

# Convolution Particle Filter for Parameter Estimation in General State-Space Models

Fabien Campillo and Vivien Rossi

## Abstract

The state-space modeling of partially observed dynamical systems generally requires estimates of unknown parameters. The dynamic state vector together with the static parameter vector can be considered as an augmented state vector. Classical filtering methods, such as the extended Kalman filter and the bootstrap particle filter, fail to estimate the augmented state vector. For these classical filters to handle the augmented state vector, a dynamic noise term should be artificially added to the parameter components or to the deterministic component of the dynamical system. However, this approach degrades the estimation performance of the filters. In this work, we propose a variant of the particle filter based on convolution kernel approximation techniques. This approach is tested on a simulated case study.

## Index Terms

Bayes procedures, Hidden Markov models, Marine vehicle detection and tracking, Monte Carlo methods, Nonlinear filters, Parameter estimation, Uncertain systems, bearings-only target motion analysis.

## I. INTRODUCTION

Consider a Markovian dynamical system in discrete time, which takes values in a general state-space. This system is partially observed:  $X_k$  is the unobserved state process and  $Y_k$  the observation process. We assume that this system depends on an unknown parameter  $\theta$ .

This work was partially completed while the authors were at IRISA/INRIA, Rennes, France. It is partially supported by the Cooperative research initiative ARC MICR funded by INRIA.

Fabien Campillo, INRIA – UMR ASB, bât. 29, 2 place Pierre Viala 34060 Montpellier cedex 1 (France) – [Fabien.Campillo@inria.fr](mailto:Fabien.Campillo@inria.fr)

Vivien Rossi, CIRAD – Campus International de Baillarguet, 34398 Montpellier cedex 5 (France) – [vivien.rossi@cirad.fr](mailto:vivien.rossi@cirad.fr)

Our goal is to estimate, in real-time, and simultaneously, the parameter  $\theta$  and the state process  $X_k$  based on the observations  $Y_{1:k} = (Y_1, \dots, Y_k)$ . In a batch context, there exists many possibilities (nonlinear filtering approximation coupled with maximum likelihood estimation techniques or the EM algorithm). In a real-time context there are two main approaches:

- (i) The non Bayesian approach consists of either minimizing a given cost function, such as the conditional least squares criterion or maximizing the likelihood function. These methods are usually performed in batch processes but can also be extended to recursive procedures [4].
- (ii) In the Bayesian approach, the augmented state variable  $(X_k, \theta)$  is processed by a filtering procedure. A prior probability distribution is prescribed for the parameter. These methods are performed recursively in  $k$ .

In both approaches, we must use an approximation of the optimal nonlinear filter. The extended Kalman filter (EKF) and its various alternatives do not always give good results and suffer from an absence of mathematical analysis. The particle filters propose a good alternative: in many practical cases they give better results, moreover their theoretical properties are becoming increasingly well understood [7], [10], [6]. It is thus particularly appealing to use particle filtering in order to estimate parameters in partially observed systems. For a review of the question, one can consult [11] or [17].

The cost function of the approach (i) must be approximated for various values of the parameter  $\theta$ . This is done via the particle approximation of the conditional probability density function (pdf) of  $Y_k$  given  $Y_{1:k-1} = (Y_1, \dots, Y_{k-1})$ . The Monte Carlo nature of this particle approximation will make optimization problematic. Indeed, the approximated cost function is not regular and does not suit classical optimization procedures like the gradient methods. However, recent work shows significant improvements of these aspects [11].

In the present work, we focus on approach (ii): compared to approach (i), its implementation is simpler and its execution faster. In approach (ii),  $\theta$  is considered as a random variable with a prescribed a priori pdf. An extended state variable  $(X_k, \theta)$  joining all the unknown quantities is considered; the posterior pdf of  $(X_k, \theta)$  given  $Y_{1:k-1} = (Y_1, \dots, Y_{k-1})$  is then approximated using particle filters.

Classical particle filtering methods, like the bootstrap particle filter, fail to estimate the augmented state vector. These classical filters can handle the augmented state vector if a dynamic

noise term is artificially added to the parameter components, but this approach degrades the estimation performance of the filters. In this work, we propose a variant of the particle filter based on the convolution particle filter introduced in [20]. The application and the convergence analysis of this filter requires weaker assumptions than the usual particle filters. This is due to the use of convolution kernels to weight the particle.

