

Approximate Nonlinear Filtering and Its Applications for GPS

B. Azimi-Sadjadi and P.S. Krishnaprasad

Institute for Systems Research
University of Maryland at College Park
College Park, MD 20742

September, 2000

Abstract

In this paper we address the problem of nonlinear filtering in the presence of integer uncertainty. In the simulation results we show that Particle Filtering is capable of resolving integer ambiguity in the given nonlinear setup. Motivated by these results we introduce a new Particle Filtering algorithm that can reduce the computational complexity for a certain class of problems. In this class, it is assumed that the conditional density of the state of the system given the observations is close to a known exponential family of densities. The proof of convergence of the approximated density to the actual density is given, and the application for GPS positioning is stated.

1 Introduction

GPS provides world wide positioning with acceptable accuracy, if four or more satellites are in view. Although the satellite constellation guarantees availability of four or more satellites (sometimes even nine) world wide, natural or man-made obstacles can easily block the satellites' signal. To overcome this vulnerability, one might think of integrating dead reckoning or INS with the GPS [1][2][3]. In this case, the INS or the dead reckoning provide positioning that is adjusted by the GPS.

Using Differential GPS (DGPS) allows the user to have a more accurate measurement. In fact, a good portion of the positioning error can be removed from the estimation. This and new technology allow the use of the carrier phase as part of the positioning information. This can increase the accuracy of the estimation to centimeter, or in the static case, to millimeter levels. This can happen only if we are able to estimate the number of full cycles of the carrier phase, which cannot be measured. This problem is called integer ambiguity resolution [4][5] [6][7].

Although Carrier Phase DGPS (CPDGPS) allows for very accurate positioning, it is very sensitive to obstacles that can block satellite signals and cycle slips. A good estimation algorithm should be able to quickly estimate the integer ambiguity on the fly. Most of the algorithms use integer least square methods for this [6][7][4]. In [4] a Kalman filter type of setup is used to estimate the integer ambiguity.

This research was supported in part by the Army Research Office under the ODDR&E MURI97 Program Grant No. DAAG55-97-1-0114 to the Center for Dynamics and Control of Smart Structures (through Harvard University), and by the National Science Foundation under the Learning and Intelligent Systems Initiative Grant CMS9720334

In most of the applications, integrated INS/GPS, dead reckoning/GPS, or vehicle dynamics/GPS, linearization of the dynamics and the GPS observation is the main tool for estimation [6][7][4]. It can be shown [3] that when the number of satellites is below a certain level or the geometry of the current constellation is near singular, the linearization causes the system to be unobservable. In this case, it is important to use a nonlinear setup for the estimation problem. In [3] this case is studied by using an approximation for nonlinear filtering [8][9].

Except for very special cases in nonlinear settings, estimating the state given the observations results in an infinite dimensional filter [10]. Therefore, approximation methods of finite dimension are very appealing. The most widely used approximation filtering method is the Extended Kalman Filter (EKF), which is a heuristic method based on the linearization of the state dynamics and observation near the nominal path [10]. EKF is computationally simple but, the convergence of the conditional distribution to the actual distribution is not guaranteed.

Projection Filtering (PrF) is another approximation method [11][12][13][14]. In PrF it is assumed that the conditional density of the state of the system can be approximated by a member of parametric family of densities. In this case, estimating the conditional density is equivalent to estimating the parameter of the family. In [11][12][13] the exponential family of densities is chosen as the parametric family. In [14] the approach is different; there a Galerkin approximation is used for solving the Fokker-Planck equation [10].

An entirely different approach to approximate the conditional density was proposed in [8][9]. This method is based on the Monte Carlo method and is called Particle Filtering (PaF). In this method, the particles at time t_i are i.i.d. random vectors that are distributed according to the empirical conditional distribution of the state, given the observations up to time t_i . These particle/state vectors are used in the state equation to find the values of particles at time t_{i+1} . Then at time t_{i+1} , the empirical distribution is evaluated according to the values of the particles. The new observation at time t_{i+1} is taken into account through Bayes' Rule to calculate the conditional empirical distribution, and this process can be repeated. In [8] it is proved that by a large enough number of particles, one can get an approximate conditional distribution that is arbitrarily close to the true conditional distribution.

In the cases where we have some prior information about the distribution, we may be able to do better if we take this information into account. By better, we mean reduction in the computational cost and 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 will be defined. Using this assumption, we replace the empirical distribution in [8] with the Maximum Likelihood Estimate (MLE) of the distribution and obtain a convergence result in Theorems (3) and (4).

To use nonlinear filtering methods for CPDGPS, one should be able to include the integer ambiguity resolution in these methods. In this paper we present some simulation results which show that PaF, with minor modifications, is capable of resolving integer uncertainties present in a problem similar to CPDGPS. One problem of PaF is the need for large number of particles. This problem is even more important for the cases where integer uncertainty is present. The writers are optimistic that PaF for exponential families of distributions is more suitable for nonlinear filtering with integer ambiguity.

In this paper, Section 2 states the nonlinear filtering problem. In Section 3 we review the results in [11][12][13] on projection filtering. In Section 4 we explain particle filtering and we state the results in [8][9]. In Sections 5 and 6 we introduce a new PaF algorithm and we state the main results of this paper. In Section 7 we apply the PaF method to a nonlinear system with integer uncertainty and we present the simulation results. In Section 8 we discuss the future research on

this subject.

2 Nonlinear Filtering, Problem Setup

Filtering problems consist of “estimating” the process $\{\mathbf{x}_t\}$ (or something about it) given the related process, $\{\mathbf{y}_t\}$, which can be observed [15]. The observation is available in an interval, i.e., $\{\mathbf{y}_s, 0 \leq s < t\}$ and the function of the state is estimated at time t . To proceed, we need to give some structure to the concerned processes.

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 there is defined an increasing family of σ -fields, $\{\mathcal{F}_t, 0 \leq t \leq T\}$. It is assumed that each process, $\{\mathbf{x}_t\}$, is adapted to \mathcal{F}_t , i.e., $\{\mathbf{x}_t\}$ is \mathcal{F}_t -measurable for all t . We assume that $\{\mathbf{x}_t\}$ is a vector diffusion process of the form

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

where $\mathbf{x}_t \in \mathcal{R}^n$, and $\mathbf{w}_t \in \mathcal{R}^q$ is a vector from an independent Brownian motion process; the second integral is in Ito sense [16], and the function $\mathbf{f}_t(\cdot)$ and the matrix $G_t(\cdot)$ have the proper sizes. The observation, \mathbf{y}_t , is a discrete time process given as follows:

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

where $\mathbf{y}_{n\tau} \in \mathcal{R}^d$, and $\mathbf{v}_n \in \mathcal{R}^d$ is a discrete time white Gaussian noise process with zero mean and known covariance matrix. The state dynamics and observation equation can be rewritten formally as follows:

$$\begin{aligned} d\mathbf{x}_t &= \mathbf{f}_t(\mathbf{x}_t)dt + G_t(\mathbf{x}_t)d\mathbf{w}_t, & \mathbf{x}_0 \\ \mathbf{y}_{n\tau} &= \mathbf{h}_n(\mathbf{x}_{n\tau}) + \mathbf{v}_n \end{aligned} \quad (3)$$

The noise processes $\{\mathbf{w}_t, t \geq 0\}$, and $\{\mathbf{v}_n, n = 0, 1, \dots\}$, and the initial condition \mathbf{x}_0 are assumed independent. We use Q_t and R_n for the covariance matrices of the processes \mathbf{w}_t and \mathbf{v}_n , respectively. We assume that R_n is invertible for all n 's. We have the following additional assumptions [12]:

(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

$$\begin{aligned} \|\mathbf{f}_t(\mathbf{x}) - \mathbf{f}_t(\mathbf{x}')\| &\leq k_r \|\mathbf{x} - \mathbf{x}'\|, & \text{and} \\ \|G_t(\mathbf{x})Q_tG_t^T(\mathbf{x}) - G_t(\mathbf{x}')Q_tG_t^T(\mathbf{x}')\| &\leq k_r \|\mathbf{x} - \mathbf{x}'\|. \end{aligned} \quad (4)$$

(A2) NON-EXPLOSION: there exists $k > 0$ such that

$$\begin{aligned} \mathbf{x}^T \mathbf{f}_t(\mathbf{x}) &\leq k(1 + \|\mathbf{x}\|^2), & \text{and} \\ \text{trace}(G_t(\mathbf{x})Q_tG_t^T(\mathbf{x})) &\leq k(1 + \|\mathbf{x}\|^2). \end{aligned} \quad (5)$$

$\forall t \in [0, T]$ and $\forall \mathbf{x} \in \mathcal{R}^n$.

Under assumptions (A1) and (A2), 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 [12].

In addition to these, we assume that the probability distribution of the state \mathbf{x}_t , given the observation up to time t , $\pi_t(d\mathbf{x}) = P(\mathbf{x}_t \in d\mathbf{x} | \mathbf{y}^t)$, where $\mathbf{y}^t = \{\mathbf{y}_n, i = 1, \dots, n, n\tau < t\}$, has a density p_t with respect to the Lebesgue measure on \mathcal{R}^n . Then $\{p_t, t > 0\}$ satisfies the following PDE and updating equations [12]:

$$\begin{aligned} \frac{\partial}{\partial t} p_t &= \mathcal{L}_t^* p_t & n\tau \leq t < (n+1)\tau, \text{ and} \\ p_{n\tau} &= c_n \Psi_n p_{n\tau-} \end{aligned} \quad (6)$$

where

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

and c_n is a normalizing factor.

Solving System (6) constitutes an infinite dimensional filter [10]. This problem can be overcome in certain special cases. In the linear Gaussian case (6) is equivalent to Kalman Filtering. Approximation of probability densities with a parametric family of distributions, in certain cases, has shown good results. In the next sections we review these results.

3 Projection Filtering on Exponential Families of Densities

This section is mainly a review of the results in [13]. We start this section with the definition of the exponential family of distributions.

