Statistics, 33, no. 5, 1983-2021, 2005.

[16] F. Le Gland and N. Oudjane, Stability and uniform approximation of nonlinear filters using the Hilbert metric, and application to particle filters, Research report RR-4215, INRIA., June, 2001.

[17] H. P. Rosenthal, On the subspaces of L^p $(p > 2)$ spanned by sequences of independent. random variables, Israel J. Math., 8, No. 3, 273-303, 1970.

[18] T. B. Schön, Estimation of Nonlinear Dynamic Systems - Theory and Applications, PhD Thesis, Linköping University, Linköping, Sweden, 2006.