The paper is organized as follows, first we present the problem. Then we recall the principle of the convolution filter for the dynamical systems without unknown parameters. Next, the Bayesian estimation approach is presented, it also relies on the convolution particle filter. Finally, this approach is tested on a simulated case study.

## II. PROBLEM STATEMENT

Consider a state process  $X_k$  and an observation process  $Y_k$  taking values in  $\mathbb{R}^d$  and  $\mathbb{R}^q$  respectively. This system depends on an unknown parameter  $\theta$  which takes values in  $\mathbb{R}^p$ . Suppose that the state process  $X_k$  is Markovian with transition kernel:

$$m_k(x'|x, \vartheta) \stackrel{\text{def}}{=} f_{X_k|X_{k-1}, \theta}(x'|x, \vartheta), \quad (1)$$

and initial pdf:

$$p_0(x) \stackrel{\text{def}}{=} f_{X_0}(x). \quad (2)$$

Throughout this paper,  $f_{X|Y}(x|y)$  will denote the conditional probability density function (pdf) of  $X$  given  $Y = y$ ;  $f_X(x)$  will denote the pdf of  $X$ .

The observation process  $Y_k$  is related to the state process  $X_k$  through the emission kernel:

$$\Psi_k(y|x, \vartheta) \stackrel{\text{def}}{=} f_{Y_k|X_k, \theta}(y|x, \vartheta). \quad (3)$$

and with the hypothesis that:

$$f_{Y_{1:k}|X_{0:k}}(y_{1:k}|x_{0:k}) = \prod_{\ell=1}^k f_{Y_\ell|X_\ell}(y_\ell|x_\ell) \quad (4)$$

for all  $k > 0$ . Note that for the sake of simplicity, the initial probability distribution (2) of the state process does not depend on the parameter  $\theta$ . Hypothesis (4) means that, conditionally to the state process, the observations are mutually independent and the observation  $Y_k$  depends on the state process only through  $X_k$ .

The model presented here is a hidden Markov model with general state–space, where the transition probability kernel and the emission kernel depend on an unknown parameter. This also includes systems with deterministic dynamics:

$$\dot{\xi}_t = f(\xi_t), \quad Y_k = h(X_k, V_k) \quad (5)$$

where  $0 = t_0 < t_1 < t_2 < \dots$ ,  $V_k$  is a white noise (independent from  $\theta$ ), and  $X_k = \xi_{t_k}$ . Here only the initial condition  $\xi_0$  is unknown. The goal is to estimate  $X_k$  at the current instant  $t_k$ , like in the case study of Section VI, or the initial condition  $X_0$ .

### III. THE CONVOLUTION FILTER

The optimal nonlinear filter is the conditional pdf:

$$\pi_k(x, \vartheta) \stackrel{\text{def}}{=} f_{X_k, \theta | Y_{1:k}}(x, \vartheta).$$

Our aim is to propose a particle approximation of  $\pi_k(x, \vartheta)$ .

For the sake of simplicity, in this section, we suppose that the parameter  $\theta$  is known. We propose a particle approximation of the optimal filter:

$$\pi_k(x) \stackrel{\text{def}}{=} f_{X_k | Y_{1:k}}(x). \quad (6)$$

The conditional pdf  $\pi_k(x)$  is a solution of the so–called sequential Bayes formula. This formula explicitly gives  $\pi_k(x)$  in terms of  $\pi_{k-1}(x)$ , in two steps. First a prediction step, known as the Chapman–Kolmogorov equation, introduces the predicted filter:

$$\pi_{k-}(x') = \int_{\mathbb{R}^d} m_k(x' | x) \pi_{k-1}(x) \, dx \quad (7)$$

where  $\pi_{k-}(x') \stackrel{\text{def}}{=} f_{X_k | Y_{1:k-1}}(x')$ . Second a correction step, namely the Bayes formula, defines the updated filter:

$$\pi_k(x) = \frac{\Psi_k(Y_k | x) \pi_{k-}(x)}{\int_{\mathbb{R}^d} \Psi_k(Y_k | x') \pi_{k-}(x') \, dx'}. \quad (8)$$

Except in the linear–Gaussian case and some other very specific cases, Equations (7) and (8) cannot be solved explicitly. Hence, in general one must get use of approximation techniques.

