A Robust Particle Filter for State Estimation – with Convergence Results

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

***Abstract*— Particle filters are becoming increasingly important and useful for state estimation in nonlinear systems. Many filter versions have been suggested, and several results on convergence of filter properties have been reported. However, apparently a result on the convergence of the state estimate itself has been lacking. This contribution describes a general framework for particle filters for state estimation, as well as a robustified filter version. For this version a quite general convergence result is established. In particular, it is proved that the particle filter estimate convergences w.p.1 to the optimal estimate, as the number of particles tends to infinity.**

I. Introduction

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

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

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

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

Note that any deterministic input signal is subsumed in the time-varying dynamics. The most commonly used estimate is the conditional expectation,

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

where $y_{1:t} \triangleq (y_1, \dots, y_t)$ and $\phi:\ \mathbb{R}^{n_x} \to \mathbb{R}$ is the function of the state that we want to estimate. In order to compute (3) we need the filtering probability density function $p(x_t|y_{1:t})$. It is well known that this density function can be expressed using multidimensional integrals [1]. The problem is that these integrals only permits analytical solutions in a few special cases. The most common special case is of course when the model (2) is linear and Gaussian and the solution is then given by the Kalman filter [2]. However, for the more interesting nonlinear/non-Gaussian case we are forced to approximations of some kind. Over the years there has been a large amount of ideas suggested on how to perform these approximations. Here, we will discuss a rather recent and popular family of methods, commonly referred to as *particle filters* (PF) or sequential Monte Carlo methods.

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

T. B. Schön and L. Ljung are with the Division of Automatic Control, Linköpings universitet, SE-581 83 Linköping, Sweden. {schon, ljung}@isy.liu.se

Whenever an approximation is used it is very important to address the issue of its convergence to the true solution and more specifically, under what conditions this convergence is valid. An extensive treatment of the currently existing convergence results can be found in the book [3] and the excellent survey papers [4], [5]. The available results prove convergence of probability measures and only treat bounded functions ϕ, effectively excluding the most commonly used state estimate, the mean value. To the best of our knowledge there are no results available for unbounded functions ϕ. For a precise definition of what is meant by an unbounded function we refer to Section V. The contribution of this paper is to provide a robust particle filter, which provably converges for a rather general class of unbounded functions.

II. Background

This section will provide the necessary background, both on dynamic systems in Section II-A and on the nonlinear filtering problem in Section II-B.

A. Dynamic Systems

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

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

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

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

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

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

$$P(X_{t+1} \in dx_{t+1}|X_t = x_t) = K(x_{t+1}|x_t)dx_{t+1},$$
$$P(Y_t \in dy_t|X_t = x_t) = \rho(y_t|x_t)dy_t. \tag{6a}$$

There is a clear relationship between the general model discussed above and the more applied model given in (2), illustrated by

$$K(x_{t+1}|x_t) = p_v(x_{t+1} - f_t(x_t), t), \tag{7a}$$
$$\rho(y_t|x_t) = p_e(y_t - h_t(x_t), t), \tag{7b}$$

where $p_v(\cdot,t)$ and $p_e(\cdot,t)$ are the probability density functions for v_t and e_t, respectively.

B. Conceptual Solution

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

$$\pi_{t|t-1}(dx_t) = \int_{\mathbb{R}^{n_x}} \pi_{t-1|t-1}(dx_{t-1})K(dx_t|x_{t-1}) \triangleq b_t(\pi_{t-1|t-1}), \tag{8a}$$

$$\pi_{t|t}(dx_t) = \frac{\rho(y_t|x_t)\pi_{t|t-1}(dx_t)}{\int_{\mathbb{R}^{n_x}} \rho(y_t|x_t)\pi_{t|t-1}(dx_t)} \triangleq a_t(\pi_{t|t-1}), \tag{8b}$$

