Commun. Comput. Phys.
doi: 10.4208/cicp.180313.130214a

Vol. **16**, No. 2, pp. 382-402
August 2014

An Implicit Algorithm of Solving Nonlinear Filtering Problems

Feng Bao[1], Yanzhao Cao[1,2,*] and Xiaoying Han[1]

[1] *Department of Mathematics and Statistics, Auburn University, Auburn, AL 36849, USA.*
[2] *School of Mathematics and Computational Sciences, Sun Yat-sen University, China.*

Received 18 March 2013; Accepted (in revised version) 13 February 2014

Communicated by Qiang Du

Available online 9 May 2014

Abstract. Nonlinear filter problems arise in many applications such as communications and signal processing. Commonly used numerical simulation methods include Kalman filter method, particle filter method, etc. In this paper a novel numerical algorithm is constructed based on samples of the current state obtained by solving the state equation implicitly. Numerical experiments demonstrate that our algorithm is more accurate than the Kalman filter and more stable than the particle filter.

AMS subject classifications: 60G35, 62M20, 93E11

Key words: Kalman filter, particle filter, implicit filter, Monte Carlo method, stochastic differential equations.

1 Introduction

The main purpose of numerical simulations of a filtering process is to obtain, recursively in time, a good estimate for the probability density function (pdf) of the state of a dynamical system based on noisy observations. The first major breakthrough in this classical problem of signal analysis is the landmark work of Kalman and Bucy (Kalman filter) [16] on linear filtering (see also [9, 19, 20, 22]), under the assumption of linearity of the system and Gaussianity of the noise, and the conditional distribution of the state, given the observations, is Gaussian. This conditional distribution gives the best estimate of the

*Corresponding author. *Email addresses:* fzb0005@auburn.edu (F. Bao), yzc0009@auburn.edu (Y. Cao), xzh0003@auburn.edu (X. Han)

statistical description of the state of the system based on all the available observation information up to the current time. Largely because of the success of Kalman filters, linear and nonlinear filters have been applied in the various engineering and scientific areas, including communications such as positioning in wireless networks, signal processing such as tracking and navigation, economics and business, and many others.

In most of the practical application problems, however, linearity assumption is not valid because of the nonlinearity in the model specification process as well as the observation process. Two of the widely used methods for nonlinear filtering problems are extended Kalman filter (EKF) [2,8,13–15] and particle filter method (PFM) [3–5,11,21,25]. In EFK the estimation problem is linearized around the predicted state so that the standard Kalman filter can be applied. The central idea of the particle filter method is to represent the desired pdf of the system state with a set of random samples. As the number of samples becomes very large, PFM provides a representation of the pdf. Since the seminal work of Gordon, Salmond and Smith [11], there have been significant development in both practical applications and theoretical analysis on PFM. Other efforts of solving nonlinear filtering problems include the Gaussian sum filter [1,17,24], moment methods based on approximations of the first two moments of the density [18] and Zakai filter which represents the pdf as the solution of a parabolic type stochastic partial differential equation (Zakai equation) [6,10,12,23,26,27].

While the aforementioned methods have been remarkably successful in attacking the nonlinear filtering problem, each of them has its drawbacks and limitations. For instance, when the state equation describing the signal process and the observation equation are highly non-linear, the extended Kalman filter can give particularly poor performance. PFM has a number of advantages over EFK, including its ability to represent arbitrary densities, adaptively focus on the most probable regions of state-space. However, it also has a number of disadvantages, including high computational complexity, degeneracy for long period simulation and its difficulty of determining optimal number of particles.

The goal of this paper is to construct a new algorithm for numerical simulations of nonlinear filtering problems. The general framework of our algorithm is adopted from the Bayesian filtering theory which constructs the pdf of the state based on all the available information. We still use the general framework of Kalman filters and particle filters which solve the problem by two stages: prediction stage and update stage. For each time recursive step, the prediction stage gives the estimation for the prior pdf of the future state based on the currently available observation information while update stage gives the posterior pdf from the updated observation information and the result obtained in the prediction stage. However, instead of attempting to search for a representation of the pdf as in PFM, we approximate the pdf as a function over a grid in state space. Specifically, at the prediction stage, we attempt to seek the predicted pdf of the future state variable through a Monte Carlo method by evaluating the conditional expectation of the future state with respect to the current stage. Since the sample points for the current state is computed by solving the state equation implicitly, we name our method as an "implicit filter method". The following two items summarize the novelty of our approach:

(i) We propose an accurate implicit scheme for prediction purpose. The implicit scheme has stabilizing effect on the proposed numerical algorithm. This is verified in our numerical experiments.

(ii) Based on the Bayesian theory, we apply a novel Monte-Carlo like method to approximate the conditional expectation in the update stage to compute the prior pdf.

The paper is organized as follows. In the next section, we describe our methodology and the implicit filter algorithm. In Section 3, we provide the convergence analysis of our algorithm. Finally in Section 4 we demonstrate the accuracy and long time stability of our algorithm through numerical experiments.

2 Methodology and the implicit filter algorithm

Consider the following state and observation equations in the dynamic state-space form:

$$\frac{dX_t}{dt} = f(t, X_t; W_t), \tag{2.1}$$

$$Y_t = g(t, X_t; V_t), \tag{2.2}$$

where $X_t \in \mathbb{R}^{n_x}$ denotes the state vector, $Y_t \in \mathbb{R}^{n_y}$ denotes the measurement vector, $W_t \in \mathbb{R}^{n_w}$ is a random vector representing the uncertainties in the model, and $V_t \in \mathbb{R}^{n_v}$ denotes the random measurement error. In many applications, the noise from measurement can be assumed to be additive, and the problem can be formulated in a discrete manner as

$$X_{t+1} = f_t(X_t, W_t), \tag{2.3}$$

$$Y_t = g_t(X_t) + V_t, \tag{2.4}$$

where $\{W_t\}_{t\in\mathbb{N}} \in \mathbb{R}^{n_w}$ and $\{V_t\}_{t\in\mathbb{N}\setminus\{0\}} \in \mathbb{R}^{n_v}$ are mutually independent white noises and the subscript t indexes the discrete time level at which the functions are evaluated. In data assimilation, the observation Y_t arrives sequentially in time and the goal is to estimate the state vector X_t given the information of $\{Y_s,\ 0<s\leq t\}$. In what follows, we provide a brief review on the formulation of Bayesian optimal filter.

2.1 Bayesian optimal filter

First we introduce some notations that will be used throughout the rest of the paper. Denote $Z_{m:n}$ as $(Z_m, Z_{m+1}, \cdots, Z_n)$ and denote $X_t \sim p(x_t)$ if the pdf of a random variable X_t is $p(x_t)$. Write

$$X_t \,|\, (X_{t-1} = x_{t-1}) \sim p(x_t \,|\, x_{t-1}), \tag{2.5}$$

where $X_t \,|\, X_{t-1}$ denotes the conditional expectation. When the context is clear, notations similar to (2.5) will be introduced without formal explanations.

The dynamical model is Markovian such that any future X_t is independent of the past given the present X_{t-1}:

$$p(x_t|x_{1:t-1},y_{1:t-1}) = p(x_t|x_{t-1}),$$

and the measurements are conditionally independent given x_t:

$$p(y_t|x_{1:t},y_{1:t-1}) = p(y_t|x_t).$$

Denote by $\mathcal{I}_t \doteq \{y_1,y_2,\cdots,y_t\}$ the information observed before time t. Given a prior distribution $p(x_0)$, Bayesian optimal filter is to construct the distribution $p(x_t|\mathcal{I}_t)$ recursively in two stages: prediction and update.