We propose a convolution particle approximation filter. We suppose that we know how to sample from the conditional laws  $m_k(x' | x)$ ,  $\Psi_k(y | x)$ , for any given  $x \in \mathbb{R}^d$ , and also from the initial pdf  $p_0(x)$ . In contrast with the classical bootstrap particle filter, we do not suppose that the pdf  $\Psi_k$  should be stated explicitly.

### A. Kernel estimation

Convolution particle approximation filters are based on kernel estimation techniques, we briefly introduce them now.

A kernel  $K : \mathbb{R}^d \mapsto \mathbb{R}$  is a bounded positive symmetric application such that  $\int K(x) dx = 1$ .

Let:

$$K_{h_N}(x) \stackrel{\text{def}}{=} \frac{1}{h_N^d} K\left(\frac{x}{h_N}\right)$$

where  $h_N > 0$  is the bandwidth parameter. The Gaussian kernel is  $K(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^d e^{-|x|^2/2}$ . A *Parzen-Rosenblatt* kernel is a kernel such that  $|x|^d K(x) \rightarrow 0$  as  $|x| \rightarrow \infty$ .

Let  $X_1, \dots, X_N$  be i.i.d. random variables with common pdf  $\phi$ . The kernel estimator  $\phi_N$  of  $\phi$  associated with the kernel  $K$  is:

$$\phi_N(x) \stackrel{\text{def}}{=} \frac{1}{N h_N^d} \sum_{i=1}^N K\left(\frac{x-X_i}{h_N}\right).$$

### B. The convolution filter

Let  $\{x_0^i\}_{i=1:N}$  be a sample of size  $N$  of  $p_0$ . We describe now the iteration  $k-1 \rightarrow k$  of the algorithm. Starting from  $\{x_{k-1}^i\}_{i=1:N}$ , we can simulate the following samples:

$$x_{k-}^i \sim m_k(\cdot | x_{k-1}^i), \quad y_{k-}^i \sim \Psi_k(\cdot | x_k^i) \quad \text{for } i = 1, \dots, N.$$

We deduce the following empirical estimate of the joint conditional pdf of  $(X_k, Y_k)$  given  $Y_{1:k-1}$ :

$$\begin{aligned} f_{X_k, Y_k | Y_{1:k-1}}(x, y) \\ \simeq \frac{1}{N} \sum_{i=1}^N \{K_{h_N}^X(x - x_{k-}^i) K_{h_N}^Y(y - y_{k-}^i)\} \end{aligned} \quad (9)$$

where  $K_{h_N}^X$  and  $K_{h_N}^Y$  are suitably chosen kernels of appropriate dimensions. Here we use Parzen-Rosenblatt kernels. Note that in  $K_{h_N}^X$  (resp.  $K_{h_N}^Y$ )  $h_N$  could implicitly depend on  $N$ ,  $d$  and  $x_{k-}^{1:N}$  (resp.  $N$ ,  $q$  and  $y_{k-}^{1:N}$ ). From (9) we deduce the following convolution approximation  $\pi_k^N(x)$  of the optimal filter  $\pi_k(x)$ :

$$\pi_k^N(x) \stackrel{\text{def}}{=} \frac{\sum_{i=1}^N K_{h_N}^X(x - x_{k-}^i) K_{h_N}^Y(Y_k - y_k^i)}{\sum_{i=1}^N K_{h_N}^Y(Y_k - y_k^i)}. \quad (10)$$

Then, we define  $x_k^i \sim \pi_k(\cdot)$  for  $i = 1, \dots, N$ .

---

```

filter initialization:  $x_0^1 \cdots x_0^N \stackrel{\text{iid}}{\sim} p_0(\cdot)$ 
for  $k = 1, 2, \dots$  do
   $x_{k-}^i \sim m_k(\cdot | x_{k-1}^i)$  for  $i = 1 : N$  {state sampling}
   $y_{k-}^i \sim \Psi_k(\cdot | x_{k-}^i)$  for  $i = 1 : N$  {observation
sampling}

  filter updating:
  
$$\pi_k^N(x) = \frac{\sum_{i=1}^N K_{h_N^Y}^Y(Y_k - y_{k-}^i) \times K_{h_N^X}^X(x - x_{k-}^i)}{\sum_{i=1}^N K_{h_N^Y}^Y(Y_k - y_{k-}^i)}$$

   $x_k^1 \cdots x_k^N \sim \pi_k^N(\cdot)$  {resampling}

end for

```

---

TABLE I  
THE CONVOLUTION PARTICLE FILTER.

