Available online at www.sciencedirect.com
SCIENCE DIRECT®
Automatica 41 (2005) 945–956
automatica
www.elsevier.com/locate/automatica

Approximate nonlinear filtering and its application in navigation☆

Babak Azimi-Sadjadi[a, b], P.S. Krishnaprasad[b, *]

[a] *Electrical, Computer, and Systems Engineering Department, Rensselaer Polytechnic Institute, NY 12180, USA*
[b] *Institute for Systems Research, University of Maryland, College Park, MD 20742, USA*

Received 5 April 2002; received in revised form 16 September 2004; accepted 14 December 2004
Available online 19 March 2005

Abstract

In this paper, we introduce for the first time particle filtering for an exponential family of densities. We prove that under certain conditions the approximated conditional density of the state converges to the true conditional density. In the realistic setting where the conditional density does not lie in an exponential family but stays close to it, we show that under certain assumptions the error of the estimate given by an approximate nonlinear filter (which we call the *projection particle filter*), is bounded. We use projection particle filtering in state estimation for a combination of inertial navigation system (INS) and global positioning system (GPS), referred to as integrated INS/GPS. We illustrate via numerical experiments that projection particle filtering outperforms regular particle filtering in navigation performance, and extended Kalman filter as well when satellite loss-of-lock occurs.

Keywords: Nonlinear filtering; Sequential Monte Carlo; GPS; INS

1. Introduction

Filtering problems consist of "estimating" a process $\{\mathbf{x}_t\}$ (or something about it) given a related observed process, $\{\mathbf{y}_t\}$ (Davis & Marcus, 1981). The observation is available on an interval, i.e., $\{\mathbf{y}_s, 0 \leqslant s < t\}$ and the function of the state is estimated at time t. Except for the linear Gaussian system and very special cases in nonlinear settings, estimating the state given the observations results in an infinite dimensional filter (Maybeck, 1982). Therefore, approximation methods of finite dimension are very appealing.

The most widely used approximate filtering method is the Extended Kalman Filter (EKF), which is a heuristic method based on the linearization of the state dynamics and the observation near a nominal path (Maybeck, 1982).

EKF is computationally simple but, the convergence of the estimated conditional density to the actual conditional density is not guaranteed.

Projection Filtering (PrF) is another approximation method (Brigo, Hanzon, & LeGland, 1995; Brigo, 1996; Beard, Gunther, Lawton, & Stirling, 1999). In PrF it is assumed that the conditional density of the state of the system can be approximated by a member of a parametric family of densities. In this case, estimating the conditional density is equivalent to estimating the parameter of the family. In Brigo et al. (1995) and Brigo (1996) the exponential family of densities is chosen as the parametric family. In Beard et al. (1999) the approach is different; there a Galerkin approximation is used to solve the Fokker–Planck equation (Maybeck, 1982).

An entirely different approach to approximate the conditional density is simulation based filtering. Grid-free simulation based filtering, now known by many different names such as Particle Filtering (PaF) (Del Moral,

☆ This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor R. Boel under the direction of Editor I. Petersen. This research was supported in part by Army Research Office under ODDR&E MURI97 Program Grant No. DAAG55-97-1-0114 to the Center for Dynamics and Control of Smart Structures (through Harvard University), and under ODDR&E MURI01 Program Grant No. DAAD19-01-1-0465 to the Center for Communicating Networked Control Systems (through Boston University), and by the National Science Foundation Learning and Intelligent Systems Initiative Grant CMS9720334.

* Corresponding author. Tel.: +1 301 405 6843; fax: +1 301 314 9920.
E-mail addresses: babak@ecse.rpi.edu (B. Azimi-Sadjadi), krishna@isr.umd.edu (P.S. Krishnaprasad).

0005-1098/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2004.12.013

1996; Doucet, de Freitas, & Gordon, 2001; Arulampalam, Maskell, Gordon, & Clapp, 2002) the Condensation Algorithm (Isard & Blake, 1996), the Sequential Monte Carlo (SMC) Method (Fearnhead, 1998), and Bayesian Bootstrap Filtering (Gordon, Salmond, & Smith, 1993) was first introduced in Gordon et al. (1993) and then it was rediscovered independently in (Isard & Blake, 1996) and (Kitagawa, 1996). Henceforth we refer to this filtering method as Particle Filtering. The results in Gordon et al. (1993) are an extension of the results in Rubin (1988) and Smith and Gelfand (1992) to the dynamic case and the idea is based on a method called Sampling/Importance Resampling (SIR). SIR is a key element of the grid-free simulation-based filtering methods which allows these methods to have automatically high resolution grids in the areas where the conditional density is significant and low resolution in the areas where the conditional density is small.

In the cases where we have some prior information about the distribution, we should expect to achieve higher performance if we take this information into account. By higher performance, we mean a reduction in the computational cost and an increase in the convergence rate. Here we assume that the conditional distribution has a density in an exponential family of densities, or at least stays close to it in a sense that we will define. Using this assumption, we replace the empirical distribution in Del Moral (1996) with the Maximum Likelihood Estimate (MLE) of the parameters of an exponential density. We call this new method *projection particle filtering*. In Theorem 5 we show that if the conditional density of the state given the observations lies in an exponential family of densities then the estimated conditional density converges to the true conditional density in a sense that will be defined. In Theorem 8 for the case where the true conditional density stays close to an exponential family of densities we show that the error of the estimate given by projection particle filtering is bounded. Approximating the conditional density of the state given the observation by an exponential family of densities is addressed in Brigo (1996). Unlike our approach, in Brigo (1996) no estimates on the closeness of the true conditional density to the approximate conditional density is reported.

One of the applications of the new particle filtering method introduced in this paper is position estimation in an integrated INS/GPS system. In Carvalho, Del Moral, Monin, and Salut (1997) it was shown that when the number of GPS satellites visible to a receiver drops below a critical number, there, EKF could no longer provide a reasonable estimate for the position, in fact the estimate given by EKF would diverge. Here we show for these critical cases, projection PaF for an exponential family of densities can provide an estimate of the position that is accurate as well as smooth. We show that, on average, this filter can perform better than regular PaF.

In this paper, Section 2 states the nonlinear filtering problem. In Section 3 we describe PaF. In Sections 4 and 5 we introduce the new projection PaF algorithm and we state the main results of this paper. In Section 6 we apply projection PaF to an exponential family of densities to estimate the position in an integrated INS/GPS. We conclude this paper in Section 7.

2. Nonlinear filtering, problem setup

We assume that all stochastic processes are defined on a fixed probability space (Ω, F, P), and a finite time interval, $[0, T]$, on which an increasing family of σ-fields, $\{\mathscr{F}_t, 0 \leqslant t \leqslant T\}$ is defined. It is assumed that each process, $\{\mathbf{x}_t\}$, is adapted to $\mathscr{F}_t$, i.e., $\{\mathbf{x}_t\}$ is $\mathscr{F}_t$-measurable for all t. We assume that $\{\mathbf{x}_t\}$ is a vector diffusion process of the form

$$\mathbf{x}_t = \mathbf{x}_0 + \int_0^t \mathbf{f}_s(\mathbf{x}_s)\,\mathrm{d}s + \int_0^t G_s(\mathbf{x}_s)\,\mathrm{d}\mathbf{w}_s, \tag{1}$$

where $\mathbf{x}_t \in \Re^n$, and $\mathbf{w}_t \in \Re^q$ is a vector from an independent Brownian motion process; the second integral is in the Ito sense (Sobczyk, 1991). Function $\mathbf{f}_t(\cdot)$ and matrix $G_t(\cdot)$ have the proper dimensions. The observation, $\mathbf{y}_t$, is a discrete time process given by

$$\mathbf{y}_{n\tau} = \mathbf{h}_n(\mathbf{x}_{n\tau}) + \mathbf{v}_n, \tag{2}$$

where $\mathbf{y}_{n\tau} \in \Re^d$, and $\mathbf{v}_n \in \Re^d$ is a discrete time white zero-mean Gaussian noise process with known covariance matrix. Given the distribution of $\mathbf{x}_0$, the state dynamics and observation equations are formally rewritten

$$\begin{aligned} \mathrm{d}\mathbf{x}_t &= \mathbf{f}_t(\mathbf{x}_t)\,\mathrm{d}t + G_t(\mathbf{x}_t)\,\mathrm{d}\mathbf{w}_t, \\ \mathbf{y}_{n\tau} &= \mathbf{h}_n(\mathbf{x}_{n\tau}) + \mathbf{v}_n. \end{aligned} \tag{3}$$

The noise processes $\{\mathbf{w}_t, t \geqslant 0\}$, and $\{\mathbf{v}_n, n = 0, 1, \ldots\}$, and the initial condition $\mathbf{x}_0$ are assumed independent. Q_t and R_n (R_n is assumed invertible for all n) are the covariance matrices of the processes $\mathbf{w}_t$ and $\mathbf{v}_n$, respectively. We have the following extra **assumptions** (Hasminskii, 1980):

A1 (*Local Lipschitz continuity*). $\forall \mathbf{x}, \mathbf{x}' \in B_r$ and $t \in [0, T]$, where B_r is a ball of radius r, we have

$$\begin{aligned} &\|\mathbf{f}_t(\mathbf{x}) - \mathbf{f}_t(\mathbf{x}')\| \leqslant k_r \|\mathbf{x} - \mathbf{x}'\|, \\ &\|G_t(\mathbf{x}) Q_t G_t^{\mathrm{T}}(\mathbf{x}) - G_t(\mathbf{x}') Q_t G_t^{\mathrm{T}}(\mathbf{x}')\| \leqslant k_r \|\mathbf{x} - \mathbf{x}'\|. \end{aligned} \tag{4}$$

A2 (*Non-explosion*). There exists $k > 0$ such that $\forall t \in [0, T]$ and $\forall \mathbf{x} \in \Re^n$,

$$\begin{aligned} &\mathbf{x}^{\mathrm{T}} \mathbf{f}_t(\mathbf{x}) \leqslant k(1 + \|\mathbf{x}\|^2), \\ &\mathrm{trace}(G_t(\mathbf{x}) Q_t G_t^{\mathrm{T}}(\mathbf{x})) \leqslant k(1 + \|\mathbf{x}\|^2). \end{aligned} \tag{5}$$

Under assumptions (A1) and (A2) above, there exists a unique solution $\{\mathbf{x}_t, t \in [0, T]\}$ to the state equation, and $\mathbf{x}_t$ has finite moment of any order (Hasminskii, 1980).