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

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

$$(\nu, \phi) \triangleq \int \phi(x)\nu(dx). \tag{9}$$

Hence, $\mathrm{E}(\phi(x_t)|y_{1:t}) = (\pi_{t|t}, \phi)$. Using this notation, by (8), for any function $\phi : \mathbb{R}^{n_x} \to \mathbb{R}$, we have the following recursive form for the optimal filter $\mathrm{E}(\phi(x_t)|y_{1:t})$,

$$(\pi_{t|t-1}, \phi) = (\pi_{t-1|t-1}, K\phi), \tag{10a}$$

$$(\pi_{t|t}, \phi) = \frac{(\pi_{t|t-1}, \phi\rho)}{(\pi_{t|t-1}, \rho)}. \tag{10b}$$

Here it is worth noticing that we have to require that $(\pi_{t|t-1}, \rho) > 0$, otherwise the optimal filter (10) will not exist. In general it is, as previously mentioned, impossible to obtain an explicit solution for the optimal filter $\mathrm{E}(\phi(x_t)|y_{1:t})$. This implies that we have to resort to numerical methods, such as particle filters, to approximate the optimal filter.

III. Particle Filters

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

A. Introduction

Roughly speaking, particle filtering algorithms are numerical methods used to approximate the conditional filtering distribution $\pi_{t|t}(dx_t)$ using an empirical distribution, consisting of a cloud of particles at each time t. The main reason for using particles to represent the distributions is that this allows us to approximate the integral operators by finite sums. Hence, the difficulty inherent in (8) has successfully been removed. Since there are two integral operators in (8), any practical particle filter has to sample particles at least twice to proceed from time $t-1$ to t. The basic particle filter, as it was introduced by [6] is given in Algorithm 1 and it is briefly described below. For a more thorough introduction, see e.g., [6], [7], [8], [9] where the latter contains a straightforward MATLAB implementation of the particle filter. There are also several books available on the particle filter [10], [11], [12], [3].

Algorithm 1: Particle filter

1) Initialize the particles, $\{x_0^i\}_{i=1}^N \sim \pi_0(dx_0)$.
2) Predict the particles by drawing independent samples according to
$$\tilde{x}_t^i \sim K(dx_t|x_{t-1}^i), \qquad i = 1, \dots, N.$$
3) Compute the importance weights $\{w_t^i\}_{i=1}^N$,
$$w_t^i = \rho(y_t|\tilde{x}_t^i), \quad i = 1, \dots, N,$$
and normalize $\tilde{w}_t^i = w_t^i / \sum_{j=1}^N w_t^j$.
4) Draw N new particles, with replacement (resampling), for each $i = 1, \dots, N$
$$P(x_t^i = \tilde{x}_t^j) = \tilde{w}_t^j, \quad j = 1, \dots, N.$$
5) Set $t := t+1$ and iterate from step 2.

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

$$\pi_{t-1|t-1}^N(dx_{t-1}) \triangleq \frac{1}{N}\sum_{i=1}^N \delta_{x_{t-1}^i}(dx_{t-1}), \tag{11}$$

where $\delta_x(dx_{t-1})$ denotes the delta-Dirac mass located in x. In step 2, the particles from time $t-1$ are predicted to time t using the dynamic equations in the Markov transition kernel K. When step 2 has been performed we have computed the empirical one-step ahead prediction distribution,

$$\tilde{\pi}_{t|t-1}^N(dx_t) \triangleq \frac{1}{N}\sum_{i=1}^N \delta_{\tilde{x}_t^i}(dx_t), \tag{12}$$

which constitutes an estimate of $\pi_{t|t-1}(dx_t)$. In step 3 the information in the present measurement y_t is used. This step can be understood simply by substituting (12) into (8b), resulting in the following approximation of $\pi_{t|t}(dx_t)$

$$\tilde{\pi}_{t|t}^N(dx_t) \triangleq \frac{\rho(y_t|x_t)\tilde{\pi}_{t|t-1}^N(dx_t)}{\int_{R^{n_x}} \rho(y_t|x_t)\tilde{\pi}_{t|t-1}^N(dx_t)} = \frac{\sum_{i=1}^N \rho(y_t|\tilde{x}_t^i)\delta_{\tilde{x}_t^i}(dx_t)}{\sum_{i=1}^N \rho(y_t|\tilde{x}_t^i)}. \tag{13}$$

