A BASIC CONVERGENCE RESULT FOR PARTICLE FILTERING

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

* *Institute of Systems Science*
Chinese Academy of Sciences
Beijing, 100080, P.R. China
e-mail: xlhu@amss.ac.cn
** *Division of Automatic Control*
Linköpings universitet
SE-581 83 Linköping, Sweden
e-mail: {schon, ljung}@isy.liu.se

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

Keywords: Nonlinear filters, particle filter, convergence, dynamic systems

1. INTRODUCTION

The nonlinear filtering problem is formulated as follows. Consider the system with state x_t, input u_t and output y_t:

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

where v_t and e_t are sequences of independent random variables. The inputs and outputs are observed for $t = 1, 2, \dots$ and the problem is to estimate the state based on these observations.

We will in this contribution assume that the input u is a deterministic sequence, so we could as well subsume the input in a time-varying dynamics:

$$x_{t+1} = f(x_t, t) + v_t, \tag{2a}$$
$$y_t = h(x_t, t) + e_t, \tag{2b}$$

Let $p_v(\cdot, t)$ and $p_e(\cdot, t)$ denote the probability density function for the noise v and e, respectively. The, in many respects, optimal estimate at time t is the conditional expectation

$$\hat{x}_t = \mathrm{E}(x_t | y_{1:t}), \tag{3}$$

where $y_{1:t} \triangleq (y_1, \dots, y_t)$.

Now, while there exist several formulas that characterize the (posterior) distribution of $\hat{x}_t$ (Jazwinski 1970) it is well known that except in some quite special cases it is not possible to compute $\hat{x}_t$ with finite computations. This has lead to a large number of approximation methods, like extended Kalman filtering, Gaussian sum approximations, point-mass filters, etc., see e.g., (Jazwinski 1970, Sorenson and Alspach 1971, Bucy and Senne 1971). Recently there has been considerable interest in a certain approximation technique based on Monte Carlo methods, usually called *Particle Filter*, (Gordon *et al.* 1993, Doucet *et al.* 2000, Doucet *et al.* 2001). A basic particle filter will be defined in the next section, but in short the main idea is to generate many random instances ('particles') of x that follow (2) and promote those that are in good accordance with the observed y. We shall denote the particle filter estimate that is based on N particles by

$$\hat{x}_t^N. \tag{4}$$

Clearly, it is desired that $\hat{x}_t$ and $\hat{x}_t^N$ are close and that the distance tends to zero as N tends to infinity. There are many papers dealing with such analysis, see e.g., the excellent survey (Crisan and Doucet 2002) and the recent book (Del Moral 2004), but they mostly deal with an estimate like

$$\mathrm{E}(\phi(x_t)|y_{1:t}), \tag{5}$$

where $\phi : \mathcal{R}^{n_x} \to \mathcal{R}$ is a *bounded* scalar-valued function. We are not aware of any convergence results for unbounded ϕ (such as $\phi(x) = x^{[i]}$ for component i of x) that are applicable to implemented particle filters. The main result of this paper is a theorem showing particle filter convergence for unbounded functions ϕ.

For a more complete picture of the information about x_t, it is natural to consider the *posterior density* of x_t, given $\{y_1, \ldots, y_t\}$. This will be denoted by

$$p(x_t|y_{1:t}). \tag{6}$$

Clearly, $\hat{x}_t$ is the mean of this posterior density.