In addition to these, we assume that the probability distribution of the state $\mathbf{x}_t$, given the observation up to time t, $\pi_t(\mathrm{d}\mathbf{x}) = P(\mathbf{x}_t \in \mathrm{d}\mathbf{x} | \mathbf{y}^t)$, where $\mathbf{y}^t = \{\mathbf{y}_n, i = 1, \ldots, n, n\tau < t\}$,

has a density p_t with respect to the Lebesgue measure on $\Re^n$. Then $\{p_t, t>0\}$ satisfies the following PDE and updating equations (Brigo, 1996):

$$\frac{\partial}{\partial t}p_t = \mathscr{L}_t^* p_t \quad n\tau \leqslant t < (n+1)\tau,$$
$$p_{n\tau} = \beta_n \Psi_n p_{n\tau^-}, \tag{6}$$

where $\beta_n = (\int_{\mathbf{x}} \Psi_n(\mathbf{x}) p_{n\tau^-}(\mathbf{x})\,\mathrm{d}\mathbf{x})^{-1}$ is a normalizing factor, $n\tau^-$ is the notation we use for the moment before updating, and

$$\mathscr{L}_t^*(\Phi) = -\sum_{i=1}^n \frac{\partial}{\partial \mathbf{x}_i}[f_t^i \Phi] + \frac{1}{2}\sum_{i,j=1}^n \frac{\partial^2}{\partial \mathbf{x}_i \partial \mathbf{x}_j}[a_t^{ij}\Phi],$$
$$[a_t^{ij}] = G_t Q_t G_t^{\mathrm{T}},$$
$$\Psi_n(\mathbf{x}) \triangleq \exp\left(-\frac{1}{2}(\mathbf{y}_{n\tau} - \mathbf{h}_n(\mathbf{x}))^{\mathrm{T}} R_n^{-1}(\mathbf{y}_{n\tau} - \mathbf{h}_n(\mathbf{x}))\right).$$

Except for the linear Gaussian case and special nonlinear cases, solving (6) constitutes an infinite dimensional filter (Maybeck, 1982). Hence in practice it is necessary to approximate the conditional density in (6). In the projection filtering approximation, one projects (6) to a finite dimensional manifold of densities (Brigo, 1996). See for instance the arguments in the proof of Theorem 5 below and also Brigo (1996).

3. Particle filtering (PaF)

PaF, in brief, is an approximation method that mimics the propagation of the conditional density to calculate the conditional density with a finite number of operations using the Monte Carlo method. The procedure for PaF is as follows (Gordon et al., 1993; Del Moral, 1996):

Algorithm 1. *Particle Filtering*

- *Step* 1: *Initialization. Sample* $\mathbf{x}_0^1, \ldots, \mathbf{x}_0^N$, *N i.i.d. random vectors with the distribution* $P_0(\mathbf{x})$.
- *Step* 2: *Diffusion/Prediction. Find* $\hat{\mathbf{x}}_{n+1}^1, \ldots, \hat{\mathbf{x}}_{n+1}^N$ *from the given* $\mathbf{x}_n^1, \ldots, \mathbf{x}_n^N$ (*customarily called* particles), *using the dynamics in* (3).
- *Step* 3: *Find the empirical distribution*

$$P_{(n+1)^-}^N(\mathbf{x}) = \frac{1}{N}\sum_{j=1}^N \delta_{\hat{\mathbf{x}}_{n+1}^j}(\mathbf{x})$$

- *Step* 4: *Updating/Bayes' Rule.*

$$P_{(n+1)}^N(\mathbf{x}) = \frac{\frac{1}{N}\sum_{j=1}^N \delta_{\hat{\mathbf{x}}_{n+1}^j}(\mathbf{x}) \cdot \Psi_{n+1}(\mathbf{x})}{\frac{1}{N}\sum_{j=1}^N \delta_{\hat{\mathbf{x}}_{n+1}^j}(\hat{\mathbf{x}}_{n+1}^j) \cdot \Psi_{n+1}(\hat{\mathbf{x}}_{n+1}^j)}$$

- *Step* 5: *Resample. Sample* $\mathbf{x}_{n+1}^1, \ldots, \mathbf{x}_{n+1}^N$ *according to* $P_{n+1|n+1}^N(\mathbf{x})$.
- *Step* 6: $n \leftarrow n+1$; *go to Step* (2)

where $\delta_{\mathbf{v}}(\mathbf{w}) = 1$ if $\mathbf{w} = \mathbf{v}$ and 0 otherwise, and $\Psi_n(\mathbf{x})$ is the conditional density of the observation $\mathbf{y}_n$ given $\mathbf{x}$.

Let $\hat{\mathbf{x}}_n^1, \ldots, \hat{\mathbf{x}}_n^N$ be the distinct particles at time n before incorporating the observation. The probability of each particle is $(1/N)$, i.e., they are uniformly distributed. After using the observations, the conditional probability of each particle changes. Some will have small, and some large probabilities. Therefore, in the process of resampling, it is very likely that some particles will never be used and instead some other particles (with high probabilities) will be sampled more than once. Hence, after resampling, some particles have repeated versions, but in the diffusion phase they go through different paths and at the end of the diffusion phase, it is very likely, we will have N distinct particles. This automatically makes the approximation one of the better resolutions in the areas where the probability is higher.

In Del Moral (1996) it is shown that under some conditions, $\lim_{N\to\infty} E(|(1/N)\sum_{i=1}^N f(\hat{\mathbf{x}}_n^i) - E_{P_n}(f(\mathbf{x}))|) = 0$ for every bounded Borel test function, $f(\cdot)$.

It must be mentioned that for a high-dimensional system, getting reasonable accuracy means using a large N, which results in a heavy computational cost.

4. PaF for exponential families of densities

In the PaF method, we saw that the conditional distribution is approximated by the empirical distribution. In most cases, the actual conditional distribution is smooth, unlike the empirical distribution. Intuition suggests that we should make effective use of any available prior knowledge of the distribution properties. Here, we assume that the conditional density lies in a parametric family of densities. With this assumption we will show the convergence of the approximated density to the actual one. After introducing our algorithm, we will present some convergence results, after which we will compare our method with the previously described methods.

For System (3), we assume that the probability density of $\mathbf{x}_t$, given the observation, is in a family of exponential densities, $\mathscr{S}$, defined as follows:

Definition 1. Let $\{c_1(\cdot), \ldots, c_p(\cdot)\}$ be affinely independent[1] scalar functions defined on $\Re^n$. Assume that $\Theta_0 = \{\theta \in \Re^p : \Upsilon(\theta) = \log\int \exp(\theta^{\mathrm{T}}\mathbf{c}(\mathbf{x}))\,\mathrm{d}\mathbf{x} < \infty\}$, is a convex set with a nonempty interior, where $\mathbf{c}(\mathbf{x}) = [c_1(\mathbf{x}), \ldots, c_p(\mathbf{x})]^{\mathrm{T}}$. Then, $\mathscr{S}$ defined as

$$\mathscr{S} = \{p(\cdot, \theta), \theta \in \Theta\}$$
$$p(\mathbf{x}, \theta) := \exp[\theta^{\mathrm{T}}\mathbf{c}(\mathbf{x}) - \Upsilon(\theta)],$$

where $\Theta \subseteq \Theta_0$ is open, is called an exponential family of probability densities.

[1] $\{c_1, \ldots, c_p\}$ are affinely independent if for distinct points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{p+1}$, $\sum_{i=1}^{p+1}\lambda_i \mathbf{c}(\mathbf{x}_i) = 0$ and $\sum_{i=1}^{p+1}\lambda_i = 0$ implies $\lambda_1 = \lambda_2 = \cdots = \lambda_{p+1} = 0$ (Crain, 1976).

We propose the new algorithm below:

Algorithm 2. *Particle filtering for an exponential family of densities.*

- *Step 1*: *Initialization. Sample* $\mathbf{x}_0^1, \ldots, \mathbf{x}_0^N$, *N i.i.d. random vectors with the density,* $p_0(\mathbf{x})$.
- *Step 2*: *Diffusion/Prediction. Find* $\hat{\mathbf{x}}_{n+1}^1, \ldots, \hat{\mathbf{x}}_{n+1}^N$ *from the given* $\mathbf{x}_n^1, \ldots, \mathbf{x}_n^N$, *using the dynamic rule in* (3):
- *Step 3*: *Find the MLE of* $\hat{\theta}_{(n+1)^-}$ *given* $\hat{\mathbf{x}}_{n+1}^1, \ldots, \hat{\mathbf{x}}_{n+1}^N$

$$\hat{\theta}_{(n+1)^-} = \arg\max_{\theta} \prod_{i=1}^{N} \exp(\theta^{\mathrm{T}} \mathbf{c}(\hat{\mathbf{x}}_{n+1}^i) - \Upsilon(\theta)).$$

- *Step 4*: *Updating/Bayes' rule.*

$$p(\mathbf{x}, \hat{\theta}_{(n+1)}) = \frac{p(\mathbf{x}, \hat{\theta}_{(n+1)^-}) \Psi_{n+1}(\mathbf{x})}{\int p(\mathbf{x}, \hat{\theta}_{(n+1)^-}) \Psi_{n+1}(\mathbf{x}) \, \mathrm{d}\mathbf{x}}.$$

- *Step 5*: *Resample. Sample* $\mathbf{x}_{n+1}^1, \ldots, \mathbf{x}_{n+1}^N$ *according to* $p(\mathbf{x}, \hat{\theta}_{n+1})$.
- *Step 6*: $n \leftarrow n+1$; *go to Step* (2).

Here, $(n+1)^-$ indicates the moment before the updating step. To generate $\mathbf{x}_{n+1}^1, \ldots, \mathbf{x}_{n+1}^N$, a Gibbs sampler can be used (Gelfand & Smith, 1990). This brings an extra computational cost, which should be taken into account when choosing Algorithm 2 over Algorithm 1.

It is instructive to discuss the structure of the ML estimator (Lehmann & Casella, 1998), since we are going to use this structure for the proof of convergence. Let $\hat{\mathbf{x}}_n^1, \ldots, \hat{\mathbf{x}}_n^N$ be the values of the particles right before the measurement at time n. The MLE of θ_n, $\hat{\theta}_n$, satisfies the first-order necessary condition,

$$\sum_{i=1}^{N} c_j(\hat{\mathbf{x}}_n^i) - N \frac{\int_{\mathbf{x}} c_j(\mathbf{x}) \exp(\hat{\theta}_n^{\mathrm{T}} \mathbf{c}(\mathbf{x})) \, \mathrm{d}\mathbf{x}}{\int_{\mathbf{x}} \exp(\hat{\theta}_n^{\mathrm{T}} \mathbf{c}(\mathbf{x})) \, \mathrm{d}\mathbf{x}} = 0,$$

so

$$1/N \sum_{i=1}^{N} c_j(\hat{\mathbf{x}}_n^i) = E_{\hat{\theta}_n}(c_j(\mathbf{x})), \quad \text{for } j = 1, \ldots, p, \tag{7}$$

where we use notation E_θ to specify the expectation with respect to the exponential density $p(\cdot, \theta)$. Eq. (7) says that the sample average of $c_j(\mathbf{x})$ and its probabilistic average, evaluated at $\hat{\theta}_n$, should be equal. Therefore, the MLE of θ is the solution to the system of equations in (7). For $j = 1, \ldots, p$, let

$$F_j(\theta) = \frac{1}{N} \sum_{i=1}^{N} c_j(\hat{\mathbf{x}}_n^i) - \frac{\int c_j(\mathbf{x}) \exp(\theta^{\mathrm{T}} \mathbf{c}(\mathbf{x})) \, \mathrm{d}\mathbf{x}}{\int \exp(\theta^{\mathrm{T}} \mathbf{c}(\mathbf{x})) \, \mathrm{d}\mathbf{x}}.$$

For simplicity we drop the index n from θ_n. Then,

$$-\frac{\partial F_i}{\partial \theta_j} = E_\theta(c_i(\mathbf{x}) c_j(\mathbf{x})) - E_\theta(c_i(\mathbf{x})) E_\theta(c_j(\mathbf{x})).$$

This shows that $g(\theta) = [g_{i,j}](\theta) = [(-\frac{\partial F_i}{\partial \theta_j})]$, where $g(\theta)$ is the Fisher information matrix of the exponential density at θ. Since $c_i(\mathbf{x})$, $i = 1, \ldots, p$ are affinely independent $g(\theta)$ is positive definite $\forall \theta \in \Theta$. Therefore (7) is the necessary and sufficient condition for optimality.