Definition 1 Let $\{c_1, \dots, c_p\}$ be affinely independent¹ scalar functions defined on \mathcal{R}^n , and assume that the convex set

$$\Theta_0 = \{\theta \in \mathcal{R}^p : \Upsilon(\theta) = \log \int \exp(\theta^T \mathbf{c}(\mathbf{x})) d\mathbf{x} < \infty\},$$

has nonempty interior. Then,

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

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

We denote by $\mathcal{S}^{\frac{1}{2}}$ the square root of the densities of \mathcal{S} , i.e., $\mathcal{S}^{\frac{1}{2}} = \{\sqrt{p(\cdot, \theta)}; \theta \in \Theta\}$. If $p(\cdot, \theta) \in \mathcal{S}$, then $\sqrt{p(\cdot, \theta)} \in L_2$. $\frac{1}{2\sqrt{p(\cdot, \theta)}} \frac{\partial p(\cdot, \theta)}{\partial \theta_i}$, $i = 1, \dots, p$ make a basis for the tangent vector space at $\sqrt{p(\cdot, \theta)}$ to the space $\mathcal{S}^{\frac{1}{2}}$, i.e., the tangent space at $\sqrt{p(\cdot, \theta)}$ is given by [17]:

$$L_{\sqrt{p(\cdot, \theta)}} \mathcal{S}^{\frac{1}{2}} = \text{span}\left\{\frac{1}{2\sqrt{p(\cdot, \theta)}} \frac{\partial p(\cdot, \theta)}{\partial \theta_1}, \dots, \frac{1}{2\sqrt{p(\cdot, \theta)}} \frac{\partial p(\cdot, \theta)}{\partial \theta_p}\right\}. \quad (7)$$

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

The inner product of any two basis elements is defined as follows

$$\begin{aligned} \left\langle \frac{1}{2\sqrt{p(\cdot, \theta)}} \frac{\partial p(\cdot, \theta)}{\partial \theta_i}, \frac{1}{2\sqrt{p(\cdot, \theta)}} \frac{\partial p(\cdot, \theta)}{\partial \theta_j} \right\rangle &= \frac{1}{4} \int \frac{1}{p(\mathbf{x}, \theta)} \frac{\partial p(\mathbf{x}, \theta)}{\partial \theta_i} \frac{\partial p(\mathbf{x}, \theta)}{\partial \theta_j} d\mathbf{x} \\ &= \frac{1}{4} g_{ij}(\theta) \end{aligned} \quad (8)$$

It can be easily seen that $g(\theta) = (g_{ij}(\theta)) = (E[c_i c_j] - E[c_i]E[c_j])$ is the Fisher information matrix of $p(\cdot, \theta)$.

Any member of L_2 can be projected to the tangent space $L_{\sqrt{p(\cdot, \theta)}} \mathcal{S}^{\frac{1}{2}}$ according to the following projection formula

$$\begin{aligned} \Pi_\theta : L_2 \supseteq V &\rightarrow L_{\sqrt{p(\cdot, \theta)}} \mathcal{S}^{\frac{1}{2}} \\ v &\rightarrow \sum_{i=1}^p \sum_{j=1}^p 4g^{ij}(\theta) \left\langle v, \frac{1}{2\sqrt{p(\cdot, \theta)}} \frac{\partial p(\cdot, \theta)}{\partial \theta_j} \right\rangle \frac{1}{2\sqrt{p(\cdot, \theta)}} \frac{\partial p(\cdot, \theta)}{\partial \theta_i}. \end{aligned} \quad (9)$$

Projection filtering seeks a solution for p_t for (6) that lies in \mathcal{S} . Of course, this solution is only an exponential density, but we hope, by choosing the proper family, we can keep the approximation error small (in L_2 sense).

If we consider the square root of the density in (6), we get

$$\frac{\partial \sqrt{p_t}}{\partial t} = \frac{1}{2\sqrt{p_t}} \frac{\partial p_t}{\partial t} = \frac{1}{2\sqrt{p_t}} \mathcal{L}_t^* p_t. \quad (10)$$

Define $\alpha_{t, \theta} = \frac{\mathcal{L}_t^* p_t(\cdot, \theta)}{p_t(\cdot, \theta)}$. We assume that for all $\theta \in \Theta$ and all $t \geq 0$, $E_{p(\cdot, \theta)}\{|\alpha_{t, \theta}|^2\} < \infty$, which implies that $\frac{\mathcal{L}_t^* p_t(\cdot, \theta)}{\sqrt{p_t(\cdot, \theta)}}$ is a vector of L_2 for all $\theta \in \Theta$ and all $t \geq 0$ [12].

Now assume that in equation (10), for $\{\sqrt{p_t}, t \geq t_0\}$, starting at time $n\tau$ from the initial condition, $\sqrt{p_{n\tau}} = \sqrt{p(\cdot, \theta_{n\tau})} \in \mathcal{S}^{\frac{1}{2}}$ for some $\theta_{n\tau} \in \Theta$. Under these assumptions, the right hand side of (10) is in L_2 , which can be projected into the finite dimensional tangent vector space $L_{\sqrt{p(\cdot, \theta_{n\tau})}} \mathcal{S}^{\frac{1}{2}}$. The projection filter for the exponential family, \mathcal{S} , in the interval $[n\tau, (n+1)\tau)$, is defined as the solution of the following differential equation in the same interval:

$$\frac{\partial \sqrt{p_t(\cdot, \theta_t)}}{\partial t} = \Pi_{\theta_t} \frac{\mathcal{L}_t^* p_t(\cdot, \theta_t)}{2\sqrt{p_t(\cdot, \theta_t)}}, \quad (11)$$

We also assume that $\mathbf{h}_n(\mathbf{x})$ in equation (2) is time invariant, i.e., $\mathbf{h}_n(\mathbf{x}) = \mathbf{h}(\mathbf{x})$, and the components of $\mathbf{h}(\mathbf{x})$, $h^i(\mathbf{x})$, and $\|\mathbf{h}(\mathbf{x})\|_{R^{-1}}^2$ are linear combinations of $c_i(\mathbf{x})$, $i = 1, \dots, p$.

$$\frac{1}{2} \|\mathbf{h}(\mathbf{x})\|_{R^{-1}}^2 = \sum_{i=1}^p \lambda_i^0 c_i(\mathbf{x}) \quad \text{and} \quad h^k(\mathbf{x}) = \sum_{i=1}^p \lambda_i^k c_i(\mathbf{x}), \quad k = 1, \dots, d \quad (12)$$

where $\|\mathbf{x}\|_A = \sqrt{\mathbf{x}^T A \mathbf{x}}$. Then if \mathbf{v}_n is stationary with the covariance matrix $R_n = R$, the likelihood function $\Psi_n(n)$ can be written as follows:

$$\begin{aligned} \Psi_n(\mathbf{x}) &= \exp\left(\frac{-1}{2} (\mathbf{y}_{n\tau}^T R^{-1} \mathbf{y}_{n\tau})\right) \exp\left(\frac{-1}{2} (\mathbf{h}^T(\mathbf{x}) R^{-1} \mathbf{h}(\mathbf{x})) + (\mathbf{y}_{n\tau}^T R^{-1} \mathbf{h}(\mathbf{x}))\right) \\ &= A_n \exp\left(-\sum_{i=1}^p \lambda_i^0 c_i(\mathbf{x}) + \sum_{k=1}^p \left(\sum_{i=1}^p \lambda_i^k z_{n\tau}^k\right) c_i(\mathbf{x})\right), \end{aligned} \quad (13)$$

where $\mathbf{z}_{n\tau} = \mathbf{y}_{n\tau}^T R^{-1}$, and A_n is a constant depending on $\mathbf{y}_{n\tau}$. Therefore the coefficient $\Psi_n(\mathbf{x})$, is a member of exponential family of distributions. This family is closed under multiplication. Using all of these facts, we can present the following theorem [12]:

Theorem 1 For system (3), where \mathbf{w}_t is a Brownian motion process with covariance Q_t and \mathbf{v}_i is a white Gaussian noise with covariance R , we assume (A1) and (A2) to be true. We also assume that $\frac{1}{2}\|\mathbf{h}(\mathbf{x})\|_{R^{-1}}^2 = \sum_{i=1}^p \lambda_i^0 c_i(\mathbf{x})$, $h^k(\mathbf{x}) = \sum_{i=1}^p \lambda_i^k c_i(\mathbf{x})$, for $k = 1, \dots, d$, and $E_{p(\cdot, \theta)} \|\frac{\mathcal{L}_t^* p(\cdot, \theta)}{p(\cdot, \theta)}\|^2 < \infty$, $\forall \theta \in \Theta$, $\forall t \geq 0$. Then for all $\theta \in \Theta$, and all $t \geq 0$, $\Pi_\theta \frac{\mathcal{L}_t^* p(\cdot, \theta)}{\sqrt{p(\cdot, \theta)}}$ is a vector on the exponential manifold $S^{\frac{1}{2}}$. The projection filter density, $p_t^\Pi = p_t(\cdot, \theta_t)$ is described by

$$\begin{aligned} \frac{\partial \sqrt{p_t(\cdot, \theta_t)}}{\partial t} &= \Pi_{\theta_t} \frac{\mathcal{L}_t^* p_t(\cdot, \theta_t)}{2\sqrt{p_t(\cdot, \theta_t)}}, & n\tau \leq t < (n+1)\tau \\ p_{n\tau}(\cdot, \theta_{n\tau}) &= c_n \Psi_n(\mathbf{y}_{n\tau}) p_{n\tau^-}(\cdot, \theta_{n\tau^-}), \end{aligned}$$

and the projection filter parameter satisfies the following combined differential and stochastic difference equations:

$$\begin{aligned} g(\theta_t) d\theta_t &= E_{\theta_t} \{\mathcal{L}_t \mathbf{c}\} dt, & n\tau \leq t < (n+1)\tau, \\ \theta_{n\tau} &= \theta_{n\tau^-} - \lambda_0^0 + \sum_{k=1}^d \lambda_0^k z_n^k, \end{aligned}$$

where

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

and $\lambda_0^i = [\lambda_1^i, \dots, \lambda_p^i]^T$, $i = 0, \dots, d$, and z_n^k is the k th component of $\mathbf{z}_{n\tau}^T = R^{-1} \mathbf{y}_{n\tau}$.