Assume that the required pdf $p(x_{t-1}|\mathcal{I}_{t-1})$ of previous step $t-1$ is available. The Chapman-Kolmogorov equation gives the prediction step of

$$p(x_t|\mathcal{I}_{t-1}) = \int_{\mathbb{R}^{n_x}} p(x_t|x_{t-1})p(x_{t-1}|\mathcal{I}_{t-1})dx_{t-1}. \tag{2.6}$$

At time t, as measurement y_t becomes available, the prior distribution from (2.6) can then be updated via Bayesian's formula

$$p(x_t|\mathcal{I}_t) = \frac{p(y_t|x_t)p(x_t|\mathcal{I}_{t-1})}{p(y_t|\mathcal{I}_{t-1})} = \frac{p(y_t|x_t)p(x_t|\mathcal{I}_{t-1})}{\int_{\mathbb{R}^{n_x}} p(y_t|x_t)p(x_t|\mathcal{I}_{t-1})dx_t}. \tag{2.7}$$

The exact computation of (2.6) and (2.7) is generally not possible. Exception exists where all $p(x_t|\mathcal{I}_t)$ are Gaussian and the model is linear, in which case the moments can be obtained using Kalman filter. In practically all other cases, approximate solutions are sought by numerical methods. Traditional particle filtering methods recursively generate samples (particles) following $p(x_t|\mathcal{I}_t)$ and use these samples to approximate moments of $p(x_t|\mathcal{I}_t)$. We propose here an inverse method which generates samples from the white noise and use them to approximate $p(x_t|\mathcal{I}_t)$, by utilizing inverse solutions to discretized stochastic differential equations.

2.2 An inverse algorithm

Our proposed method is based on the fact that

$$p(x_t|x_{t-1}) = \int_{\mathbb{R}^{n_w}} p(x_t|x_{t-1},w_{t-1})\cdot p(w_{t-1})dw_{t-1} = \mathbb{E}[p(x_t|x_{t-1},W_{t-1})],$$

and therefore the term $p(x_t|\mathcal{I}_{t-1})$ in (2.6) can be written as

$$\begin{aligned} p(x_t|\mathcal{I}_{t-1}) &= \int_{\mathbb{R}^{n_x}} p(x_t|x_{t-1})p(x_{t-1}|\mathcal{I}_{t-1})dx_{t-1} \\ &= \int_{\mathbb{R}^{n_x}} \mathbb{E}[p(x_t|x_{t-1},W_{t-1})]p(x_{t-1}|\mathcal{I}_{t-1})dx_{t-1}. \end{aligned} \tag{2.8}$$

We assume that a compact domain, $\mathcal{B}\subset\mathbb{R}^{n_x}$, is the region of interest. Assuming that the pdf $p(x_{t-1}|\mathcal{I}_{t-1})$ is given, to achieve the prediction step (2.8) from time $t-1$ to time t:

1. Generate M particles/paths $\{w_{t-1}^{(j)}\}_{j=1,\cdots,M}$ according to the pdf of W_{t-1} and approximate $p(\cdot|W_{t-1})$ by its empirical pdf, denoted by

$$\pi^M(\cdot|W_{t-1})\doteq\frac{1}{M}\sum_{j=1}^{M}\delta_{\cdot|w_{t-1}^{(j)}},$$

where δ_x denotes the delta-Dirac mass located in x. In fact, according to Bayesian formula,

$$p(x_t|x_{t-1},w_{t-1})=\frac{p(x_{t-1},w_{t-1}|x_t)}{p(x_{t-1},w_{t-1})}\cdot p(x_t),$$

and it follows immediately that $p(x_t|x_{t-1},W_{t-1})$ is a random variable and the randomness comes from the white noise W_{t-1}, for each given x_t and $x_{t-1}\in\mathbb{R}^{n_x}$. Therefore the term $\mathbb{E}[p(x_t|x_{t-1},W_{t-1})]$ in (2.8) can be approximated by

$$\mathbb{E}[p(x_t|x_{t-1},W_{t-1})]\approx\mathbb{E}[\pi^M(x_t|x_{t-1},W_{t-1})],$$

in which $\pi^M(x_t|x_{t-1},W_{t-1})$ has the probability distribution

$$Pr(\pi^M(x_t|x_{t-1},w_{t-1}^{(j)}))=\frac{1}{M},\quad j=1,\cdots,M.$$

2. Partition region $\mathcal{B}$ by using N nodes: $\{u^{(i)}\}_{i=1,\cdots,N}$ and approximate $\pi^M(x_t|\cdot)$ by $\pi^M(u^{(i)}|\cdot)$. Step 1 together with (2.8) gives

$$p(u^{(i)}|\mathcal{I}_{t-1})\approx\int_{\mathcal{B}}\mathbb{E}[\pi^M(u^{(i)}|x_{t-1},W_{t-1})]p(x_{t-1}|\mathcal{I}_{t-1})dx_{t-1}. \tag{2.9}$$

3. Assume that f_t is invertible. Let $x_{t-1}^{(i,j)}$ be the solution to the equation $f_t(x_{t-1}^{(i,j)},w_{t-1}^{(j)})=u^{(i)}$ for $j=1,\cdots,M$ and $i=1,\cdots,N$. Then

$$\begin{aligned}\mathbb{E}[\pi^M(u^{(i)}|x_{t-1},W_{t-1})]&=\frac{1}{M}\sum_{j=1}^{M}\pi^M(u^{(i)}|x_{t-1},w_{t-1}^{(j)})\\&=\frac{1}{M}\sum_{j=1}^{M}\delta(x_{t-1}-x_{t-1}^{(i,j)}),\quad i=1,\cdots,N.\end{aligned} \tag{2.10}$$

Hence the integral on the right hand side of (2.9) can be simplified and gives that for each $i=1,\cdots,N$,

$$\begin{aligned}p(u^{(i)}|\mathcal{I}_{t-1})&\approx\int_{\mathcal{B}}\mathbb{E}[\pi^M(u^{(i)}|x_{t-1},W_{t-1})]p(x_{t-1}|\mathcal{I}_{t-1})dx_{t-1}\\&=\frac{1}{M}\sum_{j=1}^{M}p\left(x_{t-1}^{(i,j)}|\mathcal{I}_{t-1}\right).\end{aligned} \tag{2.11}$$

4. The last step of prediction is to use interpolation to construct a piecewise approximation $\rho(x_t|\mathcal{I}_{t-1})$ of $p(x_t|\mathcal{I}_{t-1})$, from $p(u^{(i)}|\mathcal{I}_{t-1})$ obtained in step 3. Denote by $T[\{\cdot\}]$ the piecewise linear function

that connects the points $(u^{(i)},\{\cdot\})$, then we obtain the approximation of $p(x_t|\mathcal{I}_{t-1})$ via

$$\begin{aligned} p(x_t|\mathcal{I}_{t-1}) &\approx T\left[\{p(u^{(i)}|\mathcal{I}_{t-1})\}_{i=1}^N\right] \\ &\approx T\left[\left\{\frac{1}{M}\sum_{j=1}^M p\left(x_{t-1}^{(i,j)}|\mathcal{I}_{t-1}\right)\right\}_{i=1}^N\right] \doteq \rho(x_t|\mathcal{I}_{t-1}), \end{aligned} \tag{2.12}$$

where $p(x_{t-1}^{(i,j)}|\mathcal{I}_{t-1})$ is the value of the pdf $p(x_{t-1}|\mathcal{I}_{t-1})$ at the point $x_{t-1}=x_{t-1}^{(i,j)}$.

Finally in the update step, we update the prior pdf $\rho(x_t|\mathcal{I}_{t-1})$ at $x_t\in\mathcal{B}$ by using the Bayes formula. Thus the approximation of $p(x_t|\mathcal{I}_t)$, denoted by $\rho(x_t|\mathcal{I}_t)$ is given as

$$p(x_t|\mathcal{I}_t)\approx\rho(x_t|\mathcal{I}_t)=\frac{p(y_t|x_t)\rho(x_t|\mathcal{I}_{t-1})}{\int_{\mathcal{B}}p(y_t|x_t)\rho(x_t|\mathcal{I}_{t-1})dx_t}. \tag{2.13}$$

In this way, we obtain the approximation of $p(x_t|\mathcal{I}_t)$ on nodes $\{u^{(i)}\}_{i=1,\cdots,N}$.

Remark 2.1. It is also worthy nothing that our algorithm estimates the moments based on the information of the deterministic space points which cover all possible states instead of evaluating the state density through randomly selected particles in the particle filter method, which partially explains the fact that in the numerical experiments, our algorithm is more stable than the standard particle filter method.