In this section, we prove that if the conditional density stays in exponential family of densities $\mathscr{S}$, then by using Algorithm 2 the expectation of the error of the estimate, $E\|\theta - \hat{\theta}\|$, is bounded. We show this bound can be made arbitrarily small by decreasing the step size in approximation of the Stochastic Differential Equation (SDE) in (3) and increasing the number of particles. We use the result in Milstein (1995) to approximate the SDE which is used in step 2 of Algorithm 2. In Lemma 4 of this section we prove the projection error in step 3 of Algorithm 2 can be made arbitrarily small by increasing the number of particles and decreasing the step size in approximating the SDE. In Theorem 5 we extend the result of Lemma 4 and we show the error of estimate can be made arbitrarily small $\forall t \in [0, T]$.

We recall the following fact (Lehmann & Casella, 1998):

Fact 1. *For the family of densities* $\mathscr{S}$, *the Fisher information matrix* $g(\theta) = [E(c_i(\mathbf{x})c_j(\mathbf{x})) - E(c_i(\mathbf{x}))E(c_j(\mathbf{x}))]$ *is positive definite. Also the likelihood function,* $l(\theta) = \theta^{\mathrm{T}}\mathbf{C}(\mathbf{x}) - \Upsilon(\theta)$, *is strictly concave. Therefore, for* $c_j(\mathbf{x}) = E_\theta[c_j(\mathbf{x})]$, $j = 1, \ldots, p$, *if a solution exists,*[2] *it is unique. In addition if* $\mathbf{x}_1, \ldots, \mathbf{x}_N$ *are N i.i.d. random variables distributed according to* $p(\mathbf{x}, \theta)$, *then the MLE of* θ, $\hat{\theta}_N$, *is asymptotically normal, i.e.* $\sqrt{N}(\hat{\theta}_N - \theta) \sim \mathscr{N}(0, g^{-1}(\theta))$ *where* $\hat{\theta}_N = \arg\max_\theta \prod_{i=1}^N p(\mathbf{x}_i, \theta)$.

Therefore, $E(\|\hat{\theta}_N - \theta\|^2) \sim (1/N)\mathrm{trace}(g^{-1}(\theta))$, so $\hat{\theta}_N$ converges to θ in the mean square sense as $N \to \infty$. On the other hand, $\hat{\theta}_N$ is the solution to (7). Using the strong law of large numbers (Billingsley, 1995), when $N \to \infty$ the left-hand side in (7) goes to $E_\theta(c_j(\mathbf{x}))$, $j = 1, \ldots, p$, with probability one. In other words, when the left-hand side is the exact $E_\theta(c_j(\mathbf{x}))$, $j = 1, \ldots, p$, solution to (7) gives the exact solution for θ. Using this argument, we expect that a good estimate of the left-hand side of (7) will result in a good estimate of θ. In each iteration of this algorithm the estimate of the left-hand side of (7) is found by using the Monte Carlo method and the approximate solution for the SDE in (3).

We rewrite the SDE in (3) to employ the approximation method used in Milstein (1995).

$$\mathrm{d}\mathbf{x}_t = \mathbf{f}_t(\mathbf{x}_t) \, \mathrm{d}t + \sum_{r=1}^{q} \mathbf{g}_t^r(\mathbf{x}_t) \, \mathrm{d}w_t^r, \tag{8}$$

where $\mathbf{g}_t^r(\cdot)$ is the rth column of the matrix $G_t(\cdot)$, and w_t^r is the rth component of $\mathbf{w}_t$. We introduce the

[2] In Crain (1976) it is shown that if $N > p$, the solution to (7) exists almost surely.

operators $\Lambda_r u = (\mathbf{g}^r, \partial/\partial\mathbf{x})u$, $Lu = (\partial/\partial t + (\mathbf{f}, \partial/\partial\mathbf{x}) + \frac{1}{2}\sum_{r=1}^{q}\sum_{i=1}^{n}\sum_{j=1}^{n}(g_i^r g_j^r)\partial^2/\partial\mathbf{x}_i\partial\mathbf{x}_j)u$, where $(\mathbf{a}, \partial/\partial\mathbf{x}) = \sum_{i=1}^{n} a_i \partial/\partial\mathbf{x}_i$. Then

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \sum_{r=1}^{q} \mathbf{g}_{t_k}^r \xi_k^r h^{1/2} + \mathbf{f}_{t_k} h + (L\mathbf{f})_{t_k}\frac{h^2}{2} + \sum_{r=1}^{q}\sum_{i=1}^{q}(\Lambda_r \mathbf{g}^r)_{t_k}\xi_k^{ir} h + \frac{1}{2}\sum_{r=1}^{q}(L\mathbf{g}^r + \Lambda_r \mathbf{f})_{t_k}\xi_k^r h^{3/2}, \tag{9}$$

is an approximate solution for (8), where h is the step size and the coefficients $\mathbf{g}_{t_k}^r$, $\mathbf{f}_{t_k}$, $(\Lambda_i \mathbf{g}^r)_{t_k}$, are computed at the point $(t_k, \mathbf{x}_k)$. The sets of random variables ξ_k^r, ξ_k^{ir} are independent for distinct k and are modelled by

$$\xi^{ij} = \frac{1}{2}\xi^i\xi^j - \frac{1}{2}\gamma_{ij}\zeta^i\zeta^j, \quad \gamma_{ij} = \begin{cases} -1, & i < j \\ 1, & i \geqslant j \end{cases}$$

for each k, where ξ^i and ζ^j are independent random variables satisfying $E\xi^i = E{\xi^i}^3 = E{\xi^i}^5 = 0$, $E{\xi^i}^2 = 1$, $E{\xi^i}^4 = 3$, $E\zeta^j = E{\zeta^j}^3 = 0$, $E{\zeta^j}^2 = {\zeta^j}^4 = 1$.

Definition 2. A function $u(\cdot)$ belongs to the class $\mathscr{F}$, written $u \in \mathscr{F}$, if we can find constants, $k > 0$, and $\kappa > 0$, such that for all $\mathbf{x} \in \Re^n$, $\|u(\mathbf{x})\| \leqslant k(1 + \|\mathbf{x}\|^\kappa)$.

Before presenting the convergence results, we denote the probability spaces in which the random variables are defined. The dynamics and the observation equation are defined on a fixed probability space (Ω, F, P), and E is its associated expectation. In Algorithm 2, the generated particles form a Markov process whose probability space we denote $(\Omega', F', P'_{[Y]})$ (the subindex Y emphasizes that this Markov process is driven by the observation), and $E'_{[Y]}$ is its associated expectation. A set of random variables, ξ^i, ζ^k, are defined for the numerical approximation of the SDE (8). We denote the probability space associated to these by (Ω'', F'', P''), and E'' is its associated expectation. Finally, we define $(\tilde{\Omega}, \tilde{F}, \tilde{P})$ to be the probability space associated to the joint distribution of the random variables defined in the three probability spaces, with expectation $\tilde{E}$.

The following theorem summarizes the weak approximation results for (9).

Theorem 3 (*Milstein, 1995*). *Suppose* (A1) (*Section* 2), *and suppose that the functions* $\mathbf{f}(\cdot)$, $\mathbf{g}^r(\cdot)$, $r = 1, \ldots, q$ *along with partial derivatives of sufficiently high order, belong to class* $\mathscr{F}$. *Also, suppose that the functions* $\Lambda_i \mathbf{g}^r$, $L\mathbf{g}^r$, $\Lambda_r \mathbf{f}$, *and* $L\mathbf{f}$ *grow at most as a linear function in* $\|\mathbf{x}\|$. *Then, if function* $u(\cdot)$ *and all its derivatives up to order* 6 *belong to class* $\mathscr{F}$, *approximation* (9) *has order of accuracy* 2, *in the sense of weak approximation, i.e.,*

$$\|\tilde{E}u(\mathbf{x}_{0,\mathbf{x}_0}(t_k)) - \tilde{E}u(\hat{\mathbf{x}}_{0,\mathbf{x}_0}(t_k))\| \leqslant Kh^2, \quad t_k \in [0, T],$$

where K *is a constant* (*which depends on* T) *and* $\mathbf{x}_{0,\mathbf{x}_0}(\cdot)$ *and* $\hat{\mathbf{x}}_{0,\mathbf{x}_0}(\cdot)$ *are the exact and approximate solutions for the SDE, respectively.*

The Monte Carlo approximation of $\tilde{E}u(\mathbf{x}_{0,\mathbf{x}_0}(t_k))$ brings in another error term. The combination of these errors can be expressed as follows:

$$\left\|\tilde{E}u(\mathbf{x}_{0,\mathbf{x}_0}(t_k)) - \frac{1}{N}\sum_{i=1}^{N} u(\hat{\mathbf{x}}_{0,\mathbf{x}_0^i}(t_k))\right\| \leqslant \|\tilde{E}u(\mathbf{x}_{0,\mathbf{x}_0}(t_k)) - \tilde{E}u(\hat{\mathbf{x}}_{0,\mathbf{x}_0}(t_k))\| + \left\|\tilde{E}u(\hat{\mathbf{x}}_{0,\mathbf{x}_0}(t_k)) - 1/N\sum_{i=1}^{N} u(\hat{\mathbf{x}}_{0,\mathbf{x}_0^i}(t_k))\right\|.$$

If the variance of $u(\hat{\mathbf{x}}_{0,\mathbf{x}_0}(t_k))$ is bounded, we have

$$\tilde{E}\left\|\tilde{E}u(\mathbf{x}_{0,\mathbf{x}_0}(t_k)) - \frac{1}{N}\sum_{i=1}^{N} u(\hat{\mathbf{x}}_{0,\mathbf{x}_0^i}(t_k))\right\| \leqslant Kh^2 + \frac{k'}{N^{1/2}}, \tag{10}$$

where K and k' are constants, and h is the step size for the numerical approximation of the solution of the SDE.

The next lemma relates the approximate solution to the SDE and the estimate of the parameter θ. This lemma is the main building block for our result in this section.