From now on we use E_θ and $E_{p(\cdot, \theta)}$ interchangeably.

As can be seen from the result of the theorem, the calculation of the conditional probability density is reduced to the calculation of the parameter of the exponential family. But still, solving the differential equation in the theorem is not an easy task. At each moment $g(\theta_t)$ and $E_{\theta_t} \{\mathcal{L}_t \mathbf{c}\}$ need to be calculated. This requires a heavy computational load. In this paper, we introduce a Monte Carlo method to calculate the parameter of the exponential family with a more affordable computational load.

Although the PrF gives a better solution than EKF, there is no known error bound with which we can compare the distance between the real distribution and the distribution given by the PrF. In the next section we review PaF as an alternative to optimal nonlinear filtering.

4 Particle Filtering

Consider either the continuous dynamics and discrete observation in (3) or the discrete case,

$$\begin{aligned} \mathbf{x}_{n+1} &= \mathbf{f}_n(\mathbf{x}_n) + G_n(\mathbf{x}_n) \mathbf{w}_n, & \mathbf{x}_0 \\ \mathbf{y}_n &= \mathbf{h}_n(\mathbf{x}_n) + \mathbf{v}_n. \end{aligned} \tag{14}$$

We assume that in both cases, the initial distribution for \mathbf{x}_0 is given. The propagation of the conditional density, at least conceptually, can be expressed as follows [10]:

- Step 1 . Initialization:

$$p_{0|0}(\mathbf{x}_0 | \mathbf{y}_0) = p(\mathbf{x}_0).$$

- Step 2 . Diffusion:

$$p_{n+1|n}(\mathbf{x}_{n+1}|\mathcal{Y}_n) = \int p(\mathbf{x}_{n+1}|\mathbf{x}_n)p_{n|n}(\mathbf{x}_n|\mathcal{Y}_n)d\mathbf{x}_n,$$

where $\mathcal{Y}_n = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\}$.

- Step 3 . Bayes' rule update:

$$p_{n+1|n+1}(\mathbf{x}_{n+1}|\mathcal{Y}_{n+1}) = \frac{p(\mathbf{y}_{n+1}|\mathbf{x}_{n+1})p_{n+1|n}(\mathbf{x}_{n+1}|\mathcal{Y}_n)}{\int p(\mathbf{y}_{n+1}|\mathbf{x}_{n+1})p_{n+1|n}(\mathbf{x}_{n+1}|\mathcal{Y}_n)d\mathbf{x}_{n+1}},$$

- Step 4 . . $n \leftarrow n + 1$; go to Step (2).

The conditional density given by the above steps is exact, but in general it can be viewed as an infinite dimensional filter, thus, not implementable. PaF, in brief, is an approximation method that mimics the above calculations with a finite number of operations using Monte Carlo method. The procedure for Paf is as follows [18][8]:

Algorithm 1 *Particle Filtering*

- *Step 1 . Initialization*

◊ Sample $\mathbf{x}_0^1, \dots, \mathbf{x}_0^N$, N i.i.d. random vectors with the distribution $P_0(\mathbf{x})$.

- *Step 2 . Diffusion*

◊ Find $\hat{\mathbf{x}}_{n+1}^1, \dots, \hat{\mathbf{x}}_{n+1}^N$ from the given $\mathbf{x}_n^1, \dots, \mathbf{x}_n^N$, using the dynamics rules:

$$d\mathbf{x}_t = \mathbf{f}_t(\mathbf{x}_t)dt + G_t(\mathbf{x}_t)d\mathbf{w}_t, \quad n\tau \leq t < (n+1)\tau$$

or

$$\mathbf{x}_{n+1} = \mathbf{f}_n(\mathbf{x}_n) + G_n(\mathbf{x}_n)\mathbf{v}_n.$$

- *Step 3 . Find the empirical distribution*

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

- *Step 4 . Use Bayes Rule*

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

- *Step 5 . Resample*

◊ Sample $\mathbf{x}_{n+1}^1, \dots, \mathbf{x}_{n+1}^N$ 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.

It is customary to call $\mathbf{x}_n^1, \dots, \mathbf{x}_n^N$ particles. In the next few lines, we try to explain in words the evolution of these particles using the above algorithm.

Let $\hat{\mathbf{x}}_n^1, \dots, \hat{\mathbf{x}}_n^N$ be the distinct particles at time n before incorporating the observation at time n . The probability of each particle is $\frac{1}{N}$, and it is 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. Therefore, 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 would have N distinct particles. This automatically makes the approximation one of better resolution in the areas where the probability is higher.

In [8] it is proved under some conditions that

$$E \left(\left| \frac{1}{N} \sum_{i=1}^N f(\hat{\mathbf{x}}_n^i) - E_{P_n}(f(\mathbf{x})) \right| \right) \rightarrow 0 \quad (15)$$

$$N \rightarrow 0$$

for every bounded Borel test function, $f(\cdot)$.

One problem in using PaF method is the computational cost. Specially for high dimensional system, getting reasonable accuracy means using a large N , which causes heavy computation. In the next section, we propose a method that can reduce the number of particles for a certain class of problems.

5 Particle Filtering for Exponential Families of Densities

In the previous sections, we saw two approximation methods for nonlinear filtering. 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. It seems that if, before hand, we know some properties about the distribution, we can do better in performance than just using the empirical distribution. In the following, we assume that the conditional density lies in a parametric family of densities. We will see that with this assumption, we can show the convergence of the approximated density to the actual one. Forcing the density to lie in a parametric family induces some error to the estimate of the density, but we hope to find the proper family that results in an acceptable error.

In this section, after introducing our algorithm, we present some convergence results, and after that, we compare our method with the methods introduced in the previous sections.

For System (3), we assume that the probability density of \mathbf{x}_t , given the observation, is in a family of exponential densities \mathcal{S} . This assumption is rather strong. We will drop this assumption later, and we will only assume that there exists a known family of densities that approximates the real density well, i.e., with acceptable accuracy.

With this assumption, the proposed algorithm is as follows:

Algorithm 2 *Particle Filtering for an Exponential Family of Densities.*

- *Step 1 . Initialization*

◇ *Sample $\mathbf{x}_0^1, \dots, \mathbf{x}_0^N$, N i.i.d. random variable with the densities, $p_0(\mathbf{x})$.*

- *Step 2 . Diffusion*

◊ Find $\hat{\mathbf{x}}_{n+1}^1, \dots, \hat{\mathbf{x}}_{n+1}^N$ from the given $\mathbf{x}_n^1, \dots, \mathbf{x}_n^N$, using the dynamics rule:

$$d\mathbf{x}_t = \mathbf{f}_t(\mathbf{x}_t)dt + G_t(\mathbf{x}_t)d\mathbf{w}_t, \quad i\tau \leq t < (i+1)\tau$$

- *Step 3 . Find the MLE of $\hat{\theta}_{n+1}$ given $\hat{\mathbf{x}}_{n+1}^1, \dots, \hat{\mathbf{x}}_{n+1}^N$ [19]*

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

- *Step 4 . Use Bayes Rule*

$$p_{n+1|n+1}(\mathbf{x}, \hat{\theta}_{n+1}) = \frac{\exp(\hat{\theta}_{n+1}^T \mathbf{c}(\mathbf{x}) - \Upsilon(\hat{\theta}_{n+1})) \Psi_{n+1}(\mathbf{x})}{\int \exp(\hat{\theta}_{n+1}^T \mathbf{c}(\mathbf{x}) - \Upsilon(\hat{\theta}_{n+1})) \Psi_{n+1}(\mathbf{x}) d\mathbf{x}}$$

- *Step 5 . Resample*

◊ Sample $\mathbf{x}_{n+1}^1, \dots, \mathbf{x}_{n+1}^N$ according to $p_{n+1|n+1}(\mathbf{x}, \hat{\theta}_{n+1})$.

- *Step 6 . $n \leftarrow n + 1$; go to Step (2).*

To generate $\mathbf{x}_{n+1}^1, \dots, \mathbf{x}_{n+1}^N$, a Gibbs sampler can be used [20]. This brings an extra computational cost, which should be taken into account when choosing Algorithm 2 over Algorithm 1.

It is constructive to discuss the structure of the ML estimator. We are going to use this structure for the proof of convergence.

Let $\hat{\mathbf{x}}_n^1, \dots, \hat{\mathbf{x}}_n^N$ be the value of the particles, right before the measurement at time n . The MLE of $\theta_n, \hat{\theta}_n$, satisfies the first order necessary condition

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

Therefore, we get

$$\frac{1}{N} \sum_{i=1}^N c_j(\hat{\mathbf{x}}_n^i) = E_{\hat{\theta}_n}(c_j(\mathbf{x})), \quad \text{for } j = 1, \dots, p \quad (16)$$

where from now on, by $E_{\theta}(u(\mathbf{x}))$ we mean the expectation of $u(\mathbf{x})$ under the probability density $p(\mathbf{x}, \theta)$. Equation (16) suggests 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 (16). Let $F_j(\theta)$ be as follows:

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

For simplicity we drop the index n from θ_n . It is easy to see that

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

This shows that $(-\frac{\partial F_i}{\partial \theta_j})_{i,j} = g(\theta)$, where $g(\theta)$ is the Fisher information matrix of the exponential density at θ , and, by assumption positive definite. Therefore (16) is the necessary and sufficient condition for optimality.

In the next few pages, we prove the convergence of the MLE of θ_n , $\hat{\theta}_n$, to θ_n in the mean square sense. This results in the convergence of the density in the weak sense.