3 Weak convergence

In this section we will study the convergence of the pdf obtained by our algorithm converges to the Bayesian optimal filter on $\mathcal{B}$. In general, given a measure μ and a function φ, we define

$$\langle\mu(\cdot),\varphi\rangle=\int\varphi(x)\mu(\cdot)\,dx.$$

Definition 3.1. Let $\{\mu_n\}_{n=1}^\infty$ be a sequence of probability densities on $\mathcal{P}(\mathcal{B})$, where $\mathcal{P}(\mathcal{B})$ is the space of all probability measures over $\mathcal{B}$. We say that

- μ_n converges to $\mu\in\mathcal{P}(\mathcal{B})$ uniformly and write $\lim_{n\to\infty}\mu_n=\mu$ if for any $\varepsilon>0$, there exists N_0, such that $|\mu_n(z)-\mu(z)|<\varepsilon$ for all $z\in\mathcal{B}$ and $n>N_0$.
- μ_n converges to $\mu\in\mathcal{P}(\mathcal{B})$ weakly and write $\lim_{n\to\infty}\mu_n\overset{\varphi}{=}\mu$ if

$$\lim_{n\to\infty}\langle\mu_n,\varphi\rangle=\langle\mu,\varphi\rangle,\quad\forall\varphi\in C_b(\mathcal{B}),$$

  where $C_b(\mathcal{B})$ is the set of all continuous bounded functions on $\mathcal{B}$.

We will prove the weak convergence of $\rho(x_t|\mathcal{I}_t)$ to $p(x_t|\mathcal{I}_t)$, i.e., the convergence of $\langle\rho(x_t|\mathcal{I}_t),\varphi\rangle$ to $\langle p(x_t|\mathcal{I}_t),\varphi\rangle$.

To guarantee that the Bayes' formula in (2.7) is well defined and can be fulfilled in our algorithm, we make the following standing assumptions:

(A1) For given $\mathcal{I}_t$, the denominator in (2.7) (normalization constants) satisfies

$$\int_{\mathcal{B}} p(y_t|x_t)p(x_t|\mathcal{I}_{t-1})\,dx_t > \xi > 0.$$

(A2) The conditional kernel densities $p(x_t|x_{t-1})$ and $p(y_t|x_t)$ are uniformly continuous, bounded and strictly positive, i.e., given $\mathcal{I}_t$,

$$0 < p(x_t|x_{t-1}) < 1, \quad 0 < p(y_t|x_t) < 1.$$

For simplicity we denote $K_{t|t-1} := \mathbb{E}[p(x_t|x_{t-1},W_{t-1})]$, denote the true pdf's by

$$p_{t-1|t-1} := p(x_{t-1}|\mathcal{I}_{t-1}), \quad p_{t|t-1} := p(x_t|\mathcal{I}_{t-1}), \quad p_{t|t} := p(x_t|\mathcal{I}_t);$$

and similarly denote the simulated pdf's by

$$\rho_{t-1|t-1} := \rho(x_{t-1}|\mathcal{I}_{t-1}), \quad \rho_{t|t-1} := \rho(x_t|\mathcal{I}_{t-1}), \quad \rho_{t|t} := \rho(x_t|\mathcal{I}_t).$$

We first define $a_t : \mathcal{P}(\mathcal{B}) \to \mathcal{P}(\mathcal{B})$ to be the mapping

$$a_t(\mu)(x_t) = \int_{\mathcal{B}} K_{t|t-1}\mu(x_{t-1})\,dx_{t-1}, \quad \forall \mu \in \mathcal{P}(\mathcal{B}). \tag{3.1}$$

Then we have

$$\langle a_t(\mu),\varphi\rangle = \langle \mu, K_{t|t-1}\varphi\rangle, \quad \forall \varphi \in C_b(\mathcal{B}), \tag{3.2}$$

and it holds that

$$a_t(p_{t-1|t-1}) = p_{t|t-1}.$$

It's natural to assume that a_t is continuous, since in the context of filtering two realizations of the signal that start from "close" positions will remain "close" at subsequent times. In fact, when the transition kernel $p(x_t|x_{t-1})$ is Feller, i.e., $p(x_t|x_{t-1})\varphi$ is a continuous bounded function for any continuous bounded function φ, we have according to [7] that if $\lim_{n\to\infty}\mu_n \overset{\varphi}{=} \mu$ then

$$\lim_{n\to\infty}\langle a_t(\mu_n),\varphi\rangle = \lim_{n\to\infty}\langle \mu_n, K_{t|t-1}\varphi\rangle = \langle \mu, K_{t|t-1}\varphi\rangle = \langle a_t(\mu),\varphi\rangle, \quad \forall\ \varphi \in C_b(\mathcal{B}). \tag{3.3}$$

Define $b_t : \mathcal{P}(\mathcal{B}) \to \mathcal{P}(\mathcal{B})$ to be the mapping

$$b_t(\mu) = \frac{p(y_t|x_t)\mu(x_t)}{\int_{\mathcal{B}} p(y_t|x_t)\mu(x_t)dx_t}, \quad \forall \mu \in \mathcal{P}(\mathcal{B}). \tag{3.4}$$

Then we have

$$\langle b_t(p_{t|t-1}),\varphi\rangle = \langle p_{t|t-1},p(y_t|x_t)\rangle^{-1}\cdot\langle p_{t|t-1},\varphi p(y_t|x_t)\rangle,\quad \forall\varphi\in C_b(\mathcal{B}), \tag{3.5}$$

and it holds that

$$b_t(p_{t|t-1}) = p_{t|t}.$$

It is also natural to assume that b_t is continuous, which means that a slight variation in two distributions will not result in a large variation in the distributions when observations are taken into account. In fact, assuming that $p(y_t|\cdot)$ is a continuous bounded strictly positive function, we have according to [7] that $\lim_{n\to\infty}\mu_n \overset{\varphi}{=} \mu$ then

$$\begin{aligned}\lim_{n\to\infty}\langle b_t(\mu_n),\varphi\rangle &= \lim_{n\to\infty}\langle \mu_n,p(y_t|x_t)\rangle^{-1}\cdot\langle\mu_n,\varphi p(y_t|x_t)\rangle\\ &= \langle\mu,p(y_t|x_t)\rangle^{-1}\cdot\langle\mu,\varphi p(y_t|x_t)\rangle = \langle b_t(\mu),\varphi\rangle,\quad \forall\ \varphi\in C_b(\mathcal{B}).\end{aligned} \tag{3.6}$$

We next define two approximation operators, the sampling operator, and the interpolation operator that appear in the prediction step.

1. Denote by ψ a function of W_{t-1} from $\mathcal{P}(\mathcal{B})$ to $\mathcal{P}(\mathcal{B})$. At each step t, we draw M samples, $w_{t-1}^{(1)},\cdots,w_{t-1}^{(M)}$, which are i.i.d. random variables with common distribution W_{t-1}. The Monte Carlo estimate of $\mathbb{E}[\psi(W_{t-1})]$ can be obtained to be

$$\tilde{\psi}(w_{t-1}) = \frac{1}{M}\sum_{j=1}^{M}\psi(w_{t-1}^{(j)}).$$

   Define the sampling operator $s^M:\mathcal{P}(\mathcal{B})\to\mathcal{P}(\mathcal{B})$ to be

$$s^M(\langle\mathbb{E}[\psi],\mu\rangle) = \langle\mathbb{E}[\tilde{\psi}],\mu\rangle,\quad \forall\mu\in\mathcal{P}(\mathcal{B}). \tag{3.7}$$

   Then

$$s^M\circ a_t(\mu) = \int_{\mathcal{B}}\frac{1}{M}\sum_{j=1}^{M}\pi^M(x_t|x_{t-1},w_{t-1}^{(j)})\cdot\mu(x_{t-1})\,dx_{t-1},\quad \forall\mu\in\mathcal{P}(\mathcal{B}).$$

2. Given the rectangular spatial partition nodes $\{u^{(i)}\}_{i=1}^N$. Define $T^N:\mathcal{P}(\mathcal{B})\to\mathcal{P}(\mathcal{B})$ to be the interpolation operator

$$T^N(\mu)(x_t) = \sum_{i=1}^{N}\psi_i^N(x_t)\mu(u^{(i)}),$$

   for each $x_t\in\mathcal{B}$, where $\{\psi_i\}_{i=1}^N$ are basis functions for interpolation operator T^N. Therefore

$$T^N\circ s^M\circ a_t(p_{t-1|t-1}) = \rho_{t|t-1},$$

where $\rho_{t|t-1}$ is the piecewise linear approximation of $p_{t|t-1}$ satisfying

$$\rho_{t|t-1}(u^{(i)}) = \frac{1}{M}\sum_{j=1}^{M} p\left(x_{t-1}^{(i,j)}|\mathcal{I}_{t-1}\right), \quad \text{for each } i=1,\cdots,N.$$