Lemma 4. *For the SDE*

$$\mathrm{d}\mathbf{x}_t = \mathbf{f}_t(\mathbf{x}_t)\,\mathrm{d}t + G_t(\mathbf{x}_t)\,\mathrm{d}\mathbf{w}_t, \quad \mathbf{x}_0, \quad t \in [0, t_f],$$

assume that the probability density of the state $\mathbf{x}_t$ *lies in the family* $\mathscr{S}$ *for* Θ *bounded, with* $g(\theta) \geqslant \vartheta I$ *for some* $\vartheta > 0$. *We also assume the conditions in* `Fact 1` *and in* `Theorem 3` *with* $\mathbf{c}(\mathbf{x})$ *replacing* $u(\mathbf{x})$. *Then, there exist* k_1 *and* k_2 *such that*

$$\tilde{E}[\|\theta_t - \hat{\theta}_t\|] \leqslant k_1 h^2 + \frac{k_2}{N^{1/2}}, \quad t \in [0, t_f], \tag{11}$$

where $\hat{\theta}_t$ *is the estimate of* θ_t, *and* N *and* h *are the number of particles and the time step, respectively.*

Proof. Let θ_0 be the initial condition for θ. At $t = 0$, N independent initial conditions are generated based on the density $p(\mathbf{x}, \theta_0)$, and the approximation method (9) is applied. From (10) we know that

$$\tilde{E}\left\|E_{\theta_t}\mathbf{c}(\mathbf{x}_t) - \frac{1}{N}\sum_{i=1}^{N}\mathbf{c}(\hat{\mathbf{x}}_t^i)\right\| \leqslant Kh^2 + \frac{k'}{N^{1/2}}.$$

On the other hand, from (7), we know that $\hat{\theta}$ is a solution to the system of equations $(1/N)\sum_{i=1}^{N} c_j(\hat{\mathbf{x}}_t^i) =$

$E_{\hat{\theta}_t}(c_j(\mathbf{x}_t))$, for $j=1,\ldots,p$. From Fact 1, the solution is exact if we replace $1/N\sum_{i=1}^{N}c_j(\hat{\mathbf{x}}_t^i)$ by $E_{\theta_t}(c_j(\mathbf{x}_t))$. Subtracting $E_{\theta_t}(c_j(\mathbf{x}))$ from both sides of the preceding equation and using the vector form for it, we get

$$1/N\sum_{i=1}^{N}\mathbf{c}(\hat{\mathbf{x}}_t^i)-E_{\theta_t}(\mathbf{c}(\mathbf{x}_t))=E_{\hat{\theta}_t}(\mathbf{c}(\mathbf{x}_t))-E_{\theta_t}(\mathbf{c}(\mathbf{x}_t)).$$

On the other hand, we know that $E_\theta(\mathbf{c}(\mathbf{x}))$ is a differentiable and one to one function of θ (see Fact 1). The derivative of this function, $g(\theta)$, is positive definite and by assumption $g(\theta)\geqslant\vartheta I$. Furthermore, since Θ is bounded, $\exists\alpha>0$ such that

$$\begin{aligned}\|\theta_t-\hat{\theta}_t\|&\leqslant\alpha\|E_{\theta_t}(\mathbf{c}(\mathbf{x}_t))-E_{\hat{\theta}_t}(\mathbf{c}(\mathbf{x}_t))\|\\&=\alpha\left\|E_{\theta_t}(\mathbf{c}(\mathbf{x}_t))-1/N\sum_{i=1}^{N}\mathbf{c}(\hat{\mathbf{x}}_t^i)\right\|.\end{aligned}$$

Here, taking the expectation on both sides,

$$\begin{aligned}\tilde{E}\|\theta_t-\hat{\theta}_t\|&\leqslant\alpha\tilde{E}\left\|\frac{1}{N}\sum_{i=1}^{N}\mathbf{c}(\hat{\mathbf{x}}_t^i)-E_{\theta_t}(\mathbf{c}(\mathbf{x}_t))\right\|\\&\leqslant\alpha\left(Kh^2+\frac{k'}{N^{1/2}}\right)\\&=k_1h^2+\frac{k_2}{N^{1/2}}.\qquad\square\end{aligned}$$

Now we can present the main result of this section.

Theorem 5. *For System* (3) *assume that* $\forall t\in[0,T]$ *the conditional probability density of the state* $\mathbf{x}_t$, *conditioned on the observations* $\mathbf{y}^t$, *lies in the family* $\mathscr{S}$ *for* Θ *bounded with* $g(\theta)\geqslant\vartheta I$ *for some* $\vartheta>0$. *Also, assume the conditions in* Fact 1 *and in* Theorem 3 *with* $\mathbf{c}(\mathbf{x})$ *replacing* $u(\mathbf{x})$. *Then, if* $g^{-1}(\theta_t)E_{\theta_t}(\mathscr{L}_t\mathbf{c}(\mathbf{x}))$ *is Lipschitz with the Lipschitz constant L for the operator* $\mathscr{L}_t$ *defined by*

$$\mathscr{L}_t=\sum_{i=1}^{n}f_t^i\frac{\partial}{\partial x_i}+\frac{1}{2}\sum_{i,j=1}^{n}a_t^{ij}\frac{\partial^2}{\partial x_i\partial x_j},$$

$\exists l_1,l_2$ *such that*

$$\tilde{E}\|\theta_n-\hat{\theta}_n\|\leqslant\sum_{i=0}^{n-1}\exp(Li\tau)\left(l_1h^2+\frac{l_2}{N^{1/2}}\right),\quad n\tau\in[0,T],$$

where $\hat{\theta}_n$ *is the estimate of* θ_n, *and N and h are the number of particles and the time step, respectively. This inequality implies convergence of the parameter estimate,* $\hat{\theta}_n$, *to the true parameter,* θ_n, *as* $h\longrightarrow 0$ *and* $N\longrightarrow\infty$.

Proof. Let θ_t and $\hat{\theta}_t$ be the actual and the estimated values of the parameter of the density at time $t=n\tau$, respectively. At time $t'=(n+1)\tau$ the error in the estimate of $\theta_{t'}$ is a combination of the error in the estimate in $\hat{\theta}_t$ and the error added in the time interval $[t,t']$. If the conditional density stays in the exponential family, θ_t must satisfy the following differential equation:

$$\frac{\mathrm{d}\theta}{\mathrm{d}s}=g^{-1}(\theta)E_{\theta_s}(\mathscr{L}_s\mathbf{c}(\mathbf{x})),$$

for $t=n\tau\leqslant s<t'=(n+1)\tau$, and $\theta(t)=\hat{\theta}_t$ (see Theorem 4.5.1 in Brigo (1996)). Let $\tilde{\theta}_{t'}$ be the solution of the differential equation evaluated at $s=t'$. By the assumptions of this theorem, $g^{-1}(\theta)E_{\theta_s}(\mathscr{L}_s\mathbf{c}(\mathbf{x}))$ is Lipschitz with constant L. Hence by continuity of the solution of the ODE with respect to the initial condition (Khalil, 1996),

$$\|\theta_{t'}-\tilde{\theta}_{t'}\|\leqslant\|\theta_t-\hat{\theta}_t\|\mathrm{e}^{L(t'-t)}.$$

Therefore, $\tilde{E}\|\theta_{t'}-\tilde{\theta}_{t'}\|\leqslant\tilde{E}\|\theta_t-\hat{\theta}_t\|\mathrm{e}^{L(t'-t)}$. Also from Lemma 4, $\exists k_1(t')$ and $k_2(t')$ such that

$$\tilde{E}[\|\tilde{\theta}_{t'}-\hat{\theta}_{t'}\|]\leqslant k_1(t')h^2+k_2(t')/N^{1/2}.$$

Therefore using the triangle inequality we have

$$\begin{aligned}\tilde{E}\|\theta_{t'}-\hat{\theta}_{t'}\|&\leqslant\tilde{E}\|\theta_{t'}-\tilde{\theta}_{t'}\|+\tilde{E}\|\tilde{\theta}_{t'}-\hat{\theta}_{t'}\|\\&\leqslant\tilde{E}\|\theta_t-\hat{\theta}_t\|\mathrm{e}^{L(t'-t)}+k_1(t')h^2\\&\quad+k_2(t')/N^{1/2}.\end{aligned}$$

The observation noise $\mathbf{v}_n$ and the function $\mathbf{h}(\cdot)$ are such that Bayes' Rule does not introduce any further error in the estimate of $\hat{\theta}_{t'}$. More precisely, $\Psi_n(\mathbf{x})$ is assumed to be a member of $\mathscr{S}$. Therefore, after applying Bayes' Rule to $p(\mathbf{x},\theta_{t'})$ and $p(\mathbf{x},\hat{\theta}_{t'})$ parameters $\theta_{t'}$ and $\hat{\theta}_{t'}$ are shifted by the same vector and therefore, $\|\theta_{t^{+'}}-\hat{\theta}_{t^{+'}}\|=\|\theta_{t'}-\hat{\theta}_{t'}\|$ (cf. Section 4.5.2 of Brigo (1996). Here $t^{+'}$ represents the time right after the Bayes' correction. Therefore, starting from the initial condition θ_0 we get

$$\tilde{E}\|\theta_n-\hat{\theta}_n\|\leqslant\sum_{i=0}^{n-1}\exp(Li\tau)\left(l_1h^2+\frac{l_2}{N^{1/2}}\right),\quad n\tau\in[0,T]$$

where $l_i=\max_n k_i(n\tau)$, $n\tau\in[0,T]$, $i=1,2$. $\square$

Here, we make a few remarks:

- The result of Theorem 5 can be easily extended to convergence in the mean square sense.
- If $\|u(\mathbf{x})\|\leqslant k(1+\|\mathbf{x}\|^\kappa)$ for $k>0$ and $\kappa>0$, then $\lim_{\substack{N\to\infty\\h\to 0}}\tilde{E}\|E_\theta u(\mathbf{x})-E_{\theta^*}u(\mathbf{x})\|=0$.
- In Brigo (1996) the observation equation is considered to be time invariant. Here, the time-varying nature of $\mathbf{h}_n(\mathbf{x})$ does not complicate the algorithm.

5. Projection particle filtering for exponential families of densities

In this section, we drop the assumption that the conditional density of the state given the observation (6) lies in the exponential family of densities, $\mathscr{S}$. Here, we project the density into the exponential family at each measurement epoch. Also, we do not require that $\Psi_n(\mathbf{x})$ is a member of $\mathscr{S}$. Instead we use other assumptions. First we need the following definition:

Definition 6. We say that a function $u(\cdot)$ belongs to the class $\mathscr{F}_{k\kappa}$, written as $u \in \mathscr{F}_{k\kappa}$, for fixed $k > 0$ and κ, such that for all $\mathbf{x} \in \Re^n$, $\|u(\mathbf{x})\| \leqslant k(1 + \|\mathbf{x}\|^\kappa)$.

The next two **assumptions** guarantee the existence of an exponential density close to the true conditional density.

A3. The density in (6) stays close to the given exponential family $\mathscr{S}$ in a weak sense. In other words $\forall t \in [0, T]$, $\forall u \in \mathscr{F}_{k\kappa} \exists \theta_t^* \in \Theta^*$ such that

$$\tilde{E}\|E_{p_t}(u(\mathbf{x})) - E_{\theta_t^*}(u(\mathbf{x}))\| \leqslant \varepsilon, \tag{12}$$

where $t - 1 < n\tau \leqslant t$ and Θ^* is convex and compact.

A4. For $\theta_{n^-}^*$ in (A3) and $\Psi_n(\mathbf{x})$, $\exists \Psi_n^*(\mathbf{x})$ such that $p(\mathbf{x}, \theta) = p(\mathbf{x}, \theta_{n^-}^*)\Psi_n^*(\mathbf{x}) / \int p(\mathbf{x}, \theta_{n^-}^*)\Psi_n^*(\mathbf{x})\,\mathrm{d}\mathbf{x}$ is in the family $\mathscr{S}$ for some $\theta \in \Theta^*$ and we have

- $\forall \theta \in \Theta^*$ and $\forall u(\cdot) \in \mathscr{F}_{k\kappa}$, $\exists \varepsilon > 0$ such that

$$\tilde{E}\left\| \frac{E_\theta \Psi_n(\mathbf{x})u(\mathbf{x})}{E_\theta \Psi_n(\mathbf{x})} - \frac{E_\theta \Psi_n^*(\mathbf{x})u(\mathbf{x})}{E_\theta \Psi_n^*(\mathbf{x})} \right\| \leqslant \varepsilon.$$

- $\forall u(\cdot) \in \mathscr{F}_{k\kappa}$, $\exists \varepsilon > 0$ such that

$$\tilde{E}\left\| \frac{E_{\theta_{n^-}^*} \Psi_n^*(\mathbf{x})u(\mathbf{x})}{E_{\theta_{n^-}^*} \Psi_n^*(\mathbf{x})} - \frac{E_{p_{n^-}} \Psi_n(\mathbf{x})u(\mathbf{x})}{E_{p_{n^-}} \Psi_n(\mathbf{x})} \right\| \leqslant \varepsilon.$$

It is clear that if $\Psi_n^*(\cdot)$ satisfies (4), so does $c\Psi_n^*(\cdot)$ for any $c > 0$. Moreover, $\Psi_n^*(\cdot) = \exp(\overline{\alpha}^{\mathrm{T}}\mathbf{c}(\cdot))$ for some $\overline{\alpha} \in \Re^p$.