In each iteration the proposed algorithm starts from the density $p_{\hat{\theta}_t}(\mathbf{x}_t|\mathbf{y}^t)$, $t = \tau n$, where $\hat{\theta}_t$ is the best estimate θ_t according to the algorithm. After a full iteration the algorithm yields $\hat{\theta}_{t+1}$ which is the best estimate of θ_{t+1} . The error in $\hat{\theta}_{t+1}$ is a combination of the series of possible errors for which we want to find the upper bound. The first source of error is the error in $\hat{\theta}_t$ which will propagate even if no other error is considered. The other source comes from the fact that in each iteration new particles are resampled based on the estimated density which is different from the actual density. Finally the last source of error comes from the discretization of the stochastic dynamics of the system. We want to emphasize that here, we assume that the density of \mathbf{v}_n is stationary and Gaussian, and $\exp(-\frac{1}{2}(\mathbf{y}_{n\tau} - \mathbf{h}(\mathbf{x}_{n\tau}))^T R^{-1}(\mathbf{y}_{n\tau} - \mathbf{h}(\mathbf{x}_{n\tau})))$ lies in the family of the densities. Therefore, no other error is added to the estimate because of the Bayes correction.

We recall the following fact [19]:

Fact 1 *For the family of densities \mathcal{S} with probability density $p(\mathbf{x}, \theta) = \exp(\theta^T \mathbf{c}(\mathbf{x}) - \Upsilon(\theta))$, assume that the Fisher information matrix $g(\theta) = (E(c_i(\mathbf{x})c_j(\mathbf{x})) - E(c_i(\mathbf{x}))E(c_j(\mathbf{x})))_{i,j}$ is positive definite. This implies that the likelihood function*

$$l(\theta) = \theta^T \mathbf{C}(\mathbf{x}) - \Upsilon(\theta),$$

is strictly concave. Also if $c_1(\mathbf{x}), \dots, c_p(\mathbf{x})$, the components of $\mathbf{c}(\mathbf{x})$, are affinely independent almost everywhere, then for the system of equations

$$c_j(\mathbf{x}) = E_\theta[c_j(\mathbf{x})], \quad j = 1, \dots, p,$$

if a solution exists², it is unique. In addition if $\mathbf{x}_1, \dots, \mathbf{x}_N$ are N i.i.d. random variables distributed according to $f_\theta(\mathbf{x}) = \exp(\theta^T \mathbf{c}(\mathbf{x}) - \Upsilon(\theta))$, then the MLE of θ , $\hat{\theta}_N$, has the following property:

$$\begin{aligned} \hat{\theta}_N &= \arg \max_{\theta} \prod_{i=1}^N p(\mathbf{x}_i, \theta) \\ \sqrt{N}(\hat{\theta}_N - \theta) &\sim \mathcal{N}(0, g^{-1}(\theta)). \end{aligned}$$

Using this fact, it is easy to see that

$$E\left(\|\hat{\theta}_N - \theta\|^2\right) = \frac{1}{N} \text{trace}(g^{-1}(\theta)),$$

therefore, when $N \rightarrow \infty$, $\hat{\theta}_N \rightarrow \theta$ in the m.s. sense. On the other hand, $\hat{\theta}_N$ is the solution to (16). Using the strong law of large numbers [22], when $N \rightarrow \infty$ the LHS in (16) goes to $E_\theta(c_j(\mathbf{x}))$, $j = 1, \dots, p$, with probability one. In other words, the solution to (16) when the LHS is the exact $E_\theta(c_j(\mathbf{x}))$, $j = 1, \dots, p$, gives the exact solution for θ . Using this argument, one can expect that by finding a good estimate of the left hand side of (16), a good estimate of θ is accessible. In each iteration of the algorithm presented in this section the estimate of the LHS of

²In [21] it is shown that if $N > p$, the solution exists almost surely.

(16) is found by using the Monte Carlo method and the approximate solution for the stochastic differential equation (3).

To approximate the solution to the stochastic differential equation (3), we employ the method used in [23]. In the following, we review this method briefly. The stochastic differential equation in (3) can be rewritten as follows:

$$d\mathbf{x}_t = \mathbf{f}_t(\mathbf{x}_t) dt + \sum_{r=1}^q \mathbf{g}_t^r(\mathbf{x}_t) dw_t^r, \quad (17)$$

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

$$\Lambda_r u = \left(\mathbf{g}_r, \frac{\partial}{\partial \mathbf{x}} \right) u,$$

$$Lu = \left(\frac{\partial}{\partial t} + \left(\mathbf{f}, \frac{\partial}{\partial \mathbf{x}} \right) + \frac{1}{2} \sum_{r=1}^q \sum_{i=1}^n \sum_{j=1}^n g_i^r g_j^r \frac{\partial^2}{\partial \mathbf{x}_i \partial \mathbf{x}_j} \right) u,$$

where $\left(\mathbf{a}, \frac{\partial}{\partial \mathbf{x}} \right) = \sum_{i=1}^n a_i \frac{\partial}{\partial \mathbf{x}_i}$. Then, the approximate solution for the SDE can be written as follows:

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{x}_k + \sum_{r=1}^q g_{t_k}^r \xi_k^r h^{\frac{1}{2}} + \mathbf{f}_{t_k} h + \sum_{r=1}^q \sum_{i=1}^n (\Lambda_r \mathbf{g}^r)_{t_k} \xi_k^{ir} h + \\ &\quad \frac{1}{2} \sum_{r=1}^q (L\mathbf{g}^r + \Lambda_r \mathbf{f})_{t_k} \xi_k^r h^{\frac{3}{2}} + (L\mathbf{f})_{t_k} \frac{h^2}{2}, \end{aligned} \quad (18)$$

where h is the step size and the coefficients $\mathbf{g}_{t_k}^r$, \mathbf{f}_{t_k} , $(\Lambda_i \mathbf{g}^r)_{t_k}$, etc., are computed at the point (t_k, \mathbf{x}_k) , and the sets of random variables ξ_k^r , ξ_k^{ir} are independent for distinct k and can, for each k , be modeled as follows:

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

and ξ^i and ζ^j are independent random variables satisfying

$$\begin{aligned} E\xi_i &= E\xi_i^3 = E\xi_i^5 = 0, & E\xi_i^2 &= 1, & E\xi_i^4 &= 3, \\ E\zeta_j &= E\zeta_j^3 = 0, & E\zeta_j^2 &= \zeta_j^4 = 1. \end{aligned}$$

In particular, ξ_i can be modeled by the law $P(\xi = 0) = \frac{2}{3}$, $P(\xi = \sqrt{3}) = P(\xi = -\sqrt{3}) = \frac{1}{6}$, and ζ_j can be modeled by $P(\zeta = -1) = P(\zeta = 1) = \frac{1}{2}$.

Definition 2 We say that a function $u(\cdot)$ belongs to the class \mathcal{F} , written as $u \in \mathcal{F}$, if we can find constants, $k > 0$, and κ , such that for all $\mathbf{x} \in \mathcal{R}^n$, the following inequality holds:

$$\|u(\mathbf{x})\| \leq k(1 + \|\mathbf{x}\|^\kappa).$$

If a function $u(\mathbf{x}, s)$ depends not only on $\mathbf{x} \in \mathcal{R}^n$, but also on a parameter $s \in S$, then we say that $u(\mathbf{x}, s)$ belongs to \mathcal{F} (with respect to the variable \mathbf{x}) if the inequality holds uniformly in $s \in S$.

The following theorem summarizes the weak approximation results for (18)[23].

Theorem 2 *Suppose (A1) from Section (2), and suppose that the functions $\mathbf{f}(\cdot)$, $\mathbf{g}^r(\cdot)$, $r = 1, \dots, q$ together with the partial derivatives of sufficiently high order, belong to class \mathcal{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 the function $u(\cdot)$ and all its derivatives up to order 6 belong to class \mathcal{F} , the approximation (18) has the order of accuracy 2, in the sense of weak approximation, i.e.,*

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

where K is a constant 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 stochastic differential equation, respectively.

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

$$\begin{aligned} & \|Eu(\mathbf{x}_{0,\mathbf{x}_0}(t_k)) - \frac{1}{N} \sum_{i=1}^N u(\hat{\mathbf{x}}_{0,\mathbf{x}_0^i}(t_k))\| \leq \\ & \|Eu(\mathbf{x}_{0,\mathbf{x}_0}(t_k)) - Eu(\hat{\mathbf{x}}_{0,\mathbf{x}_0}(t_k))\| + \|Eu(\hat{\mathbf{x}}_{0,\mathbf{x}_0}(t_k)) - \frac{1}{N} \sum_{i=1}^N u(\hat{\mathbf{x}}_{0,\mathbf{x}_0^i}(t_k))\|. \end{aligned}$$

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

$$E\|Eu(\mathbf{x}_{0,\mathbf{x}_0}(t_k)) - \frac{1}{N} \sum_{i=1}^N u(\hat{\mathbf{x}}_{0,\mathbf{x}_0^i}(t_k))\| \leq Kh^2 + \frac{k'}{N^{1/2}}, \quad (19)$$

where K and k' are constants and h is the step size for the approximation of the solution of the stochastic differential equation.

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

Lemma 1 *For the SDE*

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

assume that $\mathbf{f}_t(\cdot)$, $G_t(\cdot)$ are such that for the Brownian motion, \mathbf{w}_t , the probability density of the state \mathbf{x}_t lies in the family \mathcal{S} for Θ bounded, with $g(\theta)$ positive definite and bounded away from zero. We also assume the conditions in **Fact 1** and in **Theorem 2** with $\mathbf{c}(\mathbf{x})$ replacing $u(\mathbf{x})$. Then, there exist k_1 and k_2 such that

$$E[\|\theta_t - \hat{\theta}_t\|] \leq k_1 h^2 + \frac{k_2}{N^{1/2}}, \quad t \in [0, t_f] \quad (20)$$

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 (18) is applied. From (19) we know that:

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

On the other hand, from (16), we know that $\hat{\theta}$ is a solution to the system of equation

$$\frac{1}{N} \sum_{i=1}^N c_j(\hat{\mathbf{x}}_t^i) = E_{\hat{\theta}_t}(c_j(\mathbf{x}_t)), \quad \text{for } j = 1, \dots, p.$$

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

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

On the other hand, we know that $E_{\theta}(\mathbf{c}(\mathbf{x}))$ is a differentiable and one to one function of θ . The derivative of this function, $g(\theta)$, is positive definite and bounded away from zero, so $\exists \alpha > 0$ such that

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

Taking the expectation on both sides of the inequality we have

$$\begin{aligned} E\|\theta_t - \hat{\theta}_t\| &\leq \alpha E \left\| \frac{1}{N} \sum_{i=1}^N \mathbf{c}(\hat{\mathbf{x}}_t^i) - E_{\theta_t}(\mathbf{c}(\mathbf{x}_t)) \right\| \\ &\leq \alpha \left(Kh^2 + \frac{k'}{N^{1/2}} \right) \\ &= \left(k_1 h^2 + \frac{k_2}{N^{1/2}} \right) \end{aligned}$$

◇

Now we are ready to present the main result of this section.