The propagation of the posterior density is the key tool for the estimation. It is well known, and follows from Bayes' theorem, (e.g., (Jazwinski 1970)) that

$$p(x_t|y_{1:t}) = \frac{p(y_t|x_t)p(x_t|y_{1:t-1})}{\int p(y_t|x_t)p(x_t|y_{1:t-1})dx_t}, \tag{7a}$$

$$p(x_t|y_{1:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1}. \tag{7b}$$

2. A PARTICLE FILTER

We are trying to compute the estimate (3),

$$\hat{x}_t = \int x_t p(x_t|y_{1:t})dx_t. \tag{8}$$

The particle filter can help us in doing this since it provides an estimate of the filtering probability density function (pdf) $p(x_t|y_{1:t})$. One way of representing a pdf is by using a set of samples $\{\tilde{x}_t^i\}_{i=1}^N$ and corresponding weights $\{\tilde{q}_t^i\}_{i=1}^N$ according to

$$\tilde{p}_N(x_t|y_{1:t}) = \sum_{i=1}^N \tilde{q}_t^i \delta(x_t - \tilde{x}_t^i). \tag{9}$$

Each sample $\tilde{x}_t^i$ describes a realization of the state and the associated weight $\tilde{q}_t^i$ tells us how good this realization is, given the information in the measurement. We can also obtain an alternative approximate representation of the pdf by *resampling* the samples according to their weights, resulting in an unweighted approximation

$$p_N(x_t|y_{1:t}) = \sum_{i=1}^N \frac{1}{N}\delta(x_t - x_t^i). \tag{10}$$

In this resampling step the samples with high weights have been replicated many times, whereas the samples with low weights have possibly been neglected. Intuitively this will provide a similar approximation of the pdf, see e.g., (Hol *et al.* 2006) for more details on various resampling algorithms.

Note that the samples x_t^i are typically referred to as *particles*, hence the name particle filter. It is worth noting that the particle filter can be used to compute many different estimates, not just the point estimate (3), due to the fact that it approximates the filtering pdf. We will now very briefly outline how the particle filter works, for a more thorough explanation, see e.g., (Doucet *et al.* 2000, Doucet *et al.* 2001, Schön 2006), where the latter contains a MATLAB implementation of the particle filter and an illustration using the example we discuss in Section 4.

Let us assume that we are given a set of unweighted particles $\{x_{t-1}^i\}_{i=1}^N$ describing N approximate realizations from a stochastic variable with pdf $p(x_{t-1}|y_{1:t-1})$. Using the dynamics of our system (2a), we can produce particles $\{\tilde{x}_t^i\}_{i=1}^N$ being approximate realizations of $p(x_t|y_{1:t-1})$ simply by propagating the particles though the dynamics,

$$\tilde{x}_t^i = f(x_{t-1}^i, t-1) + v_{t-1}^i, \tag{11}$$

where v_{t-1}^i denotes a realization from the process noise v_{t-1}. The information in the new measurement y_t can now be incorporated into the approximation by inserting $\tilde{p}_N(x_t|y_{1:t-1}) = \sum_{i=1}^N \frac{1}{N}\delta(x_t - \tilde{x}_t^i)$ from (11) into (7a), resulting in

$$p_N(x_t|y_{1:t}) = \sum_{i=1}^N \underbrace{\frac{p(y_t|\tilde{x}_t^i)}{\sum_{j=1}^N p(y_t|\tilde{x}_t^j)}}_{\tilde{q}_t^i} \delta(x_t - \tilde{x}_t^i),$$

where we have defined the so called normalized *importance weights* $\{\tilde{q}_t^i\}_{i=1}^N$. In order to help intuition it is instructive to note that $\tilde{q}_t^i$ reveals how likely particle i is given the information in the present measurement y_t. This information is used in the essential resampling step, to generate a new set of unweighted particles $\{x_t^i\}_{i=1}^N$. The resampling step was first introduced in (Gordon *et al.* 1993). Without this step the algorithm will rapidly diverge. All steps in the particle filter, save for the resampling step, have been known since the end of the 1940's (Metropolis and Ulam 1949). To sum up, we have the following algorithm.

Algorithm 1. Particle filter

(1) Initialize the particles, $\{x_0^i\}_{i=1}^N$ distributed according to $p(x_0)$.
(2) Time update: predict new particles $\tilde{x}_t^i$ by drawing new samples according to (11).
(3) Measurement update: compute the importance weights $\{q_t^i\}_{i=1}^N$,

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

and normalize $\tilde{q}_t^i = q_t^i / \sum_{j=1}^N q_t^j$
(4) Resampling: draw N particles, with replacement, for each $i = 1, \ldots, N$

$$\Pr(x_t^i = \tilde{x}_t^j) = \tilde{q}_t^j, \quad j = 1, \ldots, N$$

(5) Set $t := t + 1$ and iterate from step 2.