### C. Comments

A sufficient condition for the  $L^1$ -convergence of  $\pi_k^N$  to  $\pi_k^N$  is  $h_N^{2q} = O(N^{-\alpha})$ , with  $\alpha \in (0, 1)$  and  $N h_N^{q+d} / \log N \rightarrow \infty$  where the bandwidth parameters are the same for the state and the observation spaces, see [20] for details.

The practical use of the convolution filter requires the choice of the kernel functions  $K^X$ ,  $K^Y$  and of the bandwidth parameters  $h_N^X$ ,  $h_N^Y$ . The nature of the kernel does not appreciably affect the quality of the results.

The choice  $h_N^X = C_x \times N^{-1/(4+d)}$ ,  $h_N^Y = C_y \times N^{-1/(4+q)}$  is optimal for the mean square error criterion. The choice of the  $C$ 's is a critical issue for density estimation and sophisticated techniques have been proposed [9]. In the on-line context of nonlinear filtering these techniques are not usable. Moreover, particle filtering is aimed at "tracking" the state and not really at estimating the conditional density precisely.

The generic form  $C_x = c_x \times [\text{Cov}(x_k^1, \dots, x_k^N)]^{1/2}$ ,  $C_y = c_y \times [\text{Cov}(y_k^1, \dots, y_k^N)]^{1/2}$  with  $c_x, c_y \simeq 1$  gives good results. For the simulations in the last section, on taking a Gaussian kernel, we will see that the  $c$ 's are easily adjusted.

#### IV. BAYESIAN SEQUENTIAL PROCESSING

The constant unknown parameter  $\theta$  can be included in the augmented state variable  $(X_k, \theta_k)$  with:

$$\theta_k = \theta_{k-1}, \quad \theta_0 \sim \rho(\cdot). \quad (11)$$

The parameter could be denoted  $\theta_k$  or  $\theta$ . The prior pdf  $f_\theta$  is  $\rho$  and the posterior pdf  $f_{\theta|Y_{1:k}}$  at time  $k$  is the marginal pdf of the nonlinear filter:

$$\pi_k(x, \vartheta) \stackrel{\text{def}}{=} f_{X_k, \theta|Y_{1:k}}(x, \vartheta).$$

The constant dynamic (11) generally leads to divergence of the standard particle filters. This is due to the fact that the parameter space is only explored at the initialization step of the particle filter, causing impoverishment of the diversity of the relevant particles. Among the approaches proposed to avoid this trouble, Storvik [21] marginalizes parameters out of the posterior distribution, then assumes that the concerned parameters depend on sufficient statistics. This avoids the degeneracy of the particles. However, this approach is not practically useful for general systems. Kitagawa [16] and Higuchi [14] set an artificial dynamic on the parameter, such as  $\theta_k = \theta_{k-1} + \zeta_k$ , which increases the a priori variance on the parameter. Gilks and Berzuini [12] add a Markov chain Monte Carlo procedure to increase the particle diversity, but this is cumbersome. To avoid these additions Liu and West [17] proposes smoothing the empirical measure of the parameter posterior probability distribution with a Gaussian distribution.

More generally, regularization techniques are used to avoid particle degeneracy [18]. The regularization of the state model only, requires non-degenerate noises and an analytical expression for  $\Psi_k$ . These restrictions were dropped in [8] by the regularization the observation model. However, as the state model is not regularized, the approach remains subject to particle degeneracy.

In order to circumvent these two problems simultaneously, Rossi and Vila [20] jointly regularized the state model and the observation model. In this approach, the particles include state and observation components. The construction and the theoretical analysis of the corresponding filters differ from the standard particle approach, as they are based on the nonparametric estimate of the conditional densities by convolution kernels. The filter proposed to estimate simultaneously the state and the parameters in (11), extends the results of [20]. It is not necessary for the kernel to be Gaussian, any Parzen–Rosenblatt kernel will be valid.