Denote by $\kappa_t \doteq b_t \circ a_t$, $\kappa_t^{M,N} \doteq b_t \circ T^N \circ s^M \circ a_t$, $\kappa_{1:t} = \kappa_t \circ \kappa_{t-1} \circ \cdots \circ \kappa_1$ and $\kappa_{1:t}^{M,N} = \kappa_t^{M,N} \circ \kappa_{t-1}^{M,N} \circ \cdots \circ \kappa_1^{M,N}$, we have

$$\kappa_t(p_{t-1|t-1}) = p_{t|t}, \quad \kappa_t^{M,N}(p_{t-1|t-1}) = \rho_{t|t}, \quad \text{and} \quad \kappa_{1:t}(p_{0|0}) = p_{t|t}, \quad \kappa_{1:t}^{M,N}(p_{0|0}) = \rho_{t|t}.$$

Our goal is to show that $\kappa_{1:t}^{M,N} \xrightarrow{\varphi} \kappa_{1:t}$. This can be done by showing $\kappa_t^{M,N} \xrightarrow{\varphi} \kappa_t$ for each step t and induction.

Recalling that $\mathcal{P}(\mathcal{B})$ is the set of all probability measures on $\mathcal{B}$, we denote by $\mathcal{P}_U(\mathcal{B})$ be the set of all uniformly continuous probability measures on $\mathcal{B}$ and $\mathcal{P}_C(\mathcal{B})$ the set of all a.e. continuous probability measures on $\mathcal{B}$, for latter use.

Lemma 3.1. *$s^M \circ a_t$ converges to a_t weakly, i.e., for any $\mu_M, \mu \in \mathcal{P}(\mathcal{B})$ with $\lim_{M\to\infty} \mu_M \overset{\varphi}{=} \mu$, it holds that*

$$\lim_{M\to\infty} s^M \circ a_t(\mu_M) \overset{\varphi}{=} a_t(\mu). \tag{3.8}$$

Proof. For any $t \in \mathbb{N}$, by the Strong Law of Large Numbers,

$$\lim_{M\to\infty} \frac{1}{M}\sum_{j=1}^{M} \pi^M(x_t|x_{t-1}, w_{t-1}^{(j)}) = \mathbb{E}[p(x_t|x_{t-1}, W_{t-1})], \quad a.s. \tag{3.9}$$

Therefore, for any $\varphi \in C_b(\mathcal{B})$ we have

$$\begin{aligned}
&\left\langle s^M \circ a_t(\mu_M), \varphi \right\rangle - \langle a_t(\mu), \varphi \rangle \\
=&\left\langle s^M \circ a_t(\mu_M), \varphi \right\rangle - \langle a_t(\mu_M), \varphi \rangle + \langle a_t(\mu_M), \varphi \rangle - \langle a_t(\mu), \varphi \rangle \\
\leq& \int_{\mathcal{B}} \left(\frac{1}{M}\sum_{j=1}^{M} \pi^M(x_t|x_{t-1}, w_{t-1}^{(j)}) \mu_M(x_{t-1}) dx_{t-1} - \mathbb{E}[p(x_t|x_{t-1}, W_{t-1})] \right) \mu_M(x_{t-1}) \varphi dx_{t-1} \\
&+ \langle a_t(\mu_M), \varphi \rangle - \langle a_t(\mu), \varphi \rangle.
\end{aligned} \tag{3.10}$$

It then follows directly from Eqs. (3.9) and (3.3) that

$$\lim_{M\to\infty} s^M \circ a_t(\mu_M) \overset{\varphi}{=} a_t(\mu).$$

The proof is complete. □

Lemma 3.2. *Assume that $\{\mu_{M,N}\}_{M,N=1}^{\infty} \in \mathcal{P}_C(\mathcal{B})$ and $\mu_M \in \mathcal{P}_U(\mathcal{B})$ with $\lim_{N\to\infty} \mu_{M,N} = \mu_M$ for each $M \in \mathbb{N}$. Then, $\lim_{N\to\infty} s^M \circ a_t(\mu_{M,N}) = s^M \circ a_t(\mu_M)$ for each $M \in \mathbb{N}$. Moreover, if there exists $\lambda > 0$ such that $\left\| \frac{\partial}{\partial x} f_t^{-1} \right\| < \lambda$, then $s^M \circ a_t(\mu_{M,N}) \in \mathcal{P}_C(\mathcal{B})$ and $s^M \circ a_t(\mu_M) \in \mathcal{P}_U(\mathcal{B})$.*

Proof. For any $x\in\mathcal{B}$, by the definition of s^M and a_t we have

$$s^M\circ a_t(\mu_{M,N})(x)=\frac{1}{M}\sum_{j=1}^{M}\mu_{M,N}\left(x_{t-1}^{(x,j)}\right)$$

and

$$s^M\circ a_t(\mu_M)(x)=\frac{1}{M}\sum_{j=1}^{M}\mu_M\left(x_{t-1}^{(x,j)}\right).$$

Since $\lim_{N\to\infty}\mu_{M,N}=\mu_M$ for each $M\in\mathbb{N}$, given any $\varepsilon>0$, there exists N_0, such that for all $z\in\mathcal{B}$, $|\mu_{M,N}(z)-\mu_M(z)|<\varepsilon$ for each $M\in\mathbb{N}$. Therefore, for all $x\in\mathcal{B}$,

$$\begin{aligned}\left|s^M\circ a_t(\mu_{M,N})(x)-s^M\circ a_t(\mu_M)(x)\right|&=\left|\frac{1}{M}\sum_{j=1}^{M}\left(\mu_{M,N}\left(x_{t-1}^{(x,j)}\right)-\mu_M\left(x_{t-1}^{(x,j)}\right)\right)\right|\\&\le\frac{1}{M}\sum_{j=1}^{M}\left|\left(\mu_{M,N}\left(x_{t-1}^{(x,j)}\right)-\mu_M\left(x_{t-1}^{(x,j)}\right)\right)\right|\\&<\frac{1}{M}\sum_{j=1}^{M}\varepsilon=\varepsilon.\end{aligned}$$

This proves that $\lim_{N\to\infty}s^M\circ a_t(\mu_{M,N})=s^M\circ a_t(\mu_M)$ for all $M\in\mathbb{N}$.

We next prove that $s^M\circ a_t(\mu_{M,N})\in\mathcal{P}_C(\mathcal{B})$. In fact, for any $\varepsilon>0$ and $z_0\in\mathcal{B}$, since $\mu_{M,N}\in\mathcal{P}_C(\mathcal{B})$, there exists $\delta>0$ such that when $|z-z_0|<\delta$,

$$|\mu_{M,N}(z)-\mu_{M,N}(z_0)|<\varepsilon.$$

Fix arbitrary $x_0\in\mathcal{B}$, for any $x\in\mathcal{B}$ satisfying $|x-x_0|<\delta/\lambda$, using that $f_t(x_{t-1}^{(x,j)},w_{t-1}^{(j)})=x$ and $f_t(x_{t-1}^{(x_0,j)},w_{t-1}^{(j)})=x_0$ we have

$$\left|x_{t-1}^{(x,j)}-x_{t-1}^{(x_0,j)}\right|=|f_t^{-1}(x,w_{t-1}^{(j)})-f_t^{-1}(x_0,w_{t-1}^{(j)})|\le\left\|\frac{\partial}{\partial x}f_t^{-1}\right\|\cdot|x-x_0|<\delta,$$

and thus

$$\left|s^M\circ a_t(\mu_{M,N})(x)-s^M\circ a_t(\mu_{M,N})(x_0)\right|\le\frac{1}{M}\sum_{j=1}^{M}\left|\mu_{M,N}\left(x_{t-1}^{(x,j)}\right)-\mu_{M,N}\left(x_{t-1}^{(x_0,j)}\right)\right|<\varepsilon.$$