The estimate (3) is computed after step 3 in Algorithm 1, by inserting (9) into (8)

$$\hat{x}_t^N = \sum_{i=1}^N \tilde{q}_t^i \tilde{x}_t^i. \tag{12}$$

In order to prove the basic convergence result for $\hat{x}_t^N$ we shall consider a particle filter with the following modification,

$$\sum_{i=1}^{N} p(y_t|\tilde{x}_t^i) \geq \gamma_t > 0 \qquad \forall t, \tag{13}$$

that is suggested by technical requirements in the proof. It is also interesting to note that (13) has a strong support from an intuitive point of view as well. The modification requires that the sum of the likelihoods $p(y_t|\tilde{x}_t^i)$ is greater than γ_t, where the likelihood explains how probable a certain measurement y_t is given the current state $\tilde{x}_t^i$. Hence, if the sum of likelihoods is low that means that the current states are not able to explain the measurements. Now, condition (13) is checked after step 3 in Algorithm 1. If it is not fulfilled the particles are repropagated according to step 2 and the condition is checked again. It can be shown that

$$P\left(\sum_{i=1}^{N} p(y_t|\tilde{x}_t^i) \geq \gamma_t\right) \xrightarrow[N\to\infty]{} 1, \tag{14}$$

implying that the influence of the algorithm modification decreases as the number of particles increases. In fact, this means that the lower bound for (13) is almost always satisfied, provided that N is sufficiently large and γ_t is suitably chosen.

This modification also turns out to reduce the degeneracy of importance weights (see, e.g., (Crisan and Doucet 2002, Legland and Oudjane 2001)). Hence, we can expect a better performance in practise. The implications of this modification will be studied in Section 4.

3. THE MAIN RESULT

The optimal estimator $\hat{x}_t$ is a random variable, being a function of past outputs $y_s, s \leq t$. Furthermore, the particle filter estimator $\hat{x}_t^N$ is a random variable that depends also on the randomly drawn particles and the random propagation in (11). We will consider the random properties of the particle filter estimated in the latter respect, i.e., what happens with averaging over the particle samples, keeping the conditioning with respect to the past outputs.

The main convergence result is as follows:

Theorem 1. Consider the system (2), and assume that the joint probability density function of $y_s, s = 1, \dots, t$ exists. Assume that

$$\sup_{x_s} \left(|x_s|^4 p_e(y_s - h(x_s), s)\right) < \infty, \quad \forall y_s, s \leq t \tag{15}$$

where $p_e(\cdot, s)$ is the pdf of e_s. Let $\hat{x}_t$ be defined by (3), and let the particle filter estimate $\hat{x}_t^N$ be defined by (12). Then for almost all observation records $y_s, s \leq t$,

$$\mathrm{E}|\hat{x}_t - \hat{x}_t^N|^4 = \mathcal{O}\left(\frac{1}{N^2}\right), \tag{16a}$$

$$\hat{x}_t - \hat{x}_t^N \to 0 \; w.p.1 \tag{16b}$$

as the number of particles N tends to infinity.

PROOF. See Appendix A.

Remark: In (16), the observation record $y_s, s \leq t$ is fixed, and the probabilistic quantifiers "E" and "w.p.1" refer to the probability space of the particle filter algorithm, i.e. the random propagation in (11) and the resampling in step (4).