Theorem 3 *For System (3) assume that $\mathbf{f}_t(\cdot)$, $G_t(\cdot)$, and $\mathbf{h}(\cdot)$ are such that for Brownian motion \mathbf{w}_t , and the Gaussian noise \mathbf{v}_n , the conditional probability density of the state \mathbf{x}_t , conditioned on the observations, lies in the family \mathcal{S} for Θ bounded and for $t \in [0, T]$. We also assume the conditions in **Fact 1** and in **Theorem 2** with $\mathbf{c}(\mathbf{x})$ replacing $u(\mathbf{x})$. Then, if $g^{-1}(\theta) E_{\theta_t}(\mathcal{L}_t \mathbf{c}(\mathbf{x}))$ is Lipschitz with the Lipschitz constant L and $g(\theta)$ is positive definite and bounded away from zero, there exist l_1 and l_2 such that*

$$E\|\theta_{n\tau} - \hat{\theta}_{n\tau}\| \leq \sum_{i=0}^{n-1} \exp(Li\tau) \left(l_1 h^2 + \frac{l_2}{N^{1/2}} \right), \quad n\tau \in [0, T],$$

where $\hat{\theta}_{n\tau}$ is the estimate of $\theta_{n\tau}$, and N and h are the number of particles and the time step, respectively.

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 of the estimate of $\theta_{t'}$, $\hat{\theta}_{t'}$, is a combination of the error of the estimate in $\hat{\theta}_t$ and the error added in the time interval $[t, t']$. Let $\tilde{\theta}_{t'}$ be the estimate of $\theta_{t'}$, if the error due to resampling and the approximation of the SDE solution is not taken into account in the interval $[t, t']$, then

$$\|\theta_{t'} - \hat{\theta}_{t'}\| \leq \|\theta_{t'} - \tilde{\theta}_{t'}\| + \|\tilde{\theta}_{t'} - \hat{\theta}_{t'}\|.$$

If the conditional distribution stays in the exponential family of distributions, θ_t has to satisfy the following differential equation:

$$\dot{\theta} = g^{-1}(\theta) E_{\theta_t}(\mathcal{L}_t \mathbf{c}(\mathbf{x})) dt, \quad n\tau \leq t < (n+1)\tau.$$

By the assumption of the theorem, $g^{-1}(\theta) E_{\theta_t}(\mathcal{L}_t \mathbf{c}(\mathbf{x}))$ is Lipschitz with Lipschitz constant L , then by continuity of the solution of the differential equation with respect to the initial condition [24], we know that

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

therefore,

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

Also from the Lemma 1, $\exists k_1(t')$ and $k_2(t')$ such that

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

therefore,

$$E\|\theta_{t'} - \hat{\theta}_{t'}\| \leq E\|\theta_t - \hat{\theta}_t\| e^{L(t'-t)} + k_1(t')h^2 + \frac{k_2(t')}{N^{1/2}}.$$

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 $\hat{\theta}_{t'}$. Therefore, starting from the initial point θ_0 we get

$$E\|\theta_{n\tau} - \hat{\theta}_{n\tau}\| \leq \sum_{i=0}^{n-1} \exp(Li\tau) \left(l_1 h^2 + \frac{l_2}{N^{1/2}} \right), \quad n\tau \in [0, T]$$

where

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

◇

Here, we would like to make a few remarks:

- The result of Theorem (3) can be easily extended to convergence in the mean square sense.
- The assumption that the probability density stays in the family of densities, \mathcal{S} , does not seem very realistic. But with our approach, we should be able to get the result in [12]. In fact, in [12] the evolution of the density is forced to stay in the family at every single moment. In our method, we only force the density to be in the family at the end of each full iteration. This allows the estimated density to be closer to the actual density.
- In [12] 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. It surely affects the assumption that the density stays in the family, but as we explained earlier, this assumption is not realistic to begin with, and it will be dropped.

- If $u(\cdot)$ and its derivatives up to order six are in class \mathcal{F} , then

$$\begin{aligned} E \|E_\theta u(\mathbf{x}) - E_{\theta^*} u(\mathbf{x})\| &\longrightarrow 0, \\ N &\longrightarrow \infty \\ h &\longrightarrow 0 \end{aligned}$$

This is the criteria that was used in [8].

6 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 an exponential family of densities. Instead we have the following assumptions:

(A3) The density in (6) stays close to the given exponential family \mathcal{S} in a weak sense:

$$\forall t \in [0, T], \quad \forall u \in \mathcal{F} \quad \exists \theta_t^* \in \Theta^* \quad \text{s.t.} \quad \|E_{p_t}(u(\mathbf{x})) - E_{\theta_t^*}(u(\mathbf{x}))\| \leq \epsilon \quad (21)$$

where Θ^* is closed and bounded.

(A4) $\forall \theta_1, \theta_2 \in \Theta^*$ and $\forall u \in \mathcal{F} \quad \exists K_1, K_2$ such that

$$\begin{aligned} \|E_{\theta_1} u(\mathbf{x}) - E_{\theta_2} u(\mathbf{x})\| &\leq K_1 \|\theta_1 - \theta_2\| \\ \|\theta_1 - \theta_2\| &\leq K_2 \|E_{\theta_1} u(\mathbf{x}) - E_{\theta_2} u(\mathbf{x})\| \end{aligned}$$

In the following we go through the proof of the theorem that we state later precisely. Assume $\hat{\theta}_n$ is calculated according to Algorithm (2) and assume $p_{n|n}(\mathbf{x}, \hat{\theta}_n)$ is such that $\forall u \in \mathcal{F}$

$$E \|E_{\hat{\theta}_{n|n}} u(\mathbf{x}) - E_{\theta_{n|n}^*} u(\mathbf{x})\| \leq \delta. \quad (22)$$

where $\theta_{n|n}^*$ satisfies

$$\|E_{p_{n|n}} u(\mathbf{x}) - E_{\theta_{n|n}^*} u(\mathbf{x})\| \leq \epsilon. \quad (23)$$

Using the distribution $p_{n|n}(\mathbf{x}, \hat{\theta}_n)$, new particles $\mathbf{x}_n^1, \dots, \mathbf{x}_n^N$ are produced. The approximate solution for the SDE at time $(n+1)\tau$ maps these particles to $\hat{\mathbf{x}}_{n+1}^1, \dots, \hat{\mathbf{x}}_{n+1}^N$. From these new particles $\hat{\theta}_{n+1}$ is calculated. From (22) and (23) we have

$$E \|E_{p_{n|n}} u(\mathbf{x}) - E_{\hat{\theta}_{n|n}} u(\mathbf{x})\| \leq \delta + \epsilon. \quad (24)$$

We define the function $\mathbf{r}(\mathbf{x})$ as follows:

$$\mathbf{r}(\mathbf{x}) = E \mathbf{c}(\hat{\mathbf{x}}_{n,\mathbf{x}}((n+1)\tau))$$

where $\hat{\mathbf{x}}_{n,\mathbf{x}}((n+1)\tau)$ is the approximate solution of SDE (17) at time $(n+1)\tau$ with the given initial condition \mathbf{x} at time $n\tau$ using the method in (18). Since according to our assumption $\mathbf{c} \in \mathcal{F}$, then by using lemma 9.1 in [23], we have

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

where K_3 and μ only depend on the function $\mathbf{c}(\cdot)$ and the dimension of \mathbf{x} . We assume that $\mathbf{r} \in \mathcal{F}$. If the argument of $\mathbf{r}(\cdot)$ is a random variable, then using (24) we have

$$E\|E_{p_{n|n}} \mathbf{r}(\mathbf{x}) - E_{\hat{\theta}_{n|n}} \mathbf{r}(\mathbf{x})\| \leq \delta + \epsilon. \quad (25)$$

Therefore,

$$E\|E_{p_{n|n}} E[\mathbf{c}(\hat{\mathbf{x}}_{n,\mathbf{x}}((n+1)\tau))|\mathbf{x}] - E_{\hat{\theta}_{n|n}} E[\mathbf{c}(\hat{\mathbf{x}}_{n,\mathbf{x}}((n+1)\tau))|\mathbf{x}]\| \leq \delta + \epsilon.$$

In other words,

$$E\|E_{p_{n|n}} \mathbf{c}(\hat{\mathbf{x}}_{n,\mathbf{x}}((n+1)\tau)) - E_{\hat{\theta}_{n|n}} \mathbf{c}(\hat{\mathbf{x}}_{n,\mathbf{x}}((n+1)\tau))\| \leq \delta + \epsilon. \quad (26)$$

From Theorem (2) we have

$$E\|E_{p_{n|n}} \mathbf{c}(\mathbf{x}_{n,\mathbf{x}}((n+1)\tau)) - E_{p_{n|n}} \mathbf{c}(\hat{\mathbf{x}}_{n,\mathbf{x}}((n+1)\tau))\| \leq K_4 h^2, \quad (27)$$

for some K_4 . Using the Monte Carlo method to calculate the $E_{p_{n|n}} \mathbf{c}(\hat{\mathbf{x}}_{n,\mathbf{x}}((n+1)\tau))$ brings another error term that is due to the finite number of particles as the initial conditions for method (18). The expectation of this error is bounded, i.e. $\exists K_5$ s.t.