Using Assumptions 3 and 4 we prove the following lemma which is one of the building blocks of this section's results.

Lemma 7. *For $\theta_{n^-}^*$ and $\Psi_n^*(\mathbf{x})$ defined in* (A4) *and $\forall u(\cdot) \in \mathscr{F}_{k\kappa}$ and $\forall \theta_1, \theta_2 \in \Theta^*$, $\exists$ positive numbers k_1, k_2, k_3, k_4 independent of $\theta_{n^-}^*$, $\Psi_n^*(\mathbf{x})$, θ_1, and θ_2 such that*

(a) $k_1 \leqslant \|E_\theta \Psi_n^*(\mathbf{x})\| \leqslant k_2 \quad \forall \theta \in \Theta^*$.
(b) $\|E_\theta \Psi_n^*(\mathbf{x})u(\mathbf{x})\| \leqslant k_3 \quad \forall \theta \in \Theta^*$.
(c) $\|E_{\theta_1} \Psi_n^*(\mathbf{x})u(\mathbf{x}) - E_{\theta_2} \Psi_n^*(\mathbf{x})u(\mathbf{x})\| \leqslant k_4\|\theta_1 - \theta_2\|$.

Proof. See the Appendix.

Now, we prove the theorem that we will state precisely later. Assume $\hat{\theta}_n$ is calculated according to Algorithm 2 and assume $p(\mathbf{x}, \hat{\theta}_n)$ is such that $\forall u \in \mathscr{F}_{k\kappa}$

$$\tilde{E}\|E_{\hat{\theta}_n} u(\mathbf{x}) - E_{\theta_n^*} u(\mathbf{x})\| \leqslant \delta, \tag{13}$$

where θ_n^* (see (A3)) satisfies

$$\tilde{E}\|E_{p_n} u(\mathbf{x}) - E_{\theta_n^*} u(\mathbf{x})\| \leqslant \varepsilon. \tag{14}$$

Using the density $p(\mathbf{x}, \hat{\theta}_n)$, new particles $\mathbf{x}_n^1, \ldots, \mathbf{x}_n^N$ are generated. The approximate solution for the SDE at time $(n+1)\tau$ maps these particles to $\hat{\mathbf{x}}_{n+1}^1, \ldots, \hat{\mathbf{x}}_{n+1}^N$. From these new particles $\hat{\theta}_{n+1}$ is calculated. From (13) and (14) we have

$$\tilde{E}\|E_{p_n} u(\mathbf{x}) - E_{\hat{\theta}_n} u(\mathbf{x})\| \leqslant \delta + \varepsilon. \tag{15}$$

We define the function $\mathbf{r}(\mathbf{x})$ to be $\mathbf{r}(\mathbf{x}) = E''\mathbf{c}(\hat{\mathbf{x}}_{n,\mathbf{x}}((n+1)))$, where $\hat{\mathbf{x}}_{n,\mathbf{x}}((n+1)\tau)$ is the approximate solution of SDE (8) at time $(n+1)\tau$ with the given initial condition $\mathbf{x}$ at time $n\tau$ using the method in (9). Since according to our assumption $\mathbf{c} \in \mathscr{F}_{k\kappa}$, by using Lemma 9.1 in Milstein (1995), we have

$$\|\mathbf{r}(\mathbf{x})\| \leqslant K_3(1 + \|\mathbf{x}\|^\mu),$$

where K_3 and μ only depend on the function $\mathbf{c}(\cdot)$ and the dimension of $\mathbf{x}$. Therefore, k and κ in Definition 6 can be chosen such that $\mathbf{r} \in \mathscr{F}_{k\kappa}$. If the argument of $\mathbf{r}(\cdot)$ is a random variable, then using (15) we have

$$\tilde{E}\|E_{p_n} \mathbf{r}(\mathbf{x}) - E_{\hat{\theta}_n} \mathbf{r}(\mathbf{x})\| \leqslant \delta + \varepsilon. \tag{16}$$

More explicitly,

$$\begin{aligned}\tilde{E}\|&E_{p_n} E''[\mathbf{c}(\hat{\mathbf{x}}_{n,\mathbf{x}}((n+1)\tau))] \\ &- E_{\hat{\theta}_n} E''[\mathbf{c}(\hat{\mathbf{x}}_{n,\mathbf{x}}((n+1)\tau))]\| \leqslant \delta + \varepsilon.\end{aligned} \tag{17}$$

From Theorem (3) we have

$$\begin{aligned}\tilde{E}\|&E_{p_n}\mathbf{c}(\mathbf{x}_{n,\mathbf{x}}((n+1)\tau)) \\ &- E_{p_n} E''\mathbf{c}(\hat{\mathbf{x}}_{n,\mathbf{x}}((n+1)\tau))\| \leqslant K_4 h^2,\end{aligned} \tag{18}$$

for some $K_4 > 0$. Using the Monte Carlo method to calculate the $E_{p_n}\mathbf{c}(\hat{\mathbf{x}}_{n,\mathbf{x}}((n+1)\tau))$ brings another error term that is due to the finite number of particles as the initial conditions for method (9). The expectation of this error is bounded, i.e. $\exists K_5 > 0$ s.t.

$$\begin{aligned}\tilde{E}\Bigg\|&E_{\hat{\theta}_n} E''\mathbf{c}(\hat{\mathbf{x}}_{n,\mathbf{x}}((n+1)\tau)) \\ &-\frac{1}{N}\sum_{i=1}^{N}\mathbf{c}(\hat{\mathbf{x}}_{n,\hat{\mathbf{x}}_i^i}((n+1)\tau))\Bigg\| \leqslant K_5/N^{1/2},\end{aligned} \tag{19}$$

where $\hat{\mathbf{x}}^i$ are distributed according to $p(\mathbf{x}, \hat{\theta}_n)$. Combining (17), (18), and (19) we get

$$\begin{aligned}\tilde{E}\Bigg\|&E_{p_n}\mathbf{c}(\mathbf{x}_{n,\mathbf{x}}((n+1)\tau)) \\ &-\frac{1}{N}\sum_{i=1}^{N}\mathbf{c}(\hat{\mathbf{x}}_{n,\hat{\mathbf{x}}^i}((n+1)\tau))\Bigg\| \leqslant \delta + \varepsilon + K_4 h^2 + K_5/N^{1/2}.\end{aligned} \tag{20}$$

Based on (A3), we know that $\exists\theta^*_{(n+1)^-}$ such that

$$\tilde{E}\|E_{p_{(n+1)^-}}\mathbf{c}(\mathbf{x}) - E_{\theta^*_{(n+1)^-}}\mathbf{c}(\mathbf{x})\| \leqslant \varepsilon. \tag{21}$$

We know that, if $\mathbf{x}$ (initial condition at time $n\tau$) is distributed according to $p_n(\mathbf{x})$, then

$$E_{p_{(n+1)^-}}\mathbf{c}(\mathbf{x}) = E_{p_n}\mathbf{c}(\mathbf{x}_{n,\mathbf{x}}((n+1)\tau)). \tag{22}$$

Therefore, from (20) and (21) we get

$$\tilde{E}\left\|E_{\theta^*_{(n+1)^-}}\mathbf{c}(\mathbf{x}) - \frac{1}{N}\sum_{i=1}^{N}\mathbf{c}(\hat{\mathbf{x}}_{n,\hat{\mathbf{x}}^i}((n+1)\tau))\right\| \leqslant \delta + 2\varepsilon + K_4h^2 + K_5/N^{1/2}. \tag{23}$$

Then $\hat{\theta}_{(n+1)^-}$ given by Algorithm 2 satisfies the following inequality

$$\tilde{E}\|E_{\theta^*_{(n+1)^-}}\mathbf{c}(\mathbf{x}) - E_{\hat{\theta}_{(n+1)^-}}\mathbf{c}(\mathbf{x})\| \leqslant \delta + 2\varepsilon + K_4h^2 + K_5/N^{1/2}. \tag{24}$$

From (A4) we know that $\exists\theta \in \Theta^*$ such that

$$\tilde{E}\left\|\frac{E_{\theta^*_{(n+1)^-}}\Psi^*_{n+1}(\mathbf{x})u(\mathbf{x})}{E_{\theta^*_{(n+1)^-}}\Psi^*_{n+1}(\mathbf{x})} - \frac{E_{p_{(n+1)^-}}\Psi_{n+1}(\mathbf{x})u(\mathbf{x})}{E_{p_{(n+1)^-}}\Psi_{n+1}(\mathbf{x})}\right\| = \tilde{E}\|E_\theta u(\mathbf{x}) - E_{p_{(n+1)}}u(\mathbf{x})\| \leqslant \varepsilon.$$

Since θ satisfies the inequality in (A3), we can choose $\theta^*_{(n+1)}$ to be θ, i.e. $\theta^*_{(n+1)} = \theta$.

On the other hand we have

$$\begin{aligned}
&\|E_{\theta^*_{(n+1)}}u(\mathbf{x}) - E_{\hat{\theta}_{(n+1)}}u(\mathbf{x})\| \\
&\quad= \left\|\frac{E_{\theta^*_{(n+1)^-}}\Psi^*_{n+1}(\mathbf{x})u(\mathbf{x})}{E_{\theta^*_{(n+1)^-}}\Psi^*_{n+1}(\mathbf{x})} - \frac{E_{\hat{\theta}_{(n+1)^-}}\Psi_{n+1}(\mathbf{x})u(\mathbf{x})}{E_{\hat{\theta}_{(n+1)^-}}\Psi_{n+1}(\mathbf{x})}\right\| \\
&\quad\leqslant \left\|\frac{E_{\theta^*_{(n+1)^-}}\Psi^*_{n+1}(\mathbf{x})u(\mathbf{x})}{E_{\theta^*_{(n+1)^-}}\Psi^*_{n+1}(\mathbf{x})} - \frac{E_{\theta^*_{(n+1)^-}}\Psi^*_{n+1}(\mathbf{x})u(\mathbf{x})}{E_{\hat{\theta}_{(n+1)^-}}\Psi^*_{n+1}(\mathbf{x})}\right\| \\
&\qquad+ \left\|\frac{E_{\theta^*_{(n+1)^-}}\Psi^*_{n+1}(\mathbf{x})u(\mathbf{x})}{E_{\hat{\theta}_{(n+1)^-}}\Psi^*_{n+1}(\mathbf{x})} - \frac{E_{\hat{\theta}_{(n+1)^-}}\Psi^*_{n+1}(\mathbf{x})u(\mathbf{x})}{E_{\hat{\theta}_{(n+1)^-}}\Psi^*_{n+1}(\mathbf{x})}\right\| \\
&\qquad+ \left\|\frac{E_{\hat{\theta}_{(n+1)^-}}\Psi^*_{n+1}(\mathbf{x})u(\mathbf{x})}{E_{\hat{\theta}_{(n+1)^-}}\Psi^*_{n+1}(\mathbf{x})} - \frac{E_{\hat{\theta}_{(n+1)^-}}\Psi_{n+1}(\mathbf{x})u(\mathbf{x})}{E_{\hat{\theta}_{(n+1)^-}}\Psi_{n+1}(\mathbf{x})}\right\| \\
&\quad\leqslant \frac{\|E_{\theta^*_{(n+1)^-}}\Psi^*_{n+1}(\mathbf{x})u(\mathbf{x})\|}{\|E_{\theta^*_{(n+1)^-}}\Psi^*_{n+1}(\mathbf{x})\|\|E_{\hat{\theta}_{(n+1)^-}}\Psi^*_{n+1}(\mathbf{x})\|} \times \|E_{\theta^*_{(n+1)^-}}\Psi^*_{n+1}(\mathbf{x}) - E_{\hat{\theta}_{(n+1)^-}}\Psi^*_{n+1}(\mathbf{x})\| \\
&\qquad+ \frac{1}{\|E_{\hat{\theta}_{(n+1)^-}}\Psi^*_{n+1}(\mathbf{x})\|} \times \|E_{\theta^*_{(n+1)^-}}\Psi^*_{n+1}(\mathbf{x})u(\mathbf{x}) - E_{\hat{\theta}_{(n+1)^-}}\Psi^*_{n+1}(\mathbf{x})u(\mathbf{x})\| \\
&\qquad+ \left\|\frac{E_{\hat{\theta}_{(n+1)^-}}\Psi^*_{n+1}(\mathbf{x})u(\mathbf{x})}{E_{\hat{\theta}_{(n+1)^-}}\Psi^*_{n+1}(\mathbf{x})} - \frac{E_{\hat{\theta}_{(n+1)^-}}\Psi_{n+1}(\mathbf{x})u(\mathbf{x})}{E_{\hat{\theta}_{(n+1)^-}}\Psi_{n+1}(\mathbf{x})}\right\|.
\end{aligned}$$