4. NUMERICAL ILLUSTRATION

In order to illustrate the impact of the algorithm modification (13), implied by the convergence proof, we study the following nonlinear time-varying system,

$$x_{t+1} = \frac{x_t}{2} + \frac{25x_t}{1+x_t^2} + 8\cos(1.2t) + v_t, \tag{17a}$$

$$y_t = \frac{x_t^2}{20} + e_t, \tag{17b}$$

where $v_t \sim \mathcal{N}(0, 10), e_t \sim \mathcal{N}(0, 1)$, the initial state $x_0 \sim \mathcal{N}(0, 5)$ and $\gamma_t = 10^{-4}$. In the experiment we used 250 time instants and 500 experiments, all using the same measurement sequence. We used the particle filter given in Algorithm 1 modified according to (13) in order to compute an approximation of the estimate $\hat{x}_t = \mathrm{E}(x_t|y_{1:t})$. In accordance with both Theorem 1 and intuition the quality of the estimate improves with the number of particles used in the approximation. The purpose of the present experiment is to illustrate that the algorithm modification (13) is only active when a small amount of particles is used. That this is indeed the case is evident from Figure 1, where the average number of interventions due to violations of (13) are given as a function of the number of particles used in the filter.

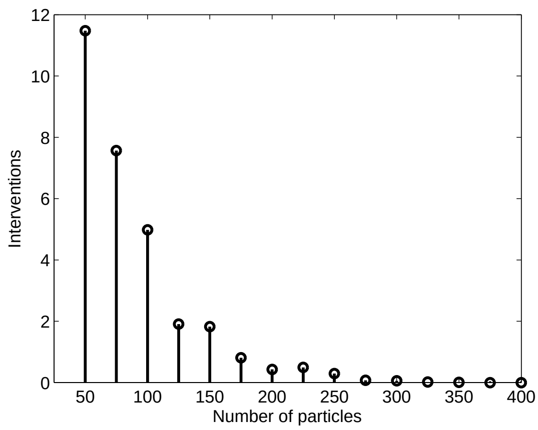

Fig. 1. *Illustration of the impact of the algorithm modification* (13). *The figure shows the number of times* (13) *was violated as a function of the number of particles used. Note that it is the average result from* 500 *experiments.*

5. CONCLUDING COMMENTS

In this contribution we have focused on properties of the optimal state estimate $\hat{x}_t$ and its relation to the particle filter estimate $\hat{x}_t^N$. Clearly this really concerns properties of the posterior density $p(x_t|y_{1:t})$ and the empirical density of the particles, $p_N(x_t|y_{1:t})$. The proof deals with the closeness of these densities. It is clear that more general statements concerning these densities can be made from the building blocks of the proof, like convergence and closeness on general functions of this density, like

$$\int \phi(x_t)p(x_t|y_{1:t})dx_t. \tag{18}$$

We refer to (Hu *et al.* 2007) for more general statements and discussion of this kind.

The basic contribution of this paper has been the extension of such convergence results to unbounded functions ϕ, which has allowed statements on the filter estimate (conditional expectation) itself. We have had to introduce a slight modification of the particle filter (eqn. (13)) in order to complete the proof. It is an interesting question to study if the modification is in fact necessary, and leads also to improved behaviour in practise. The simulation study showed that the effect of the modification decreases with increased number of particles.

ACKNOWLEDGEMENT

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

Appendix A. PROOF OF THEOREM 1

We will first establish a general result (Proposition 1) for a more general algorithm, including the one proposed in Section 2, for the proof of Theorem 1. Let $(\Omega, \mathcal{F}, P)$ be a probability space on which the n_x-dimensional state of system described by the Markov process $X = \{X_t, t = 0, 1, \ldots\}$ with transition kernel density $K(x_t|x_{t-1})$ and the n_y-dimensional observation described by $Y = \{Y_t, t = 1, 2, \ldots\}$ with observation density $\rho(y_t|x_t)$. Obviously, for the system (2) we have

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

For convenience, let us write the traditional form of the particle filter algorithm in a somewhat more abstract way,