$$E\|E_{\hat{\theta}_{n|n}} \mathbf{c}(\hat{\mathbf{x}}_{n,\mathbf{x}}((n+1)\tau)) - \frac{1}{N} \sum_{i=1}^N \mathbf{c}(\hat{\mathbf{x}}_{n,\hat{\mathbf{x}}_i}((n+1)\tau))\| \leq \frac{K_5}{N^{\frac{1}{2}}}, \quad (28)$$

where $\hat{\mathbf{x}}_i(n)$ are distributed according to $p_{n|n}(\mathbf{x}, \hat{\theta}_n)$. Combining (26), (27), and (28) we get

$$E\|E_{p_{n|n}} \mathbf{c}(\mathbf{x}_{n,\mathbf{x}}((n+1)\tau)) - \frac{1}{N} \sum_{i=1}^N \mathbf{c}(\hat{\mathbf{x}}_{n,\hat{\mathbf{x}}_i}((n+1)\tau))\| \leq \delta + \epsilon + K_4 h^2 + \frac{K_5}{N^{\frac{1}{2}}}. \quad (29)$$

Based on (A3) we know that $\exists \theta_{n+1}^*$ such that

$$\|E_{p_{n+1|n}} \mathbf{c}(\mathbf{x}((n+1)\tau)) - E_{\theta_{n+1}^*} \mathbf{c}(\mathbf{x}((n+1)\tau))\| \leq \epsilon. \quad (30)$$

We know that $E_{p_{n+1|n}} \mathbf{c}(\mathbf{x}((n+1)\tau)) = E_{p_{n|n}} \mathbf{c}(\mathbf{x}_{n,\mathbf{x}}((n+1)\tau))$, therefore from (29) and (30) we get

$$E\|E_{\theta_{n+1}^*} \mathbf{c}(\mathbf{x}((n+1)\tau)) - \frac{1}{N} \sum_{i=1}^N \mathbf{c}(\hat{\mathbf{x}}_{n,\hat{\mathbf{x}}_i}((n+1)\tau))\| \leq \delta + 2\epsilon + K_4 h^2 + \frac{K_5}{N^{\frac{1}{2}}}. \quad (31)$$

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

$$E\|E_{\theta_{n+1}^*} \mathbf{c}(\mathbf{x}((n+1)\tau)) - E_{\hat{\theta}_{n+1|n}} \mathbf{c}(\mathbf{x}((n+1)\tau))\| \leq \delta + 2\epsilon + K_4 h^2 + \frac{K_5}{N^{\frac{1}{2}}}. \quad (32)$$

From (A4) it is easy to see that

$$E\|E_{\theta_{n+1}^*} u(\mathbf{x}((n+1)\tau)) - E_{\hat{\theta}_{n+1|n}} u(\mathbf{x}((n+1)\tau))\| \leq K_1 K_2 \left(\delta + 2\epsilon + K_4 h^2 + \frac{K_5}{N^{\frac{1}{2}}} \right), \quad (33)$$

$\forall u \in \mathcal{F}$. Step (4) in Algorithm (2) incorporates the new measurement according to Bayes' Rule. Since the measurement noise is Gaussian and Θ^* is compact $\forall \theta \in \Theta^* \exists \kappa_1$ and κ_2 such that

$$\kappa_1 \leq \int \exp(\theta^T \mathbf{c}(\mathbf{x}) - \Upsilon(\theta)) \Psi_{n+1}(\mathbf{x}) d\mathbf{x} \leq \kappa_2,$$

and

$$\|E_\theta u(\mathbf{x})\Psi_{n+1}(\mathbf{x})\| \leq \kappa_3.$$

From (33) we have

$$\begin{aligned} E\|E_{\theta_{n+1|n+1}^*} u - E_{\hat{\theta}_{n+1|n+1}} u\| &= E\left\|\frac{1}{c^*} E_{\theta_{n+1|n}^*} u \Psi_{n+1} - \frac{1}{\hat{c}} E_{\hat{\theta}_{n+1|n}} u \Psi_{n+1}\right\| \\ &= E\left\|\frac{1}{c^*} E_{\theta_{n+1|n}^*} u \Psi_{n+1} - \frac{1}{\hat{c}} E_{\theta_{n+1|n}^*} u \Psi_{n+1} \right. \\ &\quad \left. + \frac{1}{\hat{c}} E_{\theta_{n+1|n}^*} u \Psi_{n+1} - \frac{1}{\hat{c}} E_{\hat{\theta}_{n+1|n}} u \Psi_{n+1}\right\| \\ &\leq \left(\frac{1}{\kappa_1} + \frac{\kappa_3}{\kappa_1^2}\right) K_1 K_2 \left(\delta + 2\epsilon + K_4 h^2 + \frac{K_5}{N^{\frac{1}{2}}}\right) \\ &= K_6 K_1 K_2 \left(\delta + 2\epsilon + K_4 h^2 + \frac{K_5}{N^{\frac{1}{2}}}\right), \end{aligned} \tag{34}$$

where $c^* = \int \exp(\theta_{n+1}^{*T} \mathbf{c}(\mathbf{x}) - \Upsilon(\theta_{n+1}^*)) \Psi_{n+1}(\mathbf{x}) d\mathbf{x}$ and $\hat{c} = \int \exp(\hat{\theta}_{n+1}^T \mathbf{c}(\mathbf{x}) - \Upsilon(\hat{\theta}_{n+1})) \Psi_{n+1}(\mathbf{x}) d\mathbf{x}$. Comparing (34) and (22) shows that by choosing δ , ϵ , and h , small and N big enough, one can achieve the desired accuracy in convergence. The next theorem summarizes our result in this section.

Theorem 4 *For the system (3) assume (A1), (A2), and (A3). We also assume (A4) and the conditions in Fact 1 and in Theorem 2 with $\mathbf{c}(\mathbf{x})$ replacing $u(\mathbf{x})$, and we assume $\mathbf{r} \in \mathcal{F}$. Then in Algorithm 2 with approximation (18), if*

$$E\|E_{\hat{\theta}_{n|n}} u(\mathbf{x}) - E_{\theta_{n|n}^*} u(\mathbf{x})\| \leq \delta$$

then

$$E\|E_{\theta_{n+1|n+1}^*} u(\mathbf{x}((n+1)\tau)) - E_{\hat{\theta}_{n+1|n+1}} u(\mathbf{x}((n+1)\tau))\| \leq K_6 K_1 K_2 \left(\delta + 2\epsilon + K_4 h^2 + \frac{K_5}{N^{\frac{1}{2}}}\right),$$

for some K_1, \dots, K_6 .

In Theorem (4) only one step of Algorithm (2) is considered, but it is straightforward to use Theorem(4) repeatedly, then for the time interval $[0, T]$, where $T = M\tau$, if $\|E_{\hat{\theta}_{0|0}} u(\mathbf{x}) - E_{\theta_{0|0}^*} u(\mathbf{x})\| \leq \delta_0$, then $\exists \alpha_1, \alpha_2, \alpha_3$, and α_4 such that

$$E\|E_{\theta_{n|n}^*} u(\mathbf{x}((n)\tau)) - E_{\hat{\theta}_{n|n}} u(\mathbf{x}((n)\tau))\| \leq \alpha_1^n \delta_0 + \sum_{i=0}^{n-1} \alpha_1^i (\alpha_2 \epsilon + \alpha_3 h^2 + \alpha_4 N^{-1/2}),$$

for $0 \leq n \leq M$.

7 Particle Filtering for Nonlinear Systems with Constant Integer Uncertainty

Consider the following nonlinear dynamics and observation

$$\begin{aligned} d\mathbf{x}_t &= \mathbf{f}_t(\mathbf{x}_t)dt + G_t(\mathbf{x}_t)d\mathbf{w}_t \\ \mathbf{y}_{n\tau} &= \mathbf{h}_n(\mathbf{x}(n\tau)) + J_n\mathbf{z} + \mathbf{v}_n \end{aligned}$$

where the assumptions and the dimensions for \mathbf{x}_t , $\mathbf{y}_{n\tau}$, \mathbf{w}_t , and \mathbf{v}_n are the same as in the previous sections. We assume that \mathbf{z} is a random integer vector, i.e. $\mathbf{z} \in \mathcal{Z}^m$ and J_n has the proper dimension. Vector \mathbf{z} is assumed to be constant in time. This problem can be set up in discrete time as well. In this case the system dynamics and the observation can be written as follows:

$$\begin{aligned} \mathbf{x}_{n+1} &= \mathbf{f}_n(\mathbf{x}_n) + G_n(\mathbf{x}_n)\mathbf{w}_n \\ \mathbf{y}_n &= \mathbf{h}_n(\mathbf{x}_n) + J_n\mathbf{z} + \mathbf{v}_n \end{aligned}$$

In both setups we assume that the integer uncertainty affects only some components of the observation, and other components are unaffected by \mathbf{z} . The affected components have associated noise components in \mathbf{v}_n that have considerably lower energy. In other words, the uncertain components of $\mathbf{y}_{n\tau}$ (or equivalently \mathbf{y}_n) are considerably more accurate than the other components if the integer ambiguity were known. This suggests that an accurate estimation of \mathbf{z} can increase the accuracy of the estimate of the state of the system significantly. With this explanation our treatment of \mathbf{z} is clear. From the state dynamics and the observation equation we first estimate \mathbf{z} and then, with fixed \mathbf{z} , we use regular nonlinear filtering methods to estimate the state of the system \mathbf{x}_t .

We augment the state \mathbf{x}_t with the integer ambiguity \mathbf{z} . Having done that, the state dynamics and the observation have the following form:

$$\begin{aligned} d \begin{bmatrix} \mathbf{x}_t \\ \mathbf{z}_t \end{bmatrix} &= \begin{bmatrix} \mathbf{f}_t(\mathbf{x}_t) \\ 0 \end{bmatrix} dt + \begin{bmatrix} G_t(\mathbf{x}_t) \\ 0 \end{bmatrix} d\mathbf{w}_t \\ \mathbf{y}_{n\tau} &= \mathbf{h}_n(\mathbf{x}(n\tau)) + J_n\mathbf{z}(n\tau) + \mathbf{v}_n \end{aligned} \tag{35}$$

We assume that the initial distribution of $(\mathbf{x}_0^T, \mathbf{z}_0^T)^T$ is known. Now with this form the state dynamics and the observation have the same form that was studied in Section (4). Therefore, we can apply particle filtering to find the conditional probability distribution of the augmented state. This setup is a special case of the setup in Section (4). In (35) there is no state transition for \mathbf{z}_t , therefore, using PaF in its original form may not be the best option. Recall that in PaF we start with N i.i.d. particles distributed according to the initial distribution. In the resampling part the low probability particles die and the high probability particles produce many particles identical to themselves. Since \mathbf{z}_t does not change, the part of the particles associated to \mathbf{z}_t tends to cover smaller and smaller portions of the state space. In fact, the state space of the integer vectors is defined by the particles at the initial time. This problem can be overcome by modifying the algorithm mentioned in Section (4). In the new algorithm Step 5 is changed in such a way that the particles are the addition of the original particles found by Algorithm 1, with a random vector. The modification is very important for the integer values, since the integers do not have a dynamics that is driven by a random input. In [9], a similar modification has been used for the regular nonlinear filtering setup (no parameter is considered constant). It seems that the convergence results given in [9] can be applied to our case as well.