---

filter initialization:  $x_0^i \sim p_0(\cdot)$  and  $\theta_0^i \sim \rho(\cdot)$  for  $i = 1 : N$

**for**  $k = 1, 2, \dots$  **do**

prediction:

$x_{k-}^i \sim m_k(\cdot | x_{k-1}^i, \theta_{k-1}^i)$  for  $i = 1 : N$  {state sampling}

$y_{k-}^i \sim \Psi_k(\cdot | x_{k-}^i, \theta_{k-1}^i)$  for  $i = 1 : N$  {observation sampling}

$\vartheta_{k-}^i = \vartheta_{k-1}^i$  for  $i = 1 : N$  {parameter sampling}

filter updating:

$$\pi_k^N(x, \vartheta) = \frac{\sum_{i=1}^N K_{h_N}^Y(Y_k - y_{k-}^i) K_{h_N}^X(x - x_{k-}^i) K_{h_N}^\vartheta(\vartheta - \vartheta_{k-}^i)}{\sum_{i=1}^N K_{h_N}^Y(y_k - y_{k-}^i)}$$

$(x_k^i, \vartheta_k^i) \sim \pi_k^N(\cdot, \cdot)$  {resampling}

**end for**

---

TABLE II

THE CONVOLUTION PARTICLE FILTER FOR BAYESIAN ESTIMATION.

The regularization with convolution kernels can also be viewed as artificial noise. Thus, our approach is connected to the methods [16], [14] presented previously. However contrary to these methods, it respects dynamics (11) and allows convergence results. In terms of artificial noise on dynamics, we have identified a whole family of acceptable noises and we have also characterized the way in which their variance must decrease to zero.

The convolution particle filter (Table I) applied to system (11) leads to the algorithm presented in Table II. It provides consistent estimates of  $\pi_k^N(x, \vartheta)$ , but also of the marginal conditional pdf's in  $X_k$  and  $\theta$ .

In practice, the parameter prior pdf  $\rho$ , the number of particles  $N$ , the kernels  $K$  and the associated bandwidth parameters  $h_N$  must be by chosen by the user.

## V. CONVERGENCE

The  $L^1$ -convergence of  $\pi_k^N(x)$  to  $\pi_k(x)$  as  $N \rightarrow \infty$ , for any fixed  $k$ , is proven in [20]. This result applied to the particular case of the extended state-space model (11) reads:



*Theorem 1:* Suppose that  $K$  is a positive Parzen-Rosenblatt kernel, and that for all  $k$ :

- the conditional pdf  $f_{X_k, \theta | Y_{1:k}}(x, \vartheta)$  exists,
- the conditional pdf  $f_{Y_k | Y_{1:k-1}}(y)$  exists, is continuous and strictly positive a.s. (at least in the neighborhood of the actual observation),
- the conditional pdf  $f_{Y_k | X_k, \theta}(y)$  exists and is bounded a.s.,
- $\lim_{N \rightarrow \infty} N h_N^{q+d+p} / \log N = \infty$  and  $h_N^{2q} = O(N^{-\alpha})$  with  $\alpha \in (0, 1)$ .

Then, for any fixed  $k$

$$\lim_{N \rightarrow \infty} \iint |\pi_k^N(x, \vartheta) - \pi_k(x, \vartheta)| \, dx \, d\vartheta = 0 \quad \text{a.s.}$$

where

$$\pi_k^N(x, \vartheta) \stackrel{\text{def}}{=} \frac{\sum_{i=1}^N K_{h_N}^Y(Y_k - y_{k-}^i) K_{h_N}^X(x - x_{k-}^i) K_{h_N}^\theta(\vartheta - \vartheta_{k-}^i)}{\sum_{i=1}^N K_{h_N}^Y(Y_k - y_{k-}^i)}.$$

The theorem can be proved by induction on  $k$ . For  $k = 1$ , the result holds from the convergence of the convolution kernel density approximations. The proof of the recurrence  $k \rightarrow k + 1$  uses a decomposition technique of the  $L_1$ -error. This technique is presented in Theorem 2. The detailed proof of Theorem 1 can be found in [20].

The convolution particle filter leads to the following estimate of the parameter:

$$\theta_k^N \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^N \bar{\theta}_k^i \quad \text{with} \quad \bar{\theta}_k^i \sim \pi_k^{2,N}(\cdot) \quad (12)$$

where

$$\begin{aligned} \pi_k^{1,N}(x) &\stackrel{\text{def}}{=} \int \pi_k^N(x, \vartheta) \, d\vartheta, \\ \pi_k^{2,N}(\vartheta) &\stackrel{\text{def}}{=} \int \pi_k^N(x, \vartheta) \, dx \end{aligned}$$

are the marginal pdf's of the convolution particle filter.