Algorithm 2. Abstract particle filter

(1) $x_0^i \sim \pi_0(dx_0)$, $i = 1, \ldots, N$.
(2) $\tilde{x}_t^i \sim \sum_{j=1}^N \alpha_j^i K(dx_t|x_{t-1}^j)$, $i = 1, \ldots, N$.
(3) $\tilde{\pi}_{t|t}^N(dx_t) = \sum_{i=1}^N w_t^i \delta(\tilde{x}_t^i - dx_t)$,
$w_t^i = \frac{\rho(y_t|\tilde{x}_t^i)}{\sum_{i=1}^N \rho(y_t|\tilde{x}_t^i)}$.
(4) $x_t^i \sim \tilde{\pi}_{t|t}^N(dx_t)$, $i = 1, \ldots, N$.
$\pi_{t|t}^N(dx_t) = \frac{1}{N}\sum_{i=1}^N \delta(x_t^i - dx_t)$.

Denote

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

For convenience, let us introduce some more notations. Given a measure ν, a function ϕ, denote

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

Hence, $\mathrm{E}(\phi(x_t)|y_{1:t}) = (\pi_{t|t}, \phi)$.

Remark 1. When $\alpha_j^i = 1$ for $j = i$, and $\alpha_j^i = 0$ for $j \neq i$, Algorithm 2 is reduced to the traditional Algorithm 1, as introduced in Section 2, see e.g., (Gordon *et al.* 1993, Doucet *et al.* 2000, Schön 2006). When $\alpha_j^i = 1/N$ for all i and j, it turns out to be a convenient form for theoretical treatment, as introduced by nearly all authors dealing with theoretical analysis, for example (Crisan and Doucet 2002, Del Moral 1996, Del Moral and Miclo 2000, Del Moral 2004). A property is followed by the selection of α_j^i:

$$\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N \alpha_j^i K(dx_t|x_{t-1}^j) = (\pi_{t-1|t-1}^N, K). \tag{A.1}$$

The convergence results are all given for a fixed observation record, which means $\mathrm{E}(\cdot) = \mathrm{E}(\cdot|y_{1:t})$. We need the following conditions to establish the general result.

H1. $(\pi_{s|s-1}, \rho) > 0$, $\rho(y_s|x_s) < \infty$; $K(x_s|x_{s-1}) < \infty$ for given $y_{1:s}$, $s = 1, 2, \ldots, t$.

H2. The function $\phi : \mathcal{R}^{n_x} \to \mathcal{R}$ satisfy

$$\sup_{x_s} |\phi(x_s)|^4 \rho(y_s|x_s) < \infty$$

for given $y_{1:s}$, $s = 1, \ldots, t$.

Remark 2. In view of (7a), clearly, $(\pi_{s|s-1}, \rho) > 0$ in H1 is a basic requirement of Bayesian philosophy, under which the optimal filter $\mathrm{E}(\phi(x_t)|y_{1:t})$ will exist.

Remark 3. From the conditions $(\pi_{s|s-1}, \rho) > 0$ and $|\phi(x_s)|^4\rho(y_s|x_s) < \infty$, we have

$$(\pi_{s|s}, |\phi|^4) = \frac{(\pi_{s|s-1}, \rho|\phi|^4)}{(\pi_{s|s-1}, \rho)} < \infty.$$

Let us denote the set of functions $\phi : \mathcal{R}^{n_x} \to \mathcal{R}$ satisfying H2 by $L_t^4(\rho)$.

Proposition 1. If H1 and H2 hold, then for any $\phi \in L_t^4(\rho)$, there exists a constant $C_{t|t}$, independent of N, such that

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

where

$$\|\phi\|_{t,4} \triangleq \max\left\{1,\ (\pi_{s|s}, |\phi|^4)^{1/4},\ s = 0, 1, \ldots, t\right\}$$

and $\pi_{s|s}^N$ is generated by the modified version of particle filter algorithm.

By Borel-Cantelli Lemma, we have the following corollary.

Corollary 1. If H1 and H2 hold, then for any $\phi \in L_t^4(\rho)$, $\lim_{N\to\infty}(\pi_{t|t}^N, \phi) = (\pi_{t|t}, \phi)$ almost surely.