Based on the modified algorithm, we simulated a nonlinear filtering problem similar to the problem involved in the GPS system.

In a two dimensional space, three transmitters (imagine three pseudo satellites) are mounted on three known points (2000,100000), (0,100000), and (-2000,100000). The moving object can measure its distance from these transmitters. For each pseudo satellite, two types of measurement are possible. One with high measurement noise and the other with low measurement noise. For the low measurement noise, though, there is an integer ambiguity. The dynamics of the moving object for this example is considered to be in discrete time and linear time invariant. The dynamics and observation equation is given as follows:

$$\begin{pmatrix} x_1 \\ v_1 \\ x_2 \\ v_2 \end{pmatrix}_{n+1} = \begin{pmatrix} 1 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ v_1 \\ x_2 \\ v_2 \end{pmatrix}_n + \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{pmatrix}_n,$$

$$\begin{aligned} y_n^a &= \|\mathbf{x} - \mathbf{s}_i\| + v_n^a, \quad i = 1, 2, 3 \\ y_n^b &= \|\mathbf{x} - \mathbf{s}_i\| + n_i + v_n^b, \quad i = 1, 2, 3, \end{aligned}$$

where $\mathbf{x} = (x_1, x_2)^T$, \mathbf{s}_i is the position of pseudo satellite i in two dimensional space, $\Delta t = 0.1$ unit of time, n_i is the integer ambiguity of the pseudo satellite i , and $\mathbf{w} = (w_1, w_2, w_3, w_4)^T$ and $\mathbf{v} = (v_1^a, v_1^b, v_2^a, v_2^b, v_3^a, v_3^b)^T$ are zero mean white Gaussian noise with covariance matrices $\Sigma_{\mathbf{w}} = \text{diag}(1, 0.5, 1, 0.5)$ and $\Sigma_{\mathbf{v}} = \text{diag}(5, 0.2, 5, 0.2, 5, 0.2)$, respectively. In the simulation, it is assumed that the initial condition for the position is distributed in a square of size 200×200 units squared, symmetric with respect to the origin.

In brief, the simulation can be separated into two parts, initialization and the full non-linear filtering. In the initialization part, we start with the initial probability distribution for (x_1, x_2) and from a series of observations, we find an estimate for the probability distribution (v_1, v_2) . In this part, we do not use the dynamics of the moving object. Using our estimate for the probability distribution of (x_1, v_1, x_2, v_2) we find the distribution for the integer ambiguity. After this, the initialization is over, and the full non-linear filter is used. There are some minor numerical considerations that we would like to point out. In the Bayes step of the algorithm, the numbers are usually very small, without proper scaling the original algorithm would not work. In the resampling part, one can use the law of large numbers and regenerate the particles based on their probability without generating random numbers that are time consuming. The result of the simulations are shown in Figures(1), (2), (3), (4), (5), and (6). To display the estimated integers, we simply used the mean value, which is not necessarily the best choice. Of course, since we have the distribution, we can use the MAP estimate of the integers. In this simulation we forced one of the integers to have a jump. Although our algorithm is not designed for these kinds of changes, we see that it can estimate the new integer values. In future, we use special treatment for the times that these kinds of jumps happen. As you can see, the estimate for the integers are reasonably good. The reliability of the estimate for the integers depends on the energy of the noise.

8 Future Works

The simulations results show that our method is capable of estimating the integer ambiguity and the position. There are certain issues that need further investigation. In the following, we itemize these issues:

- What are the proper criteria to stop the integer ambiguity estimation part and fix the integers?

- What happens when a cycle slip happens, i.e. one or more of the integers have a jump? What change detection algorithm is proper and what is the performance of this algorithm? How can we repair the integer ambiguity efficiently?
- What happens when the number of the satellites drops from the critical number?
- How much improvement does the method of Section 5 for integer ambiguity and position estimation have over PaF?

These questions are to be answered in the future work. In addition to these, we shall be more specific in our simulations, and use real GPS data for our results.

In our lab setup we have two GPS receivers that can work in differential mode. One of the receivers is mounted on a radio controlled car. In the setup, we monitor the position of the car on a stationary monitor that is connected to the base station GPS receiver. The position estimation is done by this receiver. In future, we want to integrate the GPS observations and the car dynamics for better estimation results. The car kinematics is a modified version of kinematic car [25].

$$d \begin{pmatrix} x \\ y \\ \theta \end{pmatrix} = \begin{pmatrix} \cos \theta \\ \sin \theta \\ \frac{1}{l} \tan \phi \end{pmatrix} u_n + G(\theta, \phi, x, y) d\mathbf{w}_t$$

where x , y , θ , ϕ , and l are shown in Figure (7).

We assume that two sensors on the car can measure the control u_1 and the angle ϕ (ϕ is another control). The function G depends on the structure of the sensors and the nature of the noise. This model and the DGPS observation will be used for estimating the position and the speed of the car.

References

- [1] N. El-Sheimy and K.P. Schwarz. Navigating Urban Areas by VISAT-A Mobile Mapping System Integrating GPS/INS/Digital Camera for GIS Applications. *Navigation*, 45(4):275–285, Winter 1999.
- [2] R. Zickel and N. Nehemia. GPS Aided Dead Reckoning Navigation. *Proc. ION Nat. Tech. Meeting*, pages 577–586, 1994.
- [3] H. Carvalho, P. Del Moral, A. Monin, and G. Salut. Optimal Nonlinear Filtering in GPS/INS Integration. *IEEE Trans. on Aerospace and Electronic Systems*, 33(3):835–850, July 1997.
- [4] A. Hassibi and S. Boyd. Integer Parametric Estimation in Linear Models with Applications to GPS. *IEEE Trans. on Signal Processing*, 46(11):2938–52, Nov. 1998.
- [5] B. Hofmann-Wellenhof, H. Lichtenegger, and J. Collins. *Global Positioning System: Theory and Practice*. Springer-Verlag, Second edition, 1993.
- [6] P.J.G. Teunissen. A New Method for Fast Carrier Phase Ambiguity Estimation. *Proc. IEEE Position, Location and Navigation Symp.*, pages 562–73, 1994.
- [7] P.J.G. Teunissen. The Least Square Ambiguity Decorrelation Adjustment: a Method for Fast GPS Integer Ambiguity Estimation. *Journal of Geodesy*, 70(1-2):65–83, Nov. 1995.

- [8] P. Del Moral. Non Linear Filtering: Interacting Particle Solution . *Markov Processes and Related Fields*, 2(4):555–580, 1996.
- [9] F. LeGland, C. Musso, and N. Oudjane. An Analysis of Regularized Interacting Particle Methods in Nonlinear Filtering,. In J. Rojicek, M. Valeckova, M. Karry, and K. Warwick, editors, *Proceedings of the 3rd IEEE European Workshop on Computer-Intensive Methods in Control and Signal Processing*, volume 1, pages 167–174. Prague, September 7–9 1998.
- [10] P.M. Maybeck. *Stochastic Models, Estimation, and Control*, volume 2. Academic Press, 1982.
- [11] D. Brigo, B. Hanzon, and F. LeGland. A Differential Geometric Approach to Nonlinear Filtering : the Projection Filter. *Proceedings of the 34th IEEE Conference on Decision and Control*, pages 4006–4011, December 1995.
- [12] D. Brigo. *Filtering by Projection on the Manifold of Exponential Densities*. Ph.D. Thesis Department of Economics and Econometrics, Vrije Universiteit, Amsterdam, October 29 1996.
- [13] D. Brigo and F. LeGland. A Finite Dimensional Filter with Exponential Conditional Density. *Proceedings of the 36th IEEE Conference on Decision and Control*, pages 1643–1644, December 10–12 1997.
- [14] R. Beard, J. Gunther, J. Lawton, and W. Stirling. Nonlinear Projection Filter Based on Galerkin Approximation. *Journal of Guidance Control, and Dynamics*, 22(2):258–266, March–April 1999.
- [15] M.H.A. Davis and S.I. Marcus. An Introduction to Nonlinear Filtering. In *M. Hazewinkel and J.C. Willems edit. Stochastic Systems: The Mathematics of Filtering and Identification and Applications*, pages 53–75, 1981.
- [16] K. Sobczyk. *Stochastic Differential Equation With Application to Physics and Engineering*. Kluwer Academic Publishers, 1991.
- [17] S. Amari. *Differential-Geometrical Methods in Statistics*. Springer-Verlag, 1985.
- [18] N.J. Gordon, D.J. Salmond, and A.F.M. Smith. Novel Approach to Nonlinear/NonGaussian Bayesian State Estimation. *IEE Proceedings-F (Radar and Signal Processing)*, 140(2):107–113, 1993.
- [19] E.L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer-Verlag, New York, second edition, 1998.
- [20] A.E. Gelfand and A.F.M. Smith. Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, 85(410):398–409, 1990.
- [21] B.R. Crain. Exponential Models, Maximum Likelihood Estimation, and the Haar Condition. *Journal of the American Statistical Association*, 71(355):737–740, 1976.
- [22] P. Billingsley. *Probability and Measure*. John Wiley, New York, third edition, 1995.
- [23] G.N. Milstein. *Numerical Integration of Stochastic Differential Equations*. Kluwer Academic Publishers, New York, 1995.
- [24] H.K. Khalil. *Nonlinear Systems*. Printice-Hall, second edition, 1996.