It remains to show that $s^M\circ a_t(\mu_M)\in\mathcal{P}_U(\mathcal{B})$. In fact, given any $x_1,x_2\in\mathcal{B}$,

$$s^M\circ a_t(\mu_M)(x_1)-s^M\circ a_t(\mu_M)(x_2)=\frac{1}{M}\sum_{j=1}^{M}\left(\mu_M\left(x_{t-1}^{(x_1,j)}\right)-\mu_M\left(x_{t-1}^{(x_2,j)}\right)\right).$$

For any $\varepsilon>0$, from the uniformly continuity of μ_M, there exists $\delta>0$, such that for any $z_1,z_2\in\mathcal{B}$ with $|z_1-z_2|<\delta$, $|\mu_M(z_1)-\mu_M(z_2)|<\varepsilon$. Let $\tilde{\delta}=\frac{\delta}{\lambda}$, then $|x_1-x_2|<\tilde{\delta}$ implies that

$$\left|x_{t-1}^{(x_1,j)}-x_{t-1}^{(x_2,j)}\right|=\left|f_t^{-1}(x_1,w_{t-1}^{(j)})-f_t^{-1}(x_2,w_{t-1}^{(j)})\right|\leq\left\|\frac{\partial}{\partial x}(f_t^{-1})\right\|\cdot|x_1-x_2|<\delta.$$

Hence

$$\left|s^M\circ a_t(\mu_M)(x_1)-s^M\circ a_t(\mu_M)(x_2)\right|=\left|\frac{1}{M}\sum_{j=1}^{M}\left(\mu_M\left(x_{t-1}^{(x_1,j)}\right)-\mu_M\left(x_{t-1}^{(x_2,j)}\right)\right)\right|<\varepsilon.$$

The proof is complete. □

Lemma 3.3. *For $\{\nu_{M,N}\}_{M,N=1}^{\infty}\in\mathcal{P}_C(\mathcal{B})$ and $\nu_M\in\mathcal{P}_U(\mathcal{B})$ with $\lim_{N\to\infty}\nu_{M,N}=\nu_M$ for each $M\in\mathbb{N}$, it holds that*

$$\lim_{N\to\infty}T^N(\nu_{M,N})=\nu_M,\quad\forall M\in\mathbb{M}. \tag{3.11}$$

Proof. For any $x_t\in\mathcal{B}$,

$$\left|T^N(\nu_{M,N})(x_t)-\nu_M(x_t)\right|\leq\left|T^N(\nu_{M,N})(x_t)-T^N(\nu_M)(x_t)\right|+\left|T^N(\nu_M)(x_t)-\nu_M(x_t)\right|. \tag{3.12}$$

Since $\lim_{N\to\infty}\nu_{M,N}=\nu_M$, for any $\varepsilon>0$, there exists $N_1=N_1(M)>0$ such that when $N>N_1$,

$$|\nu_{M,N}(x_t)-\nu_M(x_t)|<\frac{\varepsilon}{2}.$$

Thus because of the linearity of T^N we have

$$\left|T^N(\nu_{M,N})(x_t)-T^N(\nu_M)(x_t)\right|=\left|T^N(\nu_{M,N}-\nu_M)(x_t)\right|<\frac{\varepsilon}{2}. \tag{3.13}$$

For the second term on the right hand side of inequality (3.12), since T^N is the linear interpolation operator and ν_M is uniformly continuous, for any $\varepsilon>0$, there exists $N_2=N_2(M)>0$ such that when $N>N_2$ we have

$$\left|T^N(\nu_M)(x_t)-\nu_M(x_t)\right|<\frac{\varepsilon}{2}. \tag{3.14}$$

In summary letting $N_0=\max\{N_1,N_2\}$ we have by (3.12), (3.13) and (3.14) that for any $\varepsilon>0$,

$$\left|T^N(\nu_{M,N})(x_t)-\nu_M(x_t)\right|<\varepsilon,\quad\forall N>N_0,\quad\forall x_t\in\mathcal{B},\quad\forall M\in\mathbb{N}. \tag{3.15}$$

The proof is complete. □

We next prove the weak convergence of the operator $\kappa_t^{M,N}$ to κ_t. Letting $\kappa_t^{M,N}$ and κ_t be the composition operators defined as above, we have the following theorem.

Theorem 3.1 (Local convergence). *Assume that the transition kernel $p(x_t|x_{t-1})$ is Feller and $p(y_t|x_t)$ is bounded, uniformly continuous, and strictly positive. Also assume that $\left\|\frac{\partial}{\partial x}f_t^{-1}\right\|$ is bounded. Then, for any $\{\mu_{M,N}\}_{M,N=1}^{\infty}\in\mathcal{P}_C(\mathcal{B})$ and $\mu_M,\mu\in\mathcal{P}_U(\mathcal{B})$ with $\lim_{N\to\infty}\mu_{M,N}=\mu_M$ for each $M\in\mathbb{N}$ and $\lim_{M\to\infty}\mu_M\overset{\varphi}{=}\mu$, it holds that*

$$\lim_{M\to\infty}\lim_{N\to\infty}\kappa_t^{M,N}(\mu_{M,N})\overset{\varphi}{=}\kappa_t(\mu). \tag{3.16}$$

Proof. Given $\lim_{M\to\infty}\mu_M\overset{\varphi}{=}\mu$, by Lemma 3.1,

$$\lim_{M\to\infty}s^M\circ a_t(\mu_M)\overset{\varphi}{=}a_t(\mu). \tag{3.17}$$

Given $\lim_{N\to\infty}\mu_{M,N}=\mu_M$ for each $M\in\mathbb{N}$, by Lemma 3.2 we have $\lim_{N\to\infty}s^M\circ a_t(\mu_{M,N})=s^M\circ a_t(\mu_M)$ and $s^M\circ a_t(\mu_{M,N})\in\mathcal{P}_C(\mathcal{B})$, $s^M\circ a_t(\mu_M)\in\mathcal{P}_U(\mathcal{B})$. Thus by letting $\nu_{M,N}\doteq s^M\circ a_t(\mu_{M,N})$ and $\nu_M\doteq s^M\circ a_t(\mu_M)$ in Lemma 3.3 we get

$$\lim_{N\to\infty}T^N\circ s^M\circ a_t(\mu_{M,N})=s^M\circ a_t(\mu_M). \tag{3.18}$$

Eqs. (3.17) and (3.18) together give

$$\lim_{M\to\infty}\lim_{N\to\infty}T^N\circ s^M\circ a_t(\mu_{M,N})\overset{\varphi}{=}a_t(\mu).$$

Therefore it follows directly from (3.6) that

$$\lim_{M\to\infty}\lim_{N\to\infty}b_t\circ T^N\circ s^M\circ a_t(\mu_{M,N})\overset{\varphi}{=}b_t\circ a_t(\mu).$$

The proof is complete. □

To prove the global weak convergence result, we also need the following lemma.

Lemma 3.4. *Assume $p(y_t|x_t)$ is bounded, uniformly continuous, and strictly positive. For $\{\gamma_{M,N}\}_{M,N=1}^{\infty}\in\mathcal{P}_C(\mathcal{B})$ and $\gamma_M\in\mathcal{P}_U(\mathcal{B})$ with $\lim_{N\to\infty}\gamma_{M,N}=\gamma_M$ for each $M\in\mathbb{N}$, if there exists a $\xi_0>0$ such that $\int_{\mathcal{B}}p(y_t|x_t)\gamma_M(x_t)dx_t\geq\xi_0$, then we have*

$$\lim_{N\to\infty}b_t(\gamma_{M,N})=b_t(\gamma_M)\in\mathcal{P}_U(\mathcal{B}),\quad\forall M\in\mathbb{N}.$$

Proof. Since $\lim_{N\to\infty}\gamma_{M,N}=\gamma_M$ for each $M\in\mathbb{N}$, then for any $0<\varepsilon<\frac{\xi_0}{2}$, there exists N_0, such that when $N>N_0$,

$$|\gamma_{M,N}(x_t)-\gamma_M(x_t)|<\varepsilon,\quad\forall x_t\in\mathcal{B}\ \text{ and }\ \forall M\in\mathbb{N}.$$