Proposition 2. If the joint pdf of $y_s, s = 1, \dots, t$ exist, $(\pi_{s|s-1}, \rho) > 0, s = 1, \dots, t$ hold for almost all observation record $\{y_s\}_{s=1}^t$, i.e., the exception is with probability 0.

For the proof of Proposition 1 we refer to (Hu *et al.* 2007).

Based on Propositions 1 and 2 and Corollary 1, Theorem 1 follows directly. We prove Proposition 1 in the following.

Before proving Proposition 1, we list some simple lemmas which we need in the proof of Proposition 1. It is worth noticing that Lemmas 1 and 4 still holds for the case of conditional independence, which is actually used in the proof of Proposition 1.

Lemma 1. Let $\{\xi_i,\ i = 1, \dots, n\}$ be independent random variables such that $\mathrm{E}\xi_i = 0$, $\mathrm{E}\xi_i^4 < \infty$. Then

$$\mathrm{E}\left|\sum_{i=1}^n \xi_i\right|^4 \le \sum_{i=1}^n \mathrm{E}\xi_i^4 + \left(\sum_{i=1}^n \mathrm{E}\xi_i^2\right)^2. \qquad \text{(A.3)}$$

Lemma 2. If $\mathrm{E}|\xi|^p < \infty$, then $\mathrm{E}|\xi - \mathrm{E}\xi|^p \le 2^p\mathrm{E}|\xi|^p$, for any $p \ge 1$.

Lemma 3. If $1 \le r_1 \le r_2$ and $\mathrm{E}|\xi|^{r_2} < \infty$, then $\mathrm{E}^{1/r_1}|\xi|^{r_1} \le \mathrm{E}^{1/r_2}|\xi|^{r_2}$.

Based on Lemmas 1 and 3, we have

Lemma 4. Let $\{\xi_i,\ i = 1, \dots, n\}$ be independent random variables such that $\mathrm{E}\xi_i = 0$, $\mathrm{E}|\xi_i|^4 < \infty$. Then

$$\mathrm{E}\left|\frac{1}{n}\sum_{i=1}^n \xi_i\right|^4 \le \frac{2\max_{1\le i\le n}\mathrm{E}\xi_i^4}{n^2}. \qquad \text{(A.4)}$$

Denote $\|\rho(x)\| = \max\{1, \sup|\rho(x)|\}$. Then $\|\rho\phi\|_{0,4} \le \|\rho\| \cdot \|\phi\|_{0,4}$.

Proof of Proposition 1. The proof is carried out using mathematical induction.

(1). 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 and 2, it is clear that

$$\mathrm{E}\left|(\pi_0^N, \phi) - (\pi_0, \phi)\right|^4 \le \frac{2}{N^2}\mathrm{E}|\phi(x_0^i) - \mathrm{E}[\phi(x_0^i)]|^4$$
$$\le \frac{32}{N^2}\|\phi\|_{0,4}^4 \triangleq C_{0|0}\frac{\|\phi\|_{0,4}^4}{N^2}. \qquad \text{(A.5)}$$

Clearly,

$$\mathrm{E}\left|(\pi_0^N, |\phi|^4)\right| \le 3\mathrm{E}|\phi(x_0^i)|^4 \triangleq M_{0|0}\|\phi\|_{0,4}^4. \qquad \text{(A.6)}$$

(2). Based on (A.5) and (A.6), we assume that for $t-1$ and $\forall\phi \in L_t^4(\rho)$

$$\mathrm{E}\left|(\pi_{t-1|t-1}^N, \phi) - (\pi_{t-1|t-1}, \phi)\right|^4 \le \frac{C_{t-1|t-1}\|\phi\|_{t-1,4}^4}{N^2} \qquad \text{(A.7)}$$

and

$$\mathrm{E}\left|(\pi_{t-1|t-1}^N, |\phi|^4)\right| \le M_{t-1|t-1}\|\phi\|_{t-1,4}^4 \qquad \text{(A.8)}$$

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

Notice that

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

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 $\pi_{t-1|t-1}^{N,\alpha_i} = \sum_{j=1}^N \alpha_j^i\delta(x_{t-1}^j - dx_{t-1})$. Let us now investigate Π_1 and Π_2.