In practice (13) is usually written using the so called normalized importance weights $\tilde{w}_t^i$, defined as

$$\tilde{\pi}_{t|t}^N(dx_t) = \sum_{i=1}^N \tilde{w}_t^i \delta_{\tilde{x}_t^i}(dx_t), \quad \tilde{w}_t^i \triangleq \frac{\rho(y_t|\tilde{x}_t^i)}{\sum_{i=1}^N \rho(y_t|\tilde{x}_t^i)}. \tag{14}$$

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

$$\pi^N_{t|t}(dx_t) = \frac{1}{N}\sum_{i=1}^{N}\delta_{x_t^i}(dx_t) \tag{15}$$

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

B. Extended Setting

In order to facilitate a theoretical analysis we will now introduce a slightly more general algorithm. The generalization is that the prediction step (step 2 in Algorithm 1) is replaced with the following

$$\tilde{x}_t^i \sim \sum_{j=1}^{N}\alpha_j^i K(dx_t|x_{t-1}^j), \tag{16}$$

where a new set of weights α^i have been introduced. These weights are defined according to

$$\alpha^i = (\alpha_1^i, \alpha_2^i, \dots, \alpha_N^i), \tag{17}$$

where

$$\alpha_j^i \geq 0, \qquad \sum_{j=1}^{N}\alpha_j^i = 1, \qquad \sum_{i=1}^{N}\alpha_j^i = 1. \tag{18}$$

Clearly,

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_j^i K(dx_t|x_{t-1}^j) &= \frac{1}{N}\sum_{j=1}^{N}\left(\sum_{i=1}^{N}\alpha_j^i K(dx_t|x_{t-1}^j)\right)\\ &= \frac{1}{N}\sum_{j=1}^{N}K(dx_t|x_{t-1}^j) = (\pi^N_{t-1|t-1}, K). \end{aligned}\tag{19}$$

Note that if $\alpha_j^i = 1$ for $j = i$, and $\alpha_j^i = 0$ for $j \neq i$, the sampling method introduced in (16) is reduced to the one employed in Algorithm 1. Furthermore, when $\alpha_j^i = 1/N$ for all i and j, (16) turns out to be a convenient form for theoretical treatment. This is exploited by nearly all existing references dealing with theoretical analysis of the particle filter, see for example [4], [13], [5], [3]. An extended particle filtering algorithm is given in Algorithm 2 below.

Algorithm 2: Extended particle filter

1) Initialize the particles, $\{x_0^i\}_{i=1}^N \sim \pi_0(dx_0)$.
2) Predict the particles by drawing independent samples according to
$$\tilde{x}_t^i \sim \sum_{j=1}^{N}\alpha_j^i K(dx_t|x_{t-1}^j), \qquad i = 1, \dots, N.$$
3) Compute the importance weights $\{w_t^i\}_{i=1}^N$,
$$w_t^i = \rho(y_t|\tilde{x}_t^i), \quad i = 1, \dots, N,$$
and normalize $\tilde{w}_t^i = w_t^i / \sum_{j=1}^N w_t^j$.
4) Resample, $x_t^i \sim \tilde{\pi}^N_{t|t}(dx_t)$, $i = 1, \dots, N$. $\pi^N_{t|t}(dx_t) = \frac{1}{N}\sum_{i=1}^N \delta_{x_t^i}(dx_t)$.