It then follows that $\left|\int_{\mathcal{B}}p(y_t|x_t)(\gamma_{M,N}(x_t)-\gamma_M(x_t))dx_t\right|<\varepsilon$ and

$$\int_{\mathcal{B}}p(y_t|x_t)\gamma_{M,N}(x_t)dx_t>\int_{\mathcal{B}}p(y_t|x_t)\gamma_M(x_t)dx_t-\varepsilon>\frac{\xi_0}{2}.$$

Thus for any $x_t \in B$ when $N > N_0$ we have

$$\begin{aligned}
&|b_t(\gamma_{M,N})(x_t) - b_t(\gamma_M)(x_t)| \\
=&p(y_t|x_t)\cdot\left|\frac{\gamma_{M,N}(x_t)\cdot\int_{\mathcal{B}}p(y_t|x_t)\gamma_M(x_t)dx_t - \gamma_M(x_t)\cdot\int_{\mathcal{B}}p(y_t|x_t)\gamma_{M,N}(x_t)dx_t}{\int_{\mathcal{B}}p(y_t|x_t)\gamma_{M,N}(x_t)dx_t\cdot\int_{\mathcal{B}}p(y_t|x_t)\gamma_M(x_t)dx_t}\right| \\
\leq&\left|\frac{\int_{\mathcal{B}}p(y_t|x_t)\left(\gamma_M(x_t)-\gamma_{M,N}(x_t)\right)dx_t}{\int_{\mathcal{B}}p(y_t|x_t)\gamma_{M,N}(x_t)dx_t\cdot\int_{\mathcal{B}}p(y_t|x_t)\gamma_M(x_t)dx_t}\right| + \left|\frac{\gamma_M(x_t)-\gamma_{M,N}(x_t)}{\int_{\mathcal{B}}p(y_t|x_t)\gamma_M(x_t)dx_t}\right| \\
<&\left(\frac{4}{\xi_0^2}+\frac{2}{\xi_0}\right)\varepsilon. \qquad (3.19)
\end{aligned}$$

Therefore $\lim_{N\to\infty} b_t(\gamma_{M,N}) = b_t(\gamma_M)$. It remains to show that $b_t(\gamma_M) \in \mathcal{P}_U(\mathcal{B})$. In fact, by the definition of b_t, we have for any $x_t^{(1)}, x_t^{(2)} \in \mathcal{B}$,

$$\left|b_t(\gamma_M)(x_t^{(1)}) - b_t(\gamma_M)(x_t^{(2)})\right| \leq \frac{2}{\xi_0}\left|p(y_t|x_t^{(1)})\gamma_M(x_t^{(1)}) - p(y_t|x_t^{(2)})\gamma_M(x_t^{(2)})\right|.$$

From the uniformly continuity property and the boundedness of γ_M and $p(x_t|x_{t-1})$ that for any $\varepsilon > 0$, there exists $\delta > 0$, such that when $x_t^{(1)}, x_t^{(2)} \in \mathcal{B}$ with $|x_t^{(1)} - x_t^{(2)}| < \delta$, we have $|\gamma_M(x_t^{(1)}) - \gamma_M(x_t^{(2)})| < \frac{\varepsilon}{2}$ and $|p(x_t^{(1)}) - p(x_t^{(2)})| < \frac{\varepsilon}{2}$. Thus,

$$\left|b_t(\gamma_M)(x_t^{(1)}) - b_t(\gamma_M)(x_t^{(2)})\right| < C\varepsilon.$$

This completes the proof. □

Note 3.1. When the standing assumption (A1) holds, it follows immediately by the definition of s^M that there exists a $\xi_0 > 0$ such that $\int_{\mathcal{B}} p(y_t|x_t)s^M \circ p(x_t|\mathcal{I}_{t-1})dx_t > \xi_0$ for M sufficient large.

Applying Theorem 3.1 to the context of filtering problems, we can obtain the weak convergence of our implicit filtering simulation to the Bayesian optimal filter. Our main result of this work is stated in the following theorem. For simplicity, we define two new operators θ_t^M and $\theta_{1:t}^M$ to be

$$\theta_t^M = b_t \circ s^M \circ a_t \quad \text{and} \quad \theta_{1:t}^M = \theta_t \circ \theta_{t-1} \circ \cdots \circ \theta_1.$$

Theorem 3.2 (Global convergence)**.** *Assume that the transition kernel $p(x_t|x_{t-1})$ is Feller and $p(x_t|x_{t-1})$ is bounded, uniformly continuous, and strictly positive. Also assume that $\left\|\frac{\partial}{\partial x}f_t^{-1}\right\|$ is bounded. Then*

$$\lim_{N\to\infty}\kappa_{1:t}^{M,N}(p_{0|0}) = \theta_{1:t}^M(p_{0|0}) \quad \forall M \in \mathbb{N}, \quad \textit{and} \quad \lim_{M\to\infty}\theta_{1:t}^M(p_{0|0}) \overset{\varphi}{=} \kappa_{1:t}(p_{0|0}),$$

which implies that

$$\lim_{M\to\infty}\lim_{N\to\infty}\rho_{t|t} \overset{\varphi}{=} p_{t|t}.$$

Proof. To prove Theorem 3.2, we use induction method.

(1) $t=1$: choose $\mu_{M,N}=\mu_M=\mu=p_{0|0}$ in Eq. (3.16). It is obviously that

$$\lim_{N\to\infty}\mu_{M,N}=\mu_M,\quad \lim_{M\to\infty}\mu_M\overset{\varphi}{=}\mu,$$

and $\mu_M=p_{0|0}$ is uniformly continuous. By Lemmas 3.2, 3.3 and 3.4 and Note 3.1,

$$\lim_{N\to\infty}\kappa_1^{M,N}(p_{0|0})=\theta_1^M(p_{0|0})\in\mathcal{P}_U(\mathcal{B}),\quad \forall M\in\mathbb{N}.$$

By Lemma 3.1 and the continuity of b_t, we have

$$\lim_{M\to\infty}\theta_1^M(p_{0|0})\overset{\varphi}{=}\kappa_1(p_{0|0})=p_{1|1}.$$

It then follows from Theorem 3.1 that

$$\lim_{M\to\infty}\lim_{N\to\infty}\kappa_1^{M,N}(p_{0|0})\overset{\varphi}{=}\kappa_1(p_{0|0})=p_{1|1}.$$

(2) Assume that

$$\lim_{N\to\infty}\kappa_{1:t-1}^{M,N}(p_{0|0})=\theta_{1:t-1}^M(p_{0|0})\in\mathcal{P}_U(\mathcal{B})\ \ \forall M\in\mathbb{N},\quad\text{and}\quad \lim_{M\to\infty}\theta_{1:t-1}^M(p_{0|0})\overset{\varphi}{=}\kappa_{1:t-1}(p_{0|0}).$$

We choose $\mu_{M,N}=\kappa_{1:t-1}^{M,N}(p_{0|0})$ and $\mu=\kappa_{1:t-1}(p_{0|0})=p_{t-1|t-1}$ in Eq. (3.16) and $\mu_M=\theta_{1:t-1}^M(p_{0|0})$ in Theorem 3.1. From the assumption,

$$\lim_{N\to\infty}\mu_{M,N}=\mu_M\quad\text{and}\quad\lim_{M\to\infty}\mu_M\overset{\varphi}{=}\mu.$$

By Lemmas 3.2, 3.3 and 3.4 and Note 3.1 that

$$\lim_{N\to\infty}\kappa_{1:t}^{M,N}(p_{0|0})=\lim_{N\to\infty}b_t\circ T^N\circ s^M\circ a_t(\kappa_{1:t-1}^{M,N}(p_{0|0}))=\theta_{1:t}^M(p_{0|0})\in\mathcal{P}_U(\mathcal{B}).\tag{3.20}$$

By Lemma 3.1 and the continuity of b_t we have

$$\lim_{M\to\infty}\theta_{1:t}^M(p_{0|0})\overset{\varphi}{=}p_{t|t}.\tag{3.21}$$

Therefore it follows from Theorem 3.1 that

$$\lim_{M\to\infty}\lim_{N\to\infty}\kappa_t^{M,N}(\kappa_{1:t-1}^{M,N}(p_{0|0}))\overset{\varphi}{=}\kappa_t(p_{t-1|t-1})=\kappa_{1:t}(p_{0|0})=p_{t|t}.$$

The proof is complete. □

4 Numerical experiments

In this section, we present two numerical examples to demonstrate the efficiency of our method. The first example involves a one dimensional nonlinear system and measurement equation while the second is a 2-D bearing-only tracking problem. We shall compare our method with the standard EKF and particle filter. In the particle filter method we use sequential important sampling with resampling (SIR) and the Newton's method as the nonlinear solver in our implicit filter method.