Using Lemma 7 and (A4) we get

$$\tilde{E}\|E_{\theta^*_{(n+1)}}u(\mathbf{x}) - E_{\hat{\theta}_{(n+1)}}u(\mathbf{x})\| \leqslant \frac{k_3k_4 + k_1k_4}{k_1^2}\tilde{E}\|\theta^*_{(n+1)^-} - \hat{\theta}_{(n+1)^-}\| + \varepsilon.$$

Therefore, from (24) and Fact 2 we get

$$\tilde{E}\|\theta^*_{(n+1)^-} - \hat{\theta}_{(n+1)^-}\| \leqslant K_2(\delta + 2\varepsilon + K_4h^2 + K_5/N^{1/2}).$$

This implies that, $\exists\iota_1, \iota_2, \iota_3, \iota_4 > 0$ such that

$$\tilde{E}\|E_{\theta^*_{(n+1)}}u(\mathbf{x}) - E_{\hat{\theta}_{(n+1)}}u(\mathbf{x})\| \leqslant \iota_1\delta + \iota_2\varepsilon + \iota_3h^2 + \iota_4N^{-1/2}.$$

The next theorem summarizes our result in this section.

Theorem 8. *For system* (3) *assume* (A1), (A2), (A3), *and* (A4). *We also assume the conditions in* `Fact 1` *and in* `Theorem 3` *with* $\mathbf{c}(\mathbf{x})$ *replacing* $u(\mathbf{x})$, *and we assume* $\mathbf{r} \in \mathscr{F}_{k\kappa}$. *Then in* `Algorithm 2` *with approximation* (9), *if* $\tilde{E}\|E_{\hat{\theta}_n}u(\mathbf{x}) - E_{\theta^*_n}u(\mathbf{x})\| \leqslant \delta$, $\forall u(\cdot) \in \mathscr{F}_{k\kappa}$ *then*

$$\tilde{E}\|E_{\theta^*_{(n+1)}}u(\mathbf{x}) - E_{\hat{\theta}_{(n+1)}}u(\mathbf{x})\| \leqslant \iota_1\delta + \iota_2\varepsilon + \iota_3h^2 + \iota_4N^{-1/2},$$

for some $\iota_1, \iota_2, \iota_3, \iota_4 > 0$.

In Theorem (8) only one step of Algorithm 2 is considered; it is straightforward to use then Theorem (8) repeatedly for the time interval $[0, T]$, where $T = M\tau$. In that case, $\|E_{\hat{\theta}_0}u(\mathbf{x}) - E_{\theta^*_0}u(\mathbf{x})\| \leqslant \delta_0$. Then $\exists\alpha_1, \alpha_2, \alpha_3$, and α_4 positive such that

$$\tilde{E}\|E_{\theta^*_n}u(\mathbf{x}((n)\tau)) - E_{\hat{\theta}_n}u(\mathbf{x}((n)\tau))\| \leqslant \alpha_1^n\delta_0 + \sum_{i=0}^{n-1}\alpha_1^i(\alpha_2\varepsilon + \alpha_3h^2 + \alpha_4N^{-1/2}),$$

for $0 \leqslant n \leqslant M$.

6. Applications of projection particle filtering to integrated INS/GPS

GPS provides world wide accurate positioning, if four or more satellites are in view of the receiver. Although the satellite constellation guarantees availability of four or more satellites world wide, natural or man-made obstacles can block the satellite signals easily. Integrating dead reckoning or INS with GPS (Abbott & Powell, 1999; Elkaim, O'Connor, Bell, & Parkinson, 1996; Zickel & Nehemia, 1994; Carvalho et al., 1997) is a method to overcome this vulnerability. Here, INS or the dead reckoning provides

positioning that is calibrated by the GPS. In this section we consider the case of an integrated INS/GPS. We show that using nonlinear filtering for positioning is essential, and we compare the proposed Projection PaF with regular PaF and EKF. One application of accurate INS/GPS navigation is in the formation flight of unmanned aerial vehicles to reduce drag (Iannotta, 2002).

We consider the observation equation provided by the ith GPS satellite at time n to have the following form:

$$y_n^i = \rho^i(p_x, p_y, p_z) - \rho^i(b_x, b_y, b_z) + c\delta + v_n^i, \tag{25}$$

where $[p_x, p_y, p_z]^{\mathrm{T}}$ and $[b_x, b_y, b_z]^{\mathrm{T}}$ are the rover and (known) base coordinates at time n, respectively, $\rho^i(x_1, x_2, x_3)$ is the distance from point $[x_1, x_2, x_3]^{\mathrm{T}}$ to satellite i, δ is a combination of the receiver clock bias in the base and the rover, c is the speed of light, and v_n^i is the measurement noise for the ith satellite signal.

Here we point out that the nonlinearity in measurement is not solely due to the function ρ. Integrated INS/GPS requires coordinate transformations between INS parameters and GPS parameters, which contributes to the nonlinearity of the measurement.

Parameters of an integrated INS/GPS are expressed in different coordinate systems. The GPS measurements are given in an Earth Centered Earth Fixed (ECEF) frame (Hofmann-Wellenhof, Lichtenegger, & Collins, 1993; Farrell & Barth, 1998). The GPS position is given either in the rectangular coordinate system (see Eq. (25)), or in the geodetic coordinate system with the familiar latitude, longitude, and height coordinate vector, $[p_\lambda, p_\phi, p_h]^{\mathrm{T}}$. For the latter, the Earth's geoid is approximated by an ellipsoid. Table 1 shows the defining parameters for the geoid according to the WGS84 reference frame. The parameters and measurements of INS are given in the local geographical frame or in the body frame system where the velocity is given by the north-east-down velocity vector, $[v_{\mathrm{N}}, v_{\mathrm{E}}, v_{\mathrm{D}}]^{\mathrm{T}}$. The transformation matrix from the body frame to the local geographical frame is given by matrix R_{b2g}. In this paper we assume the estimation problem for the gyro measurements is solved. Hence R_{b2g} is known.

The GPS clock drift and the INS equations constitute key dynamics in integrated INS/GPS. The INS dynamic equation can be expressed as follows (see Azimi-Sadjadi & Krishnaprasad (2001b) for details):

$$\mathrm{d}\begin{pmatrix} p_\lambda \\ p_\phi \\ p_h \end{pmatrix} = \begin{pmatrix} \dfrac{1}{R_\lambda + p_h} & 0 & 0 \\ 0 & \dfrac{1}{(R_\phi + p_h)\cos(p_\lambda)} & 0 \\ 0 & 0 & -1 \end{pmatrix} \times \begin{pmatrix} v_{\mathrm{N}} \\ v_{\mathrm{E}} \\ v_{\mathrm{D}} \end{pmatrix} \mathrm{d}t$$

Table 1
Definition of the parameters for WGS84 reference frame.

a	6378137.0 m	Semi major axis
b	6356752.3142 m	Semi minor axis
ω_{ie}	7.292115×10^{-5}	Earth angular velocity
e	$\dfrac{\sqrt{b(a-b)}}{a}$	Ellipsoid eccentricity

$$\mathrm{d}\begin{pmatrix} v_{\mathrm{N}} \\ v_{\mathrm{E}} \\ v_{\mathrm{D}} \end{pmatrix} = \left[\begin{pmatrix} -\frac{v_{\mathrm{E}}^2}{R_\phi + p_h}\tan(p_\lambda) - 2\omega_{ie}\sin(p_\lambda)v_{\mathrm{E}} \\ \quad + \frac{v_{\mathrm{N}}v_{\mathrm{D}}}{R_\lambda + p_h} \\ \frac{v_{\mathrm{E}}v_{\mathrm{N}}}{R_\lambda + p_h}\tan(\lambda) + \omega_{ie}\sin(p_\lambda)v_{\mathrm{N}} \\ \quad + \frac{v_{\mathrm{E}}v_{\mathrm{D}}}{R_\phi + p_h} + 2\omega_{ie}\cos(p_\lambda)v_{\mathrm{D}} \\ -\frac{v_{\mathrm{N}}^2}{R_\lambda + p_h} - \frac{v_{\mathrm{E}}^2}{R_\phi + p_h} - 2\omega_{ie}\cos(p_\lambda)v_{\mathrm{E}} \end{pmatrix} + R_{b2g}\left(\begin{pmatrix} \tilde{a}_u \\ \tilde{a}_v \\ \tilde{a}_w \end{pmatrix} + \begin{pmatrix} b_u \\ b_v \\ b_w \end{pmatrix}\right) + \begin{pmatrix} 0 \\ 0 \\ g \end{pmatrix} \right] \mathrm{d}t + \mathrm{d}\mathbf{w}_t^v,$$

where $g = 9.780327\,\mathrm{m/s^2}$ is the gravitational acceleration,

$$R_\lambda = \frac{a(1-\mathrm{e}^2)}{(1-\mathrm{e}^2\sin^2(p_\lambda))^{3/2}}, \quad R_\phi = \frac{a}{(1-\mathrm{e}^2\sin^2(p_\lambda))^{1/2}},$$

$[\tilde{a}_u, \tilde{a}_v, \tilde{a}_w]^{\mathrm{T}}$ is the accelerometer measurement and $[b_u, b_v, b_w]^{\mathrm{T}}$ is the accelerometer measurement bias both expressed in the body frame, and $\mathbf{w}^v$ is a vector-valued Brownian motion process with zero mean and known covariance matrix. The measurement bias, $\mathbf{b} = [b_u, b_v, b_w]^{\mathrm{T}}$, has the following dynamics

$$\mathrm{d}\mathbf{b} = -a_b\mathbf{b}\,\mathrm{d}t + \mathrm{d}\mathbf{w}_t^b,$$

where $\mathbf{w}_t^b$ is a vector-valued Brownian motion with zero mean and known covariance matrix, and a_b is a small positive constant. The receiver clock drift, δ_t, is represented by the integration of an exponentially correlated random process ϱ_t (Carvalho et al., 1997).

$$\mathrm{d}\varrho_t = -(1/500)\varrho_t\,\mathrm{d}t + \mathrm{d}w_t^\varrho$$
$$\mathrm{d}\delta_t = \varrho_t\,\mathrm{d}t,$$

where w_t^ϱ is a zero-mean Brownian motion process with variance $\sigma_\varrho^2 = 10^{-24}$. This dynamic model is typical for a quartz TCXO with frequency drift rate $10^{-9}\,\mathrm{s/s}$ (Carvalho et al., 1997).