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

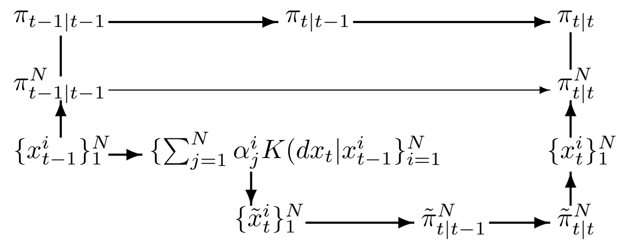

Fig. 1. Illustration of how the particle filter transforms the probability measures. The theoretical transformation (8) is given at the top. The bottom describes what happens during one pass in the particle filter.

transformations of the involved probability measures a bit further, they are

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

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

$$\tilde{\pi}^N_{t|t-1} = c(N)\bar{\circ}\Lambda b_t(\pi^N_{t-1|t-1}), \tag{20}$$

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

$$\pi^N_{t|t} = c^N \circ a_t \circ c(N)\bar{\circ}\Lambda b_t(\pi^N_{t-1|t-1}), \tag{21}$$

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

$$\begin{aligned}\pi^N_{t|t} =& c^N \circ a_t \circ c(N)\bar{\circ}\Lambda b_t \circ \cdots \circ \\ & c^N \circ a_1 \circ c(N)\bar{\circ}\Lambda b_1 \circ c^N(\pi_0).\end{aligned}\tag{22}$$

While, in the existing theoretical versions of the particle filter algorithm in [4], [13], [5], [3], as stated in [4], the transformation between time $t-1$ and t is in a somewhat simpler form,

$$\pi^N_{t|t} = c^N \circ a_t \circ c^N \circ b_t(\pi^N_{t-1|t-1}). \tag{23}$$

IV. A Robust Particle Filter

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

A. Robust Algorithm Modification

From the optimal filter recursion (10b) it is clear that we have to require that

$$(\pi_{t|t-1}, \rho) > 0, \tag{24}$$

in order for the optimal filter to exist. In the approximation to (10b) we have used (12) to approximate $\pi_{t|t-1}(dx_t)$, implying that the following is used in the particle filter algorithm

$$(\pi_{t|t-1}, \rho) \approx (\tilde{\pi}^N_{t|t-1}, \rho) = \int \rho(y_t|x_t) \frac{1}{N} \sum_{i=1}^{N} \delta_{\tilde{x}^i_t}(dx_t)$$
$$= \frac{1}{N} \sum_{i=1}^{N} \rho(y_t|x^i_t). \quad (25)$$

This is implemented in step 3 of Algorithm 1 and 2, i.e., in the importance weight computation. In order to make sure that (24) is fulfilled the algorithm has to be modified. The modification takes the following form, in sampling for $\{\tilde{x}^i_t\}_1^N$ in step 2 of Algorithm 1 and 2, it is required that the following inequality is satisfied

$$(\tilde{\pi}^N_{t|t-1}, \rho) = \sum_{i=1}^{N} \rho(y_t|\tilde{x}^i_t) \geq \gamma_t > 0. \quad (26)$$

Now, clearly, the threshold γ_t must be chosen so that the inequality may be satisfied for sufficiently large N, i.e., so that the true conditional expectation is larger than γ_t. Since this value is typically unknown, it may mean that the problem dependent constant γ_t has to be selected by trial and error and experience. If the inequality (26) holds, the algorithm proceeds as proposed, whereas if it does not hold, a new set of particles $\{\tilde{x}^i_t\}_{i=1}^N$ is generated and (26) is checked again and so on. The modified algorithm is given in Algorithm 3 below.

Algorithm 3: A robust particle filter

1) Initialize the particles, $\{x^i_0\}_{i=1}^N \sim \pi_0(dx_0)$.
2) Predict the particles by drawing independent samples according to

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

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

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

and normalize $\tilde{w}^i_t = w^i_t / \sum_{j=1}^{N} w^j_t$.
5) Resample, $x^i_t \sim \tilde{\pi}^N_{t|t}(dx_t) = \sum_{i=1}^{N} \tilde{w}^i_t \delta_{\tilde{x}^i_t}(dx_t)$, $i = 1, \ldots, N$.
6) Set $t := t + 1$ and iterate from step 2.

The reason for renaming in step 4 is that the distribution of the particles changes by the test in step 3, $\tilde{x}$ which have passed the test have a different distribution from $\bar{x}$. It is interesting to note that this modification, motivated by (10b), makes sense in its own right. Indeed, it has previously, more or less ad hoc, been used as an indicator for divergence in the particle filter and to obtain a more robust algorithm. Furthermore, this modification is related to the well known degeneracy of the particle weights, see e.g., [4], [14] for insightful discussions on this topic.