Example 4.1. In this example, we consider the following nonlinear model

$$x_k = 40\cdot\tan(x_{k-1}+10)+50w_{k-1}, \tag{4.1a}$$

$$y_k = 40\cdot\frac{x_k}{2000+x_k}+v_k, \tag{4.1b}$$

where w_k and v_k are two independent zero-mean white noise processes with variance 1.0, y_k is the noise perturbed observation of x_k. The initial position is taken to be $x_0=2$ and Fig. 1 shows a 50 step realization of the state equation in model (4.1).

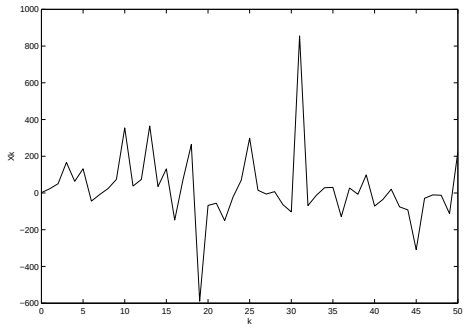

Figure 1: Original position.

Figs. 2, 3 and 4 are the simulation results obtained by using EKF and particle filter and our implicit particle filter method, respectively. The true state is represented by blue diamonds while simulation results are given as red "stars" and connected by solid lines. The prior pdf $p(x_0)$ is initialized with the standard normal distribution with the mean value x_0 and the variance 1.0.

In particle filter method, we use 500 particles (sample points) to represent the pdf and the running time is 1.711293 seconds with an Intel(R) Core(TM)2 Duo CPU P8700@2.53GHZ 2.53Ghz. In our implicit filter, we use 100 nodes to partition the region and the number of Monte-Carlo samples is $M=10$. The running time of the implicit algorithm is 1.130215 seconds.

Form the three figures, one can see that when the variation between two consecutive points is not very large, all three methods produce very accurate approximations to the true state. On the other hand, when the true state has very large variations at some time

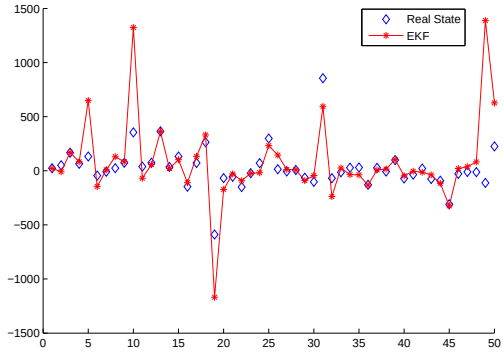

Figure 2: Extended Kalman filter.

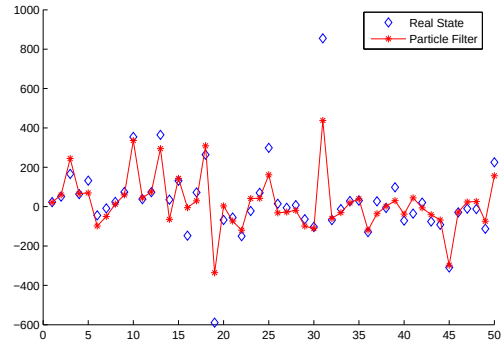

Figure 3: Particle filter.

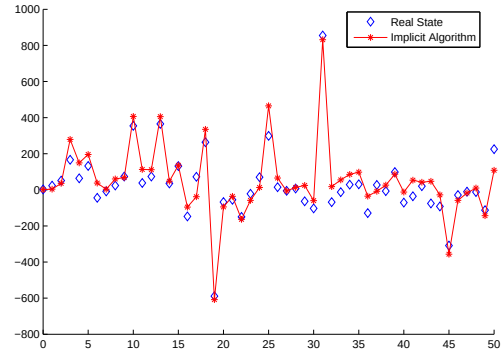

Figure 4: Implicit algorithm.

steps, i.e., the state variable has a large jump from its previous state, both EKF and particle filter fail to produce accurate approximations. However, our implicit particle algorithm still produces accurate estimations at these points.

Example 4.2. In this example, we consider the following bearing-only tracking problem

$$dX_1(t) = -\alpha X_2(t)dt + \beta \frac{X_1(t)}{(X_1(t))^2 + (X_2(t))^2} + \sigma_1 dW_1(t), \tag{4.2a}$$

$$dX_2(t) = \alpha X_1(t)dt + \beta \frac{X_2(t)}{(X_1(t))^2 + (X_2(t))^2} + \sigma_2 dW_2(t), \tag{4.2b}$$

where $W_1(t)$ and $W_2(t)$ are two independent Brownian Motions. This stochastic dynamical system may serve to model the motion of a ship which moves with a constant radial and angular velocity, perturbed by a white noise. The observations are collected by a detector located at a platform with time intervals of length $\Delta = 0.05$ and the data are angular measurements corrupted by noise.

To approximate the state variables $X = (X_1, X_2)$, we discretize the dynamical system (4.2) in time and obtain a discrete nonlinear filtering problem. Let $x_k = (x_k^1, x_k^2)$. We have the discrete system model

$$x_k^1 = x_{k-1}^1 - \alpha\Delta \cdot x_{k-1}^2 + \beta\Delta \cdot \frac{x_{k-1}^1}{(x_{k-1}^1)^2 + (x_{k-1}^2)^2} + \sigma_1\sqrt{\Delta} \cdot w_{k-1}^1, \tag{4.3a}$$

$$x_k^2 = x_{k-1}^2 + \alpha\Delta \cdot x_{k-1}^1 + \beta\Delta \cdot \frac{x_{k-1}^2}{(x_{k-1}^1)^2 + (x_{k-1}^2)^2} + \sigma_2\sqrt{\Delta} \cdot w_{k-1}^2. \tag{4.3b}$$

The mathematical formula for the measurement equation is given by

$$y_k = \arctan\left(\frac{x_k^2 - x_{platform}^2}{x_k^1 - x_{platform}^1}\right) + \sqrt{\Delta} v_k, \tag{4.4}$$

where $x_{platform} = (x_{platform}^1, x_{platform}^2)$ is the location of the platform where a detector is placed.

In the numerical simulations the model parameters are chosen as $\alpha = 5$, $\beta = 2$ and $\sigma_1 = \sigma_2 = 8$. Fig. 5 gives the target path in the x-y plan, with the position of the target at each time step shown by a diamond. The location of the detector platform is chosen as $x_{platform} = (-15, -15)$, marked by a red box.

The problem is initialized with a best guess of the target position at the initial time, which is $(x_0^1, x_0^2) = (0.5, 0.5)$. In this example, we use 8000 particles (sample points) to represent the pdf in the particle filter method and the running time is 45.7266 seconds. In the implicit filter, we use 1600 nodes to partition the region and the number of Monte-Carlo samples is $M = 10$. The running time of the implicit algorithm is 41.8192 seconds. Fig. 6 shows the simulation results of observation angle using EKF, particle filter and our implicit particle filter method. From this figure, one can see that both particle filter and the implicit particle filter produce good approximation for the relative observation angle of the target. Although the EKF provides the trend of the movement of the target, the estimation is a few steps delayed from the true target observation angle.

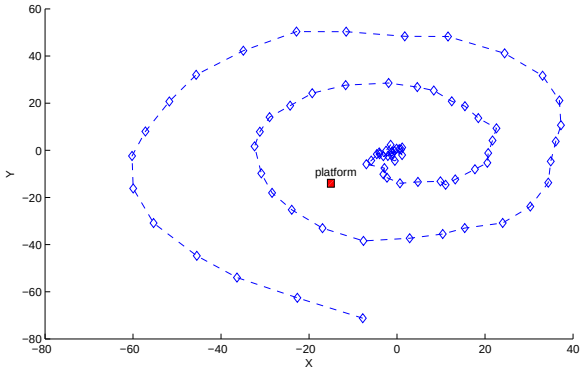

Figure 5: Target positions.