6.1. Simulation and results

For our simulations of an INS/GPS we apply three different filtering methods, EKF, PaF, and Projection PaF. Therefore, with the assumptions of the previous sections, the dimension of the dynamical system for this simulation is

eleven. The state of the dynamical system is

$$\mathbf{x} = [p_\lambda, p_\phi, p_h, v_N, v_E, v_D, b_u, b_v, b_w, \varrho, \delta]^T.$$

Here, we assume that $a_b = 0.001$, and that the covariance matrices for the Brownian motions in the INS dynamic equations, Σ_b and Σ_v, are diagonal. Specifically, $\Sigma_b = 10^{-6} I$ and $\Sigma_v = 10^{-4} I$, where I is the appropriate identity matrix. The observation vector at time n is $\mathbf{y}_n = [y_n^1, \ldots, y_n^d]^T$, where d is the number of satellites in view and y_n^i is given in (25). In (25) the observation is given as a function of the position in the ECEF rectangular coordinate system that is then transformed to the geodetic coordinate system (Azimi-Sadjadi & Krishnaprasad, 2001b).

For this simulation we simply chose an 11 dimensional Gaussian density for the Projection PaF. This choice makes random vector generation easy and computationally affordable. For the Projection PaF, we employed maximum likelihood estimates of the Gaussian density parameters before and after the Bayes' correction.

We used two Novatel GPS receivers to collect navigation data (April 2, 2000). From this data, we extracted position information for the satellites, the pseudo range, and the carrier phase measurement noise powers for the L1 frequency. From the collected information we generated the pseudo range and the carrier phase data for one static and one moving receiver (base and rover, respectively). We assume that for the carrier phase measurement the integer ambiguity problem is already solved (see Azimi-Sadjadi & Krishnaprasad (2001a) for details). The movement of the INS/GPS platform was simulation based and it was the basis for the measurement data measured by the accelerometers, the gyros, the GPS pseudo range, and the GPS carrier phase data.

We first compare the position estimation error of Projection PaF with PaF and EKF. The position and observation data are the same for the three methods and the comparison is done for *one typical run* of each method. For this run we assumed that the GPS receiver starts with 6 satellites. At time $t = 100$, the receiver experiences loss of lock for signals from 3 satellites. The receiver regains one satellite signal at $t = 400$. The loss of lock occurs when the receiver's phase lock loop cannot track the received signal for a variety of reasons including natural or manmade obstacles, low signal-to-noise ratio, and circuit failure. We emphasize that GPS requires at least 4 satellites for instantaneous stand alone positioning. We used 500 particles in the Projection PaF and PaF methods. The position error of the Projection PaF, PaF, and EKF methods are plotted in Fig. 1. When the receiver loses the signal from 3 satellites (starting at $t = 100$) the error of the EKF method grows exponentially and the position error reaches almost 400 m compared to 18 m for PaF and 4 m for Projection PaF. When the receiver gains an additional satellite signal at $t = 400$ the EKF corrects its position estimate. Unlike EKF, PaF and Projection PaF are successful in providing a reasonable position estimate even with only 3 satellites in view. Fig. 2 is a magnified section of Fig. 1 emphasizing the comparison of the errors between PaF and Projection PaF, which clearly shows that the Projection PaF method outperforms the PaF method.

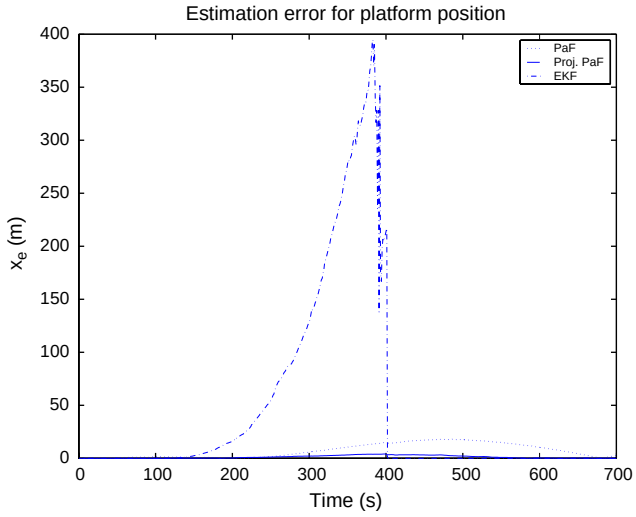


Fig. 1. Position estimation error (=Euclidean distance between cartesian position vector and its estimate) for three methods: EKF, PaF, and Projection PaF. The system starts with 6 satellites in view. At $t = 100$ s the signals from 3 satellites are lost. At $t = 400$ s the system regains the signal from one satellite.

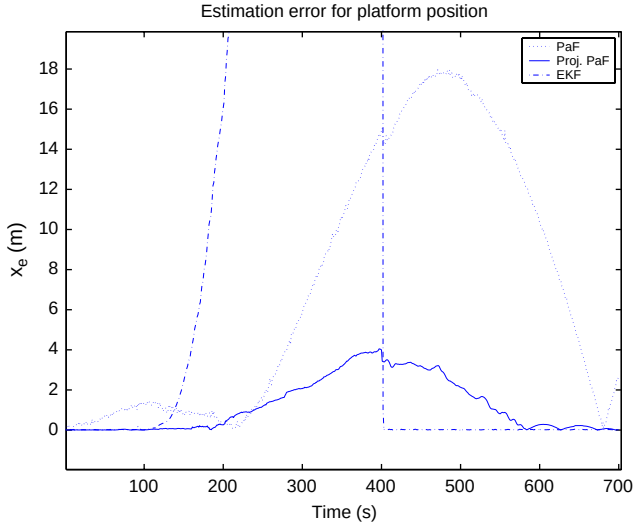


Fig. 2. Detail of Fig. 1. This compares the estimation errors of the Projection PaF and PaF methods.

Table 2
Estimation error of Projection PaF and PaF methods

Number of Particles	500	1000	2000
Projection PaF	1.2 m	1.1 m	1.0 m
PaF	17.3 m	14.3 m	14.3 m

Table 2 compares the estimation error of PaF and Projection PaF for different numbers of particles when the

integrated INS/GPS receives signals from only 3 satellites. For each method we simulated the filtering algorithm for 100 independent runs, each run 300 s long. The position error is averaged over time and the 100 runs. We see that the error of the Projection PaF is an order of magnitude smaller than that of PaF. For both methods the estimation error decreases, though with slow rate, as the number of particles grows. It should be noted that increasing the number of particles cannot indefinitely reduce the position error; the lower bound for the position error is the error of the optimum nonlinear filter. The slow rate for Projection PaF could be associated to the closeness of the error of this method to the error of the optimum nonlinear filter. For PaF the slow rate is because of sample impoverishment which results in poor approximation of the conditional density (Carpenter, Clifford, & Fearnhead, 1999).

7. Concluding remarks

Bringing together the distinct advantages of simulation and sampling-based density computation and analytic propagation in parameterized families of densities, we introduce in this paper a new approach to the problem of nonlinear filtering—the projection particle filter. We develop the filter for exponential families of densities. We provide bounds for estimation errors accounting for dependence on population size of particles, and step size used in numerical integration of dynamics. We illustrate the effectiveness of this new filter in the technologically important problem of integrating INS and GPS. We aim to investigate in future work application of projection particle filtering to stochastic control.

Appendix

Using Assumption (A3), we state this fact:

Fact 2. *$\forall\theta_1, \theta_2 \in \Theta^*$ and $\forall u \in \mathscr{F}_{k\kappa}$, $\exists K_1, K_2$ positive such that*

$$\|E_{\theta_1}u(\mathbf{x}) - E_{\theta_2}u(\mathbf{x})\| \leqslant K_1\|\theta_1 - \theta_2\| \tag{26}$$

$$\|\theta_1 - \theta_2\| \leqslant K_2\|E_{\theta_1}\mathbf{c}(\mathbf{x}) - E_{\theta_2}\mathbf{c}(\mathbf{x})\|. \tag{27}$$

Proof. To prove (26), define $f_u(\theta) = E_\theta u(\mathbf{x})$ for $u(\cdot) \in \mathscr{F}_{k\kappa}$. Then

$$\frac{\mathrm{d}}{\mathrm{d}\theta_i} f_u(\theta) = E_\theta c_i(\mathbf{x})u(\mathbf{x}) - E_\theta c_i(\mathbf{x})E_\theta u(\mathbf{x}).$$

Since $\|u(\mathbf{x})\| \leqslant k(1 + \|\mathbf{x}\|^\kappa)$ and $\theta \in \Theta^*$, where Θ^* is compact, then there exists a constant A such that $\|\mathrm{d}f_u(\theta)/\mathrm{d}\theta\| \leqslant A\ \forall u(\cdot) \in \mathscr{F}_{k\kappa}$ and $\forall\theta \in \Theta^*$. Since Θ^* is convex and compact, it is clear that $\exists K_1$ independent of $u(\cdot)$ such that $f_u(\mathbf{x})$ is Lipschitz over Θ^* with the Lipschitz constant K_1 (Khalil, 1996).

Inequality (27) follows from the fact that Θ^* is compact and the Fisher information matrix $g(\theta) > \vartheta I$ over Θ^*. □

Denote the interior of the set Θ^* by Θ^*_{int}. For Θ^*_{int} we can state the following fact.