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

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

Thus, by Lemmas 1, 2, 3, (A.1) and (A.8),

$$\mathrm{E}|\Pi_1|^4 \le 2^5\frac{\|K\|^4 M_{t-1|t-1}\|\phi\|_{t-1,4}^4}{N^2}. \qquad \text{(A.9)}$$

Furthermore, by (A.1) and (A.7),

$$\mathrm{E}|\Pi_2|^4 \le \frac{C_{t-1|t-1}\|K\|^4\|\phi\|_{t-1,4}^4}{N^2}. \qquad \text{(A.10)}$$

Then, using Minkowski's inequality, (A.1), (A.9), and (A.10), we have

$$\mathrm{E}^{1/4}\left|(\tilde{\pi}_{t|t-1}^N, \phi) - (\pi_{t|t-1}, \phi)\right|^4$$
$$\le \mathrm{E}^{1/4}|\Pi_1|^4 + \mathrm{E}^{1/4}|\Pi_2|^4$$
$$\le \|K\|\left([2^5 M_{t-1|t-1}]^{1/4} + C_{t-1|t-1}^{1/4}\right)\frac{\|\phi\|_{t-1,4}}{N^{1/2}}$$
$$\triangleq \tilde{C}_{t|t-1}^{1/4}\frac{\|\phi\|_{t-1,4}}{N^{1/2}}.$$

That is

$$\mathrm{E}\left|(\tilde{\pi}_{t|t-1}^N, \phi) - (\pi_{t|t-1}, \phi)\right|^4 \le \tilde{C}_{t|t-1}\frac{\|\phi\|_{t-1,4}^4}{N^2}. \qquad \text{(A.11)}$$

By Lemma 2, (A.8) and the use of a separation, similar to the one employed above, we have

$$\mathrm{E}\left|(\tilde{\pi}_{t|t-1}^N, |\phi|^4) - (\pi_{t|t-1}, |\phi|^4)\right|$$
$$\le \|K\|^4(3M_{t-1|t-1} + 1)\|\phi\|_{t-1,4}^4 \triangleq \tilde{M}_{t|t-1}\|\phi\|_{t-1,4}^4. \qquad \text{(A.12)}$$

(3). In this step we analyse $\mathrm{E}\left|(\tilde{\pi}_{t|t}^N, \phi) - (\pi_{t|t}, \phi)\right|^4$ and $\mathrm{E}(\tilde{\pi}_{t|t}^N, |\phi|^4)$ based on (A.11) and (A.12). Clearly,

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

where

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

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

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

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

Thus, by Minkowski's inequality and (A.11),

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

which implies

$$\mathrm{E} \left| (\tilde{\pi}^N_{t|t}, \phi) - (\pi_{t|t}, \phi) \right|^4 \le \tilde{C}_{t|t} \frac{\|\phi\|^4_{t-1,4}}{N^2}. \quad \text{(A.13)}$$

Using a separation similar to the one previously used, by (A.12), and observing that $\|\phi\|_{s,4} \ge 1$ is increasing with respect to s, we have

$$\begin{aligned} &\mathrm{E} \left| (\tilde{\pi}^N_{t|t}, |\phi|^4) \right| \\ &\quad \le 3 \max \left\{ \frac{\|\rho\phi^4\| \cdot 2\|\rho\|}{\gamma_t(\pi_{t|t-1}, \rho)}, \frac{\tilde{M}_{t|t-1} \|\rho\|}{(\pi_{t|t-1}, \rho)}, 1 \right\} \cdot \|\phi\|^4_{t,4} \\ &\quad \triangleq \tilde{M}_{t|t} \|\phi\|^4_{t,4}. \end{aligned} \quad \text{(A.14)}$$

(4). Finally, we analyse $\mathrm{E} \left| (\pi^N_{t|t}, \phi) - (\pi_{t|t}, \phi) \right|^4$ and $\mathrm{E}(\pi^N_{t|t}, |\phi|^4)$ based on (A.13) and (A.14).