Clearly, the choice of γ_t may be non-trivial. If it is chosen too large (larger than the true conditional expectation), steps $2-3$ may be an infinite loop. However, it can be proved (see [15] for details) that for a sufficiently large N such an infinite loop will not occur.

B. Numerical Illustration

In order to illustrate the impact of the algorithm modification (26), we study the following nonlinear time-varying system,

$$x_{t+1} = \frac{x_t}{2} + \frac{25x_t}{1+x_t^2} + 8\cos(1.2t) + v_t, \quad (27a)$$
$$y_t = \frac{x_t^2}{20} + e_t, \quad (27b)$$

where $v_t \sim \mathcal{N}(0, 10), e_t \sim \mathcal{N}(0, 1)$, the initial state $x_0 \sim \mathcal{N}(0, 5)$ and $\gamma_t = 10^{-4}$. In the experiment we used 250 time instants and 500 simulations, all using the same measurement sequence. We used the robust particle filter given in Algorithm 3 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 N used in the approximation. The algorithm modification (26) is only active when a small amount of particles is used. That this is indeed the case is evident from Fig. 2, where the average number of interventions due to violations of (26) are given as a function of the number of particles used in the filter.

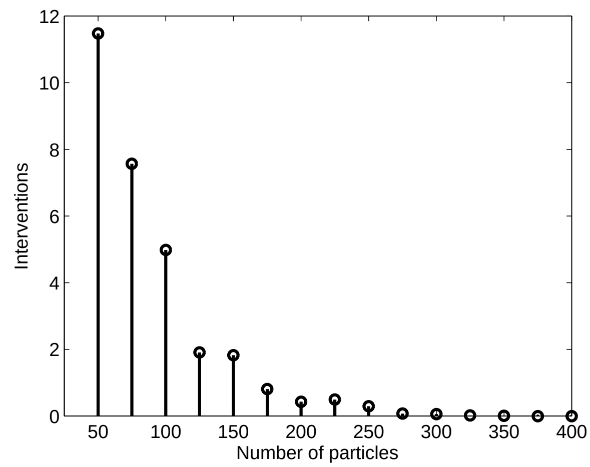

Fig. 2. *Illustration of the impact of the algorithm robustification* (26) *introduced in Algorithm 3. The figure shows the number of times* (26) *was violated and the particles had to be regenerated, as a function of the number of particles used. This is the average result from* 500 *simulations.*

V. Convergence Results

The following conditions are required for the convergence 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 $\phi(\cdot)$ satisfy $\sup_{x_s} |\phi(x_s)|^4 \rho(y_s|x_s) < C(y_{1:s}) < \infty$ for given $y_{1:s}$, $s = 1, \ldots, t$.

Let the class of functions ϕ satisfying H2 be denoted by $L^4_t(\rho)$, where ρ satisfies H1.

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

$$\mathrm{E}\left|(\pi^N_{t|t}, \phi) - (\pi_{t|t}, \phi)\right|^4 \leq C_{t|t} \frac{\|\phi\|^4_{t,4}}{N^2}, \quad (28)$$

where $\|\phi\|_{t,4} \triangleq \max\{1, (\pi_{s|s}, |\phi|^4)^{1/4}, s = 0, 1, \ldots, t\}$ and $\pi^N_{s|s}$ is generated by Algorithm 3, $s = 0, 1, \ldots, t$.

Proof: See Appendix B. ■

See [15] for a more complete treatment of the convergence proof and the premises under which it hold.

By Borel-Cantelli lemma, we have a corollary as follow.

Corollary 1: If H1 and H2 hold, then for any $\phi \in L^4_t(\rho)$,

$$\lim_{N\to\infty} (\pi^N_{t|t}, \phi) = (\pi_{t|t}, \phi), \quad \text{almost surely.} \tag{29}$$

VI. Conclusion

This paper introduced a robust particle filter and the intuition behind this filter is confirmed in the convergence proof for the algorithm. The convergence proof is valid for a rather large class of unbounded functions. We have also provided a rather general setting for discussing particle filters, containing both the applied and the theoretical versions of the filter.