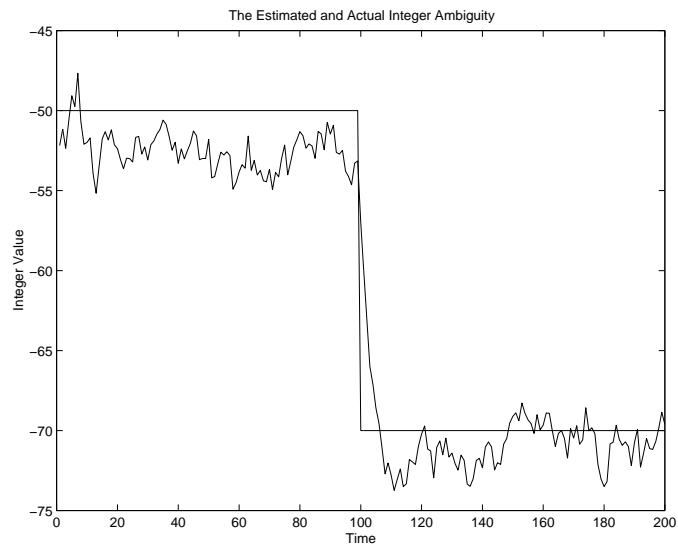


Figure 1: Estimated integer ambiguity versus the actual integer ambiguity of pseudo satellite (1). At time 100 there is a cycle slip of strength -20 for the measured phase of the carrier from pseudo satellite (1) .

- [25] R.M. Murray, Z. Li, and S.S. Sastry. *A Mathematical Introduction to Robotic Manipulation*. CRC Press, 1994.

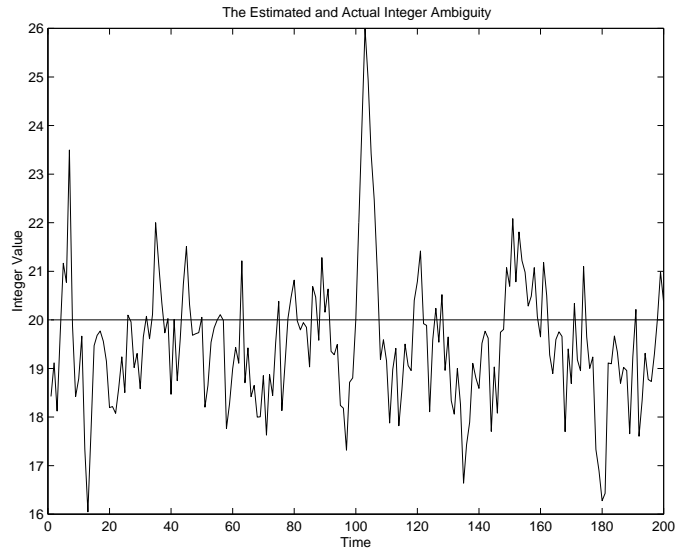


Figure 2: Estimated integer ambiguity versus the actual integer ambiguity of pseudo satellite (2). At time 100 there is a cycle slip of strength -20 for the measured phase of the carrier from pseudo satellite (1) .

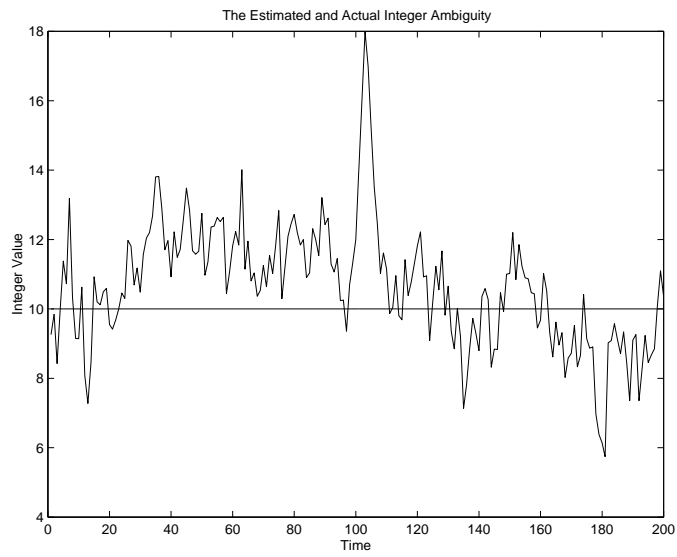


Figure 3: Estimated integer ambiguity versus the actual integer ambiguity of pseudo satellite (3). At time 100 there is a cycle slip of strength -20 for the measured phase of the carrier from pseudo satellite (1) .

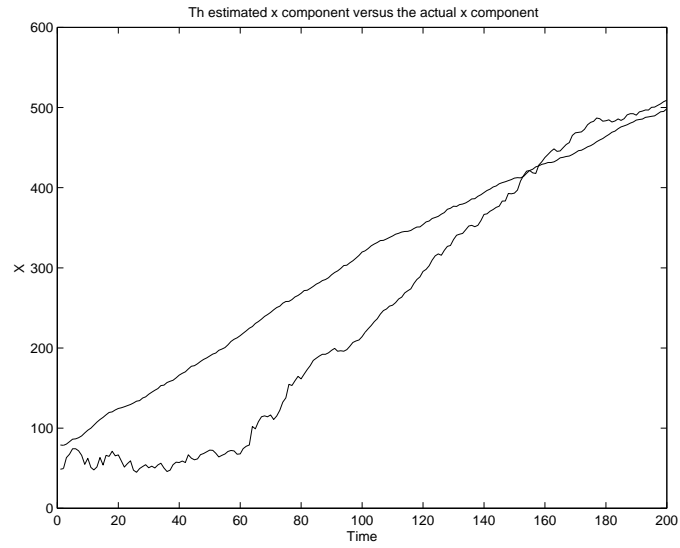


Figure 4: Estimated X component versus the actual X component of the position of the car. At time 100 there is a cycle slip of strength -20 for the measured phase of the carrier from pseudo satellite (1) .

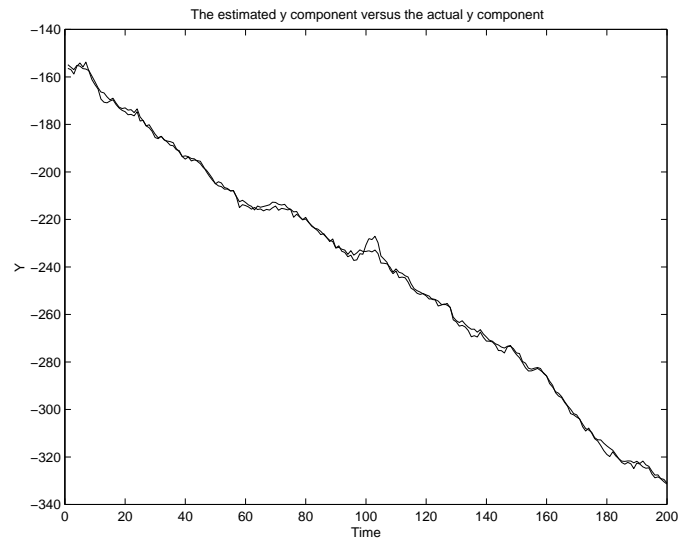


Figure 5: Estimated Y component versus the actual Y component of the position of the car. At time 100 there is a cycle slip of strength -20 for the measured phase of the carrier from pseudo satellite (1) .

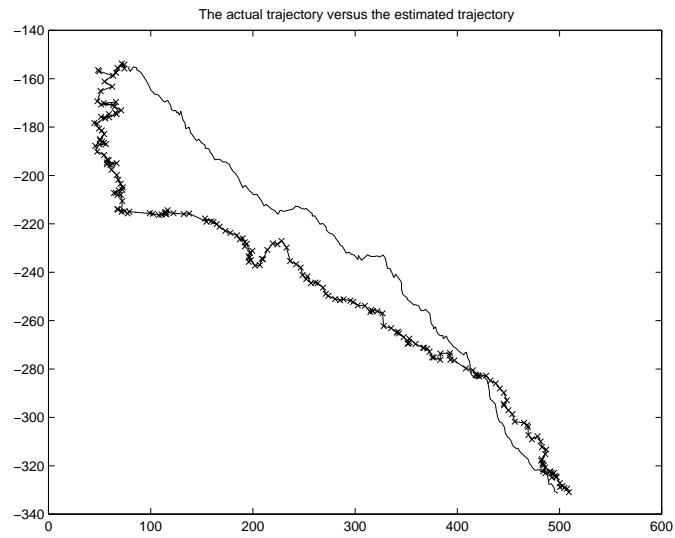


Figure 6: Estimated trajectory versus the actual trajectory of the car. At time 100 there is a cycle slip of strength -20 for the measured phase of the carrier from pseudo satellite (1) .

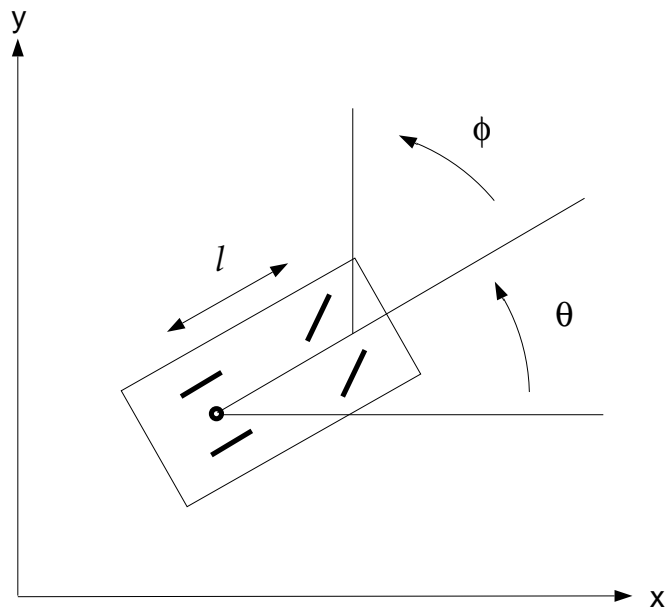


Figure 7: Kinematic car.