Fact 3. *Set $\mathscr{A}$ is closed*:

$$\mathscr{A} = \left\{\alpha : \int \exp(\alpha^{\mathrm{T}}\mathbf{c}(\mathbf{x}))\exp(\theta^{\mathrm{T}}\mathbf{c}(\mathbf{x})) < \infty, \right.$$
$$\left. \forall\theta \in \Theta^*_{\mathrm{int}} \text{ and } \alpha \in \Re^p \right\}.$$

Proof. Assume $\mathscr{A}$ is not closed. Then there exists a converging sequence $\{\alpha_i\} \subset \mathscr{A}$, $\overline{\alpha} \notin \mathscr{A}$, and $\exists\overline{\theta} \in \Theta^*_{\mathrm{int}}$ such that

$$\int \exp(\overline{\alpha}^{\mathrm{T}}\mathbf{c}(\mathbf{x}))\exp(\overline{\theta}^{\mathrm{T}}\mathbf{c}(\mathbf{x}))\,\mathrm{d}\mathbf{x} > M, \quad \forall M \in \Re.$$

Since Θ^*_{int} is an open set, $\exists\varepsilon > 0$ such that $\mathscr{N}_\varepsilon(\overline{\theta}) \in \Theta^*_{\mathrm{int}}$. Also, since $\{\alpha_i\}$ is a converging sequence, $\exists k > 0$ such that $\alpha_k \in \mathscr{N}_\varepsilon(\overline{\alpha})$. This implies that $\theta_1 \in \Theta^*_{\mathrm{int}}$ where $\theta_1 = \overline{\theta} + \overline{\alpha} - \alpha_k$. Therefore,

$$\int \exp(\alpha_k^{\mathrm{T}}\mathbf{c}(\mathbf{x}))\exp(\theta_1^{\mathrm{T}}\mathbf{c}(\mathbf{x}))\,\mathrm{d}\mathbf{x} < \infty.$$

On the other hand, we know $\exp(\alpha_k^{\mathrm{T}}\mathbf{c}(\mathbf{x}))\exp(\theta_1^{\mathrm{T}}\mathbf{c}(\mathbf{x})) = \exp(\overline{\alpha}^{\mathrm{T}}\mathbf{c}(\mathbf{x}))\exp(\overline{\theta}^{\mathrm{T}}\mathbf{c}(\mathbf{x}))$. Thus we get a contradiction, and therefore, $\mathscr{A}$ is closed. □

Proof of Lemma 7. Let $\mathscr{M}$ be a set defined as follows:

$$\mathscr{M} = \{\mathbf{m} : \mathbf{m} = \theta_1 - \theta_2, \forall\theta_1, \theta_2 \in \Theta^*\}.$$

We claim that $\mathscr{M}$ is compact. To prove this, assume that $\{\mathbf{m}_i\}$ is a sequence in $\mathscr{M}$, i.e. $\mathbf{m}_i \in \mathscr{M}$ such that $\lim_{i\to\infty}\mathbf{m}_i = \overline{\mathbf{m}}$. We know that there exist sequences $\{\theta_{1,i}\}$ and $\{\theta_{2,i}\}$ such that $\mathbf{m}_i = \theta_{1,i} - \theta_{2,i}$ and $\theta_{1,i}, \theta_{2,i} \in \Theta^*$. Since Θ^* is compact there exist converging subsequences $\{\overline{\theta}_{1,i}\}$ and $\{\overline{\theta}_{2,i}\}$ in Θ^*. This implies that $\overline{\mathbf{m}} = \overline{\theta}_1 - \overline{\theta}_2$, where $\overline{\theta}_1$ and $\overline{\theta}_2$ are the limits of the subsequences $\{\overline{\theta}_{1,i}\}$ and $\{\overline{\theta}_{2,i}\}$. Since $\overline{\theta}_1$ and $\overline{\theta}_2 \in \Theta^*$, then $\overline{\mathbf{m}} \in \mathscr{M}$, therefore $\mathscr{M}$ is closed. Since Θ^* is bounded, $\mathscr{M}$ is bounded and therefore, it is compact.

We define set $\mathscr{A}_1$ as follows:

$$\mathscr{A}_1 = \left\{\alpha : \int \exp(\alpha^{\mathrm{T}}\mathbf{c}(\mathbf{x}))\exp(\theta^{\mathrm{T}}\mathbf{c}(\mathbf{x})) < \infty, \right.$$
$$\left. \forall\theta \in \Theta^* \text{ and } \alpha \in \Re^p \right\}.$$

Clearly, $\mathscr{A}_1 \subset \mathscr{A}$, where $\mathscr{A}$ defined in Fact 3 is shown to be closed. From (A4), $\Psi^*_n(\mathbf{x}) = \exp(\overline{\alpha}^{\mathrm{T}}\mathbf{c}(\mathbf{x}))$ and $\overline{\alpha} \in \mathscr{A} \cap \mathscr{M}$. Since $\mathscr{A} \cap \mathscr{M}$ and Θ^* are compact we have

$$\min_{\theta\in\Theta^*}\min_{\alpha\in\mathscr{A}\cap\mathscr{M}} \|E_\theta\Psi^*_n(\mathbf{x})\| \leqslant \|E_\theta\Psi^*_n(\mathbf{x})\| \leqslant \max_{\theta\in\Theta^*}\max_{\alpha\in\mathscr{A}\cap\mathscr{M}} \|E_\theta\Psi^*_n(\mathbf{x})\|.$$

So (a) is true with $k_1 = \min_{\theta\in\Theta^*}\min_{\alpha\in\mathscr{A}\cap\mathscr{M}}\|E_\theta\Psi_n^*(\mathbf{x})\|$ and $k_2 = \max_{\theta\in\Theta^*}\max_{\alpha\in\mathscr{A}\cap\mathscr{M}}\|E_\theta\Psi_n^*(\mathbf{x})\|$. Similarly, since $u(\cdot)\in\mathscr{F}_{k\kappa}$, (b) is true.

Using the arguments from above and in Fact 2, we can show that $\|(\mathrm{d}/\mathrm{d}\theta)E_\theta\Psi_n^*(\mathbf{x})u(\mathbf{x})\|$ is uniformly bounded. Since Θ^* is convex and compact, (c) is true (Khalil, 1996). □

References

Abbott, E., & Powell, D. (1999). Land–vehicle navigation using GPS. *Proceedings of the IEEE*, *87*(1), 145–162.

Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, *50*(2), 174–188.

Azimi-Sadjadi, B., & Krishnaprasad, P. S. (2001a). Integer ambiguity resolution in GPS using particle filtering. *Proceedings of American control conference*, June 2001 (pp. 3761–3766).

Azimi-Sadjadi, B., & Krishnaprasad, P. S. (2001b). *Approximate nonlinear filtering and its application for GPS*. Institute for Systems Research Technical Report TR 2001-26, September.

Beard, R., Gunther, J., Lawton, J., & Stirling, W. (1999). Nonlinear projection filter based on Galerkin approximation. *Journal of Guidance Control, and Dynamics*, *22*(2), 258–266.

Billingsley, P. (1995). *Probability and measure*. (3rd ed.), New York: Wiley.

Brigo, D. (1996). *Filtering by projection on the manifold of exponential densities*. Ph.D. thesis, Department of Economics and Econometrics, Vrije Universiteit, Amsterdam.

Brigo, D., Hanzon, B., & LeGland, F. (1995). A differential geometric approach to nonlinear filtering: The projection filter. *Proceedings of the 34th IEEE conference on decision and control*, December 1995 (pp. 4006–4011).

Carpenter, J., Clifford, P., & Fearnhead, P. (1999). An improved particle filter for non-linear problems. *IEE Proceedings-Radar, Sonar and Navigation*, *146*(1), 2–7.

Carvalho, H., Del Moral, P., Monin, A., & Salut, G. (1997). Optimal nonlinear filtering in GPS/INS integration. *IEEE Transactions on Aerospace and Electronic Systems*, *33*(3), 835–850.

Crain, B. R. (1976). Exponential models, maximum likelihood estimation, and the Haar condition. *Journal of the American Statistical Association*, *71*(355), 737–740.

Davis, M. H. A., & Marcus, S. I. (1981). An introduction to nonlinear filtering. In M. Hazewinkel, & J. C. Willems (Eds.), *Stochastic systems*: *The mathematics of filtering and identification and applications* (pp. 53–75).

Del Moral, P. (1996). Non linear filtering: Interacting particle solution. *Markov Processes and Related Fields*, *2*(4), 555–580.

Doucet, A., de Freitas, N., & Gordon, N. (2001). Sequential Monte Carlo methods in practice. *Statistics for engineering and information science*. New York: Springer.

Elkaim, G., O'Connor, M., Bell, T., & Parkinson, B. (1996). System identification of a farm vehicle using carrier-phase differential GPS. *ION Conference*, September 1996 (pp. 485–494).

Farrell, J. A., & Barth, M. (1998). *The global positioning system and inertial navigation*. New York: McGraw-Hill.

Fearnhead, P. (1998). *Sequential Monte Carlo methods in filter theory*. Ph.D. thesis, Merton College, University of Oxford.

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*(410), 398–409.

Gordon, N. J., Salmond, D. J., & Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F (Radar and Signal Processing)*, *140*(2), 107–113.

Hasminskii, R. Z. (1980). *Stochastic stability of differential equations*. The Netherlands, Rockville, Maryland, USA: Sijthoff & Noordhoff Alphen aan den.

Hofmann-Wellenhof, B., Lichtenegger, H., & Collins, J. (1993). *Global positioning system: Theory and practice*. (2nd ed.), New York: Springer.

Iannotta, B. (2002). Vortex Draws Flight. *Aerospace America* (pp. 26–30).

Isard, M., & Blake, A. (1996). Contour tracking by stochastic propagation of conditional density. *Proceedings of European conference on computer vision* (pp. 343–356).

Khalil, H. K. (1996). *Nonlinear systems*. (2nd ed.), Upper Saddle River: Prentice-Hall.

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state models. *Journal of Computational and Graphical Statistics*, *5*, 1–25.

Lehmann, E. L., & Casella, G. (1998). *Theory of point estimation*. (2nd ed.), New York: Springer.

Maybeck, P. M. (1982). *Stochastic models, estimation, and control* (Vol. 2). New York: Academic Press.

Milstein, G. N. (1995). *Numerical integration of stochastic differential equations*. New York: Kluwer Academic Publishers.

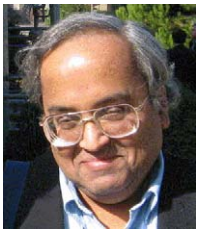
Rubin, D. B. (1988). Using the SIR algorithm to simulate posterior distributions. In: J.M. Bernardo, M.H. DeGroot, D.V. Lindley, & A.F.M. Smith (Eds.), *Bayesian statistics* (pp. 395–402). Oxford: Oxford University Press.

Smith, A. F. M., & Gelfand, A. E. (1992). Bayesian statistics without tears: A sampling-resampling perspective. *The American Statistician*, *46*, 84–88.

Sobczyk, K. (1991). *Stochastic differential equations with application to physics and engineering*. Dordrecht, MA: Kluwer Academic Publishers.

Zickel, R., & Nehemia, N. (1994). GPS aided dead reckoning navigation. *Proceedings of the ION National Technical Meeting* (pp. 577–586).

Babak Azimi-Sadjadi received his B.Sc. from Sharif University of Technology in 1989, his M.Sc. from Tehran University in 1992, and his Ph.D. from the University of Maryland at College Park in 2001 all in Electrical Engineering. He is currently with the Department of Electrical, Systems, and Computer Engineering of Rensselaer Polytechnic Institute where he is a Research Assistant Professor. He also has a joint appointment with the Institute for Systems Research of University of Maryland as a research associate. His research interests include, nonlinear filtering, networked control systems, and wireless networks.



P. S. Krishnaprasad received the Ph.D. degree from Harvard University in 1977. He was on the Faculty of the Systems Engineering Department at Case Western Reserve University from 1977 to 1980. He has been with the University of Maryland since August 1980, where he has held the position of Professor of Electrical Engineering since 1987, and a joint appointment with the Institute for Systems Research since 1988. He is also a member of the Applied Mathematics Faculty. He has held visiting positions with Erasmus University (Rotterdam); the Department of Mathematics at University of California, Berkeley; the Department of Mathematics at University of Groningen (The Netherlands); the Mathematical Sciences Institute at Cornell University; and the Mechanical and Aerospace Engineering Department at Princeton University. Krishnaprasad's interests include geometric control theory, filtering and signal processing theory, robotics, acoustics, and biologically-inspired approaches to control, sensing and computation. He is currently actively engaged in the study of collectives, i.e. communicating, networked control systems. P. S. Krishnaprasad was elected a Fellow of the IEEE in 1990.