Appendix

A. Auxiliary Lemmas

Before proving Theorem 1, we list some lemmas which we need. 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 Theorem 1.

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

$$\mathrm{E}\left|\sum_{i=1}^n \xi_i\right|^4 \leq \sum_{i=1}^n \mathrm{E}\,\xi_i^4 + \left(\sum_{i=1}^n \mathrm{E}\,\xi_i^2\right)^2. \tag{30}$$

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

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

Based on Lemmas 1 and 3, we have

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

$$\mathrm{E}\left|\frac{1}{n}\sum_{i=1}^n \xi_i\right|^4 \leq \frac{2\max_{1\leq i\leq n}\mathrm{E}\,\xi_i^4}{n^2}. \tag{31}$$

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

B. Proof of Theorem 1

The proof is carried out using mathematical induction. For the sake of brevity this proof does not handle the fact that $\bar{x}$ and $\tilde{x}$ have different distributions. The details for this can be found in the more detailed proof to be found in [15].

Proof:

(1). Let $\{x^i_0\}^N_{i=1}$ 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

$$\begin{aligned}\mathrm{E}\left|(\pi^N_0, \phi) - (\pi_0, \phi)\right|^4 &\leq \frac{2}{N^2}\,\mathrm{E}\,|\phi(x^i_0) - \mathrm{E}[\phi(x^i_0)]|^4 \\ &\leq \frac{32}{N^2}\|\phi\|^4_{0,4} \triangleq C_{0|0}\frac{\|\phi\|^4_{0,4}}{N^2}.\end{aligned} \tag{32}$$

Hence,

$$\mathrm{E}\left|(\pi^N_0, |\phi|^4)\right| \leq 3\,\mathrm{E}\,|\phi(x^i_0)|^4 \triangleq M_{0|0}\|\phi\|^4_{0,4}. \tag{33}$$

(2). Based on (32) and (33), we assume that for $t-1$ and $\forall\phi \in L^4_t(\rho)$

$$\mathrm{E}\left|(\pi^N_{t-1|t-1}, \phi) - (\pi_{t-1|t-1}, \phi)\right|^4 \leq \frac{C_{t-1|t-1}\|\phi\|^4_{t-1,4}}{N^2} \tag{34}$$

and

$$\mathrm{E}\left|(\pi^N_{t-1|t-1}, |\phi|^4)\right| \leq M_{t-1|t-1}\|\phi\|^4_{t-1,4} \tag{35}$$

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

Notice that

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

where

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

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

and $\pi^{N,\alpha_i}_{t-1|t-1} = \sum_{j=1}^N \alpha^i_j \delta_{x^j_{t-1}}(dx_{t-1})$. Let us now investigate Π_1 and Π_2.

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

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

Thus, by Lemmas 1, 2, 3, (19) and (35),

$$\mathrm{E}\,|\Pi_1|^4 \leq 2^5\frac{\|K\|^4 M_{t-1|t-1}\|\phi\|^4_{t-1,4}}{N^2}. \tag{36}$$

Furthermore, by (19) and (34),

$$\mathrm{E}\,|\Pi_2|^4 \leq \frac{C_{t-1|t-1}\|K\|^4\|\phi\|^4_{t-1,4}}{N^2}. \tag{37}$$

Then, using Minkowski's inequality, (19), (36), and (37), we have

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

That is

$$\mathrm{E}\left|(\tilde{\pi}^N_{t|t-1}, \phi) - (\pi_{t|t-1}, \phi)\right|^4 \leq \tilde{C}_{t|t-1}\frac{\|\phi\|^4_{t-1,4}}{N^2}. \tag{38}$$