Obviously

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

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 $\{\tilde{x}^i_t\}^N_{i=1}$. From the generation of x^i_t, we have,

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

and then

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

Then, by Lemmas 2, 4, and (A.14),

$$\mathrm{E}|\bar{\Pi}_1|^4 \le 2^5 \tilde{M}_{t|t} \frac{\|\phi\|^4_{t,4}}{N^2}. \quad \text{(A.15)}$$

Then by Minkowski's inequality, (A.13) and (A.15)

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

That is

$$\mathrm{E} \left| (\pi^N_{t|t}, \phi) - (\pi_{t|t}, \phi) \right|^4 \le C_{t|t} \frac{\|\phi\|^4_{t,4}}{N^2}. \quad \text{(A.16)}$$

Using a similar separation mentioned above, by (A.14),

$$\mathrm{E} \left| (\pi^N_{t|t}, |\phi|^4) \right| \le (3\tilde{M}_{t|t} + 2) \|\phi\|^4_{t,4} \triangleq M_{t|t} \|\phi\|^4_{t,4}. \quad \text{(A.17)}$$

Therefore, the proof of Proposition 1 is completed, since (A.7) and (A.8) are successfully replaced by (A.16) and (A.17).

□

REFERENCES

Bucy, R. S. and K. D. Senne (1971). Digital synthesis on nonlinear filters. *Automatica* **7**, 287–298.

Crisan, D. and A. Doucet (2002). A survey of convergence results on particle filtering methods for practitioners. *IEEE Transactions on Signal Processing* **50**(3), 736–746.

Del Moral, P. (1996). Non-linear filtering: Interacting particle solution. *Markov processes and related fields* **2**(4), 555–580.

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

Del Moral, P. and L. Miclo (2000). *Branching and Interacting Particle Systems Approximations of Feynman-Kac Formulae with Applications to Non-Linear Filtering*. pp. 1–145. Vol. 1729 of *Lecture Notes in Mathematics*. Springer-Verlag. Berlin, Germany.

Doucet, A., de Freitas, N. and Gordon, N. (Eds.) (2001). *Sequential Monte Carlo Methods in Practice*. Springer Verlag. New York, USA.

Doucet, A., S. J. Godsill and C. Andrieu (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing* **10**(3), 197–208.

Gordon, N. J., D. J. Salmond and A. F. M. Smith (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In: *IEE Proceedings on Radar and Signal Processing*. Vol. 140. pp. 107–113.

Hol, J., T. B. Schön and F. Gustafsson (2006). On resampling algorithms for particle filters. In: *Nonlinear Statistical Signal Processing Workshop*. Cambridge, United Kingdom.

Hu, X.-L., T. B. Schön and L. Ljung (2007). Basic convergences for particle filtering methods: Theory for users. Technical report. Department of Electrical Engineering, Linköping University, Sweden. Linköping, Sweden.

Jazwinski, A. H. (1970). *Stochastic processes and filtering theory*. Mathematics in science and engineering. Academic Press. New York, USA.

Legland, F. and N. Oudjane (2001). Stability and uniform approximation of nonlinear filters using the hilbert metric, and application to particle filters. Technical Report RR–4215. INRIA. Paris, France.

Metropolis, N. and S. Ulam (1949). The Monte Carlo method. *Journal of the American Statistical Association* **44**(247), 335–341.

Schön, T. B. (2006). *Estimation of Nonlinear Dynamic Systems – Theory and Applications*. Dissertations No 998. Department of Electrical Engineering, Linköping University, Sweden.

Sorenson, H. W. and D. L. Alspach (1971). Recursive Bayesian estimation using Gaussian sum. *Automatica* **7**, 465–479.