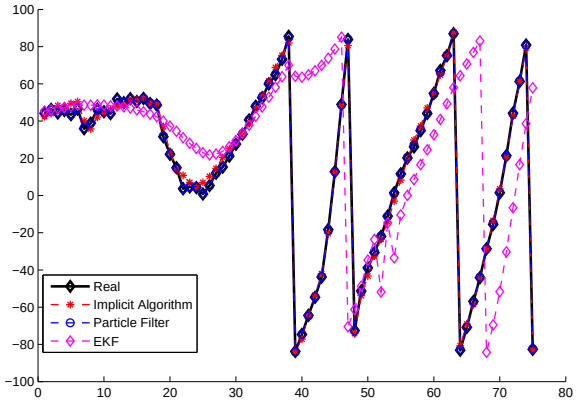

Figure 6: Comparison result for the observation angle.

Fig. 7 shows the results of system state simulations using EKF while Fig. 8 compares the performance between particle filter and the implicit particle filter. Clearly both the particle filter and implicit particle filter outperform EKF. While the accuracy of the particle filter and the implicit particle filter are very close to each other at the initial stage, the implicit particle filter becomes more accurate at the final stage of time interval.

5 Concluding remarks

We have developed an implicit filter algorithm for nonlinear filtering problems. The essential idea is to evaluate the probability distribution of the current state in the prediction step by evaluating the previous state through the state equation, given the valuate of the current state and a sample of the noise. Through rigorous analysis we proved the convergence of the algorithm. Our numerical experiments indicate that our method is more

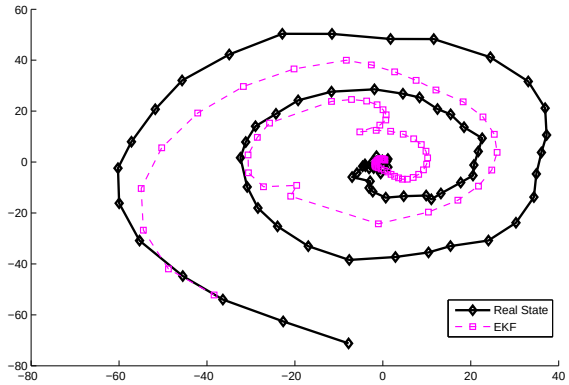

Figure 7: Simulation result of EKF.

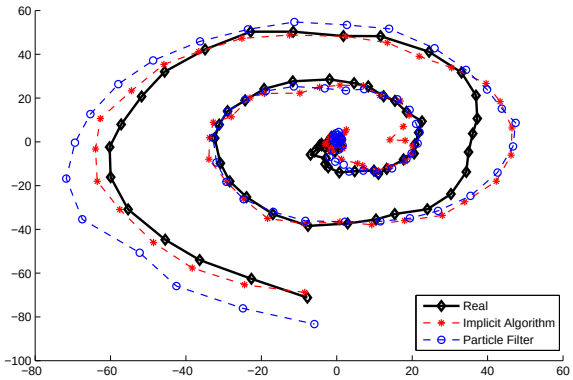

Figure 8: Comparison result between particle filter and implicit algorithm.

accurate than the standard Kalman filter. The numerical experiments also show that as an implicit scheme, our method is more stable than the standard particle filter method for long term simulations. Finally it needs to be pointed out that our method is a grid method in which the probability distributions are evaluated at all grid points. As such, the computing cost will increase exponentially as the dimension increases. Thus our method at the current form is suitable for only low dimension problems such as target tracking problems. In future research, we plan to improve the algorithm by adding an adaptive mechanism to it so that fewer grid points will be selected adaptively.

Acknowledgments

This research is partially supported by National Science Foundation under grant number DMS0914554, AFOSR under grant number FA9550-12-1-0281, and by Guangdong Provin-

cial Government of China through the Computational Science Innovative Research Team program.

References

[1] D. Alspach and H. Sorenson. Nonlinear bayesian estimation using gaussian sum approximations. IEEE Trans. Automatic Control, 17:439–448, 1972.
[2] Y. Bar-Shalom and T. E. Fortmann. Tracking and data association, in: Mathematics in Science and Engineering, volume 179. Academic Press Inc., San Diego, CA, 1988.
[3] M. Bolić, P. M. Djurić, and S. Hong. Resampling algorithms and architectures for distributed particle filters. IEEE Trans. Signal Process., 53(7):2442–2450, 2005.
[4] Z. Chen. Bayesian filtering: from kalman filters to particle filters, and beyond. Unpublished manuscript, 2011.
[5] D. Crisan and O. Obanubi. Particle filters with random resampling times. Stochastic Process. Appl., 122(4):1332–1368, 2012.
[6] D. Crisan. Exact rates of convergence for a branching particle approximation to the solution of the Zakai equation. Ann. Probab., 31(2):693–718, 2003.
[7] D. Crisan and A. Doucet. A survey of convergence results on particle filtering methods for practitioners. IEEE Trans. Sig. Proc., 50(3):736–746, 2002.
[8] J. Duník, M. Šimandl, and O. Straka. Unscented Kalman filter: aspects and adaptive setting of scaling parameter. IEEE Trans. Automat. Control, 57(9):2411–2416, 2012.
[9] G. A. Einicke. Asymptotic optimality of the minimum-variance fixed-interval smoother. IEEE Trans. Signal Process., 55(4):1543–1547, 2007.
[10] E. Gobet, G. Pagès, H. Pham, and J. Printems. Discretization and simulation of the Zakai equation. SIAM J. Numer. Anal., 44(6):2505–2538, 2006.
[11] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE PROCEEDING-F, 140(2):107–113, 1993.
[12] Y. Hu, G. Kallianpur, and J. Xiong. An approximation for the Zakai equation. Appl. Math. Optim., 45(1):23–44, 2002.
[13] A. H. Jazwinski. Stochastic Processing and Filtering Theory, volume 64. Academic Press, New York, 1973.
[14] S. J. Julier and J. J. LaViola Jr. On Kalman filtering with nonlinear equality constraints. IEEE Trans. Signal Process., 55(6, part 2):2774–2784, 2007.
[15] K. Xiong, L. Liu, and Y. Liu. Robust extended Kalman filtering for nonlinear systems with multiplicative noises. Optimal Control Appl. Methods, 32(1):47–63, 2011.
[16] R. E. Kalman and R. S. Bucy. New results in linear filtering and prediction theory. Trans. ASME Ser. D. J. Basic Engrg., 83:95–108, 1961.
[17] X. Luo, I. M. Moroz, and I. Hoteit. Scaled unscented transform Gaussian sum filter: theory and application. Phys. D, 239(10):684–701, 2010.
[18] C. J. Masreliez. Approximate non-gaussian filtering with linear state and observation relations. IEEE Trans. Automatic Control, 20:107–110, 1975.
[19] C. J. Masreliez and R. D. Martin. Robust Bayesian estimation for the linear model and robustifying the Kalman filter. IEEE Trans. Automatic Control, AC-22(3):361–371, 1977.
[20] M. Nagata and Y. Sawaragi. Error analysis of the Schmidt-Kalman filter. Internat. J. Systems Sci., 7(7):769–778, 1976.
[21] U. Orguner and F. Gustafsson. Target tracking with particle filters under signal propagation delays. IEEE Trans. Signal Process., 59(6):2485–2495, 2011.

[22] C. F. Price. An analysis of the divergence problem in the Kalman filter. IEEE Trans. Automatic Control, AC-13:699–702, 1968.
[23] J. Printems. On the discretization in time of parabolic stochastic partial differential equations. M2AN Math. Model. Numer. Anal., 35(6):1055–1078, 2001.
[24] G. Terejanu, P. Singla, T. Singh, and P. D. Scott. Adaptive Gaussian sum filter for nonlinear Bayesian estimation. IEEE Trans. Automat. Control, 56(9):2151–2156, 2011.
[25] Y. Ulker and B. Gunsel. Multiple model target tracking with variable rate particle filters. Digit. Signal Process., 22(3):417–429, 2012.
[26] M. Zakai. On the optimal filtering of diffusion processes. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 11:230–243, 1969.
[27] H. Zhang and D. Laneuville. Grid based solution of zakai equation with adaptive local refinement for bearing-only tracking. IEEE Aerospace Conference, 2008.