By Lemma 2, (35) and the use of a separation, similar to the one employed above, we have

$$\begin{aligned}&\mathrm{E}\left|(\tilde{\pi}^N_{t|t-1}, |\phi|^4) - (\pi_{t|t-1}, |\phi|^4)\right| \\ &\leq \|K\|^4(3M_{t-1|t-1} + 1)\|\phi\|^4_{t-1,4} \triangleq \tilde{M}_{t|t-1}\|\phi\|^4_{t-1,4}.\end{aligned} \tag{39}$$

(4). In this step we analyse $\mathrm{E}\left|(\tilde{\pi}^N_{t|t},\phi)-(\pi_{t|t},\phi)\right|^4$ and $\mathrm{E}(\tilde{\pi}^N_{t|t},|\phi|^4)$ based on (38) and (39). Clearly,

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

where

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

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

By condition H1 and the modification (26) introduced in Algorithm 3 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 (38),

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

which implies

$$\mathrm{E}\left|(\tilde{\pi}^N_{t|t},\phi)-(\pi_{t|t},\phi)\right|^4 \le \tilde{C}_{t|t}\frac{\|\phi\|^4_{t-1,4}}{N^2}. \qquad (40)$$

Using a separation similar to the one previously used, by (39), and observing that $\|\phi\|_{s,4}\ge 1$ is increasing with respect to s, we have

$$\begin{aligned}&\mathrm{E}\left|(\tilde{\pi}^N_{t|t},|\phi|^4)\right| \\ &\le 3\max\left\{\frac{\|\rho\phi^4\|\cdot 2\|\rho\|}{\gamma_t(\pi_{t|t-1},\rho)},\frac{\tilde{M}_{t|t-1}\|\rho\|}{(\pi_{t|t-1},\rho)},1\right\}\cdot\|\phi\|^4_{t,4} \\ &\triangleq \tilde{M}_{t|t}\|\phi\|^4_{t,4}. \qquad (41)\end{aligned}$$

(5). 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 (40) and (41).

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 (41),

$$\mathrm{E}\,|\bar{\Pi}_1|^4 \le 2^5\tilde{M}_{t|t}\frac{\|\phi\|^4_{t,4}}{N^2}. \qquad (42)$$

Then by Minkowski's inequality, (40) and (42)

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

That is

$$\mathrm{E}\left|(\pi^N_{t|t},\phi)-(\pi_{t|t},\phi)\right|^4 \le C_{t|t}\frac{\|\phi\|^4_{t,4}}{N^2}. \qquad (43)$$

Using a similar separation mentioned above, by (41),

$$\mathrm{E}\left|(\pi^N_{t|t},|\phi|^4)\right| \le (3\tilde{M}_{t|t}+2)\|\phi\|^4_{t,4} \triangleq M_{t|t}\|\phi\|^4_{t,4}. \qquad (44)$$

Therefore, the proof of Theorem 1 is completed, since (34) and (35) are successfully replaced by (43) and (44), respectively. ■

References

[1] A. H. Jazwinski, *Stochastic processes and filtering theory*, ser. Mathematics in science and engineering. New York, USA: Academic Press, 1970.

[2] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME, Journal of Basic Engineering*, vol. 82, pp. 35–45, 1960.

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

[4] 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.

[5] 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.

[6] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," in *IEE Proceedings on Radar and Signal Processing*, vol. 140, 1993, pp. 107–113.

[7] 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.

[8] 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.

[9] T. B. Schön, "*Estimation of Nonlinear Dynamic Systems – Theory and Applications*," PhD thesis No 998, Department of Electrical Engineering, Linköping University, Sweden, Feb. 2006.

[10] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. New York, USA: Springer Verlag, 2001.

[11] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, ser. Springer Series in Statistics. New York, USA: Springer, 2001.

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

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

[14] F. Legland and N. Oudjane, "Stability and uniform approximation of nonlinear filters using the hilbert metric, and application to particle filters," INRIA, Paris, France, Tech. Rep. RR–4215, 2001.

[15] X.-L. Hu, T. B. Schön, and L. Ljung, "A basic convergence result for particle filtering," *Submitted to IEEE Transactions on Signal Processing*, 2007.