Now we prove the convergence of  $\theta_k^N$ :

*Theorem 2:* In addition to the assumptions of Theorem 1, suppose that the conditional pdf  $f_{\theta_k | Y_{1:k}}(\vartheta)$  exists and that its support is included in a compact set  $C$  then

$$\lim_{N \rightarrow \infty} \theta_k^N = \mathbb{E}[\theta | Y_{1:k}] \quad \text{a.s.}$$

Theorem 2 is transposable to the state variable  $X_k$ . More precisely,

$$X_k^N \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^N \bar{x}_k^i \quad \text{with} \quad \bar{x}_k^i \sim \pi_k^{1,N}(\cdot)$$

is a consistent estimate of  $\mathbb{E}[X_k|Y_{1:k}]$ .

*Proof of Theorem 2:* Consider the pdf  $f_{\theta_k|Y_{1:k}}(\vartheta)$  and let

$$\Delta_N \stackrel{\text{def}}{=} \frac{1}{2} \int |\pi_k^{2,N}(\vartheta) - f_{\theta_k|Y_{1:k}}(\vartheta)| \, d\vartheta \in [0, 1] \text{ a.s.}$$

Consider a  $N$ -sample from the density  $\pi_k^{2,N}$ :

$$S_k^N \stackrel{\text{def}}{=} \{\bar{\theta}_k^1, \dots, \bar{\theta}_k^N\} \quad \text{with} \quad \bar{\theta}_k^i \sim \pi_k^{2,N}(\cdot).$$

We now show that there exists a subsample  $\{\bar{\theta}_k^{i_1}, \dots, \bar{\theta}_k^{i_{M^N}}\} \subset S_k^N$ , and a new sample  $\{\bar{\theta}_k^1, \dots, \bar{\theta}_k^{N_N}\}$ , which together can be considered as sampled from  $f_{\theta_k|Y_{1:k}}$ . Define

$$\gamma_1 \stackrel{\text{def}}{=} \frac{\min(\pi_k^{2,N}, f_{\theta_k|Y_{1:k}})}{1 - \Delta_N}, \quad (13)$$

$$\gamma_2 \stackrel{\text{def}}{=} \frac{\pi_k^{2,N} - \min(\pi_k^{2,N}, f_{\theta_k|Y_{1:k}})}{\Delta_N}, \quad (14)$$

$$\gamma_3 \stackrel{\text{def}}{=} \frac{f_{\theta_k|Y_{1:k}} - \min(\pi_k^{2,N}, f_{\theta_k|Y_{1:k}})}{\Delta_N}. \quad (15)$$

These functions are pdf's. Indeed, consider two pdf's  $h$  and  $g$  on  $\mathbb{R}^p$ , and let  $f = \min(h, g)$ . Define  $H$  (resp.  $G$ ) the subset of  $\mathbb{R}^p$  for which we have  $\min(h, g) = h$  (resp.  $\min(h, g) = g$ ). Let  $I = H \cap G$ . We have:

$$\begin{aligned} \int f &= \int_H h + \int_G g - \int_I h \\ &= 1 - \int_{H^c} h + \int_G g - \int_I h \\ &= 1 - [\int_G h - \int_I h] + \int_G g - \int_I h \\ &= 1 - \int_G h + \int_G g \\ &= 1 - \int_G (h - g) = 1 - \frac{1}{2} \int |h - g| \end{aligned}$$

this last equality follows from Scheffe's lemma. Hence  $(1 - \frac{1}{2} \int |h - g|)^{-1} f$  is a pdf, so  $\gamma_1$  is.

The assertion for (14) and (15) is proved the same way.

We have:

$$\begin{aligned} \pi_k^{2,N} &= \Delta_N \gamma_2 + (1 - \Delta_N) \gamma_1, \\ f_{\theta_k|Y_{1:k}} &= \Delta_N \gamma_3 + (1 - \Delta_N) \gamma_1. \end{aligned}$$

This shows that each  $\bar{\theta}_k^i$  sampled according to  $\pi_k^{2,N}$  is, with probability  $\Delta_N$ , sampled from  $\gamma_2$ .

Let

$$Z_i \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } \bar{\theta}_k^i \sim \gamma_2, \\ 0 & \text{if } \bar{\theta}_k^i \sim \gamma_1, \end{cases}$$

and

$$\mathcal{N}_N \stackrel{\text{def}}{=} \sum_{i=1}^N Z_i, \quad \mathcal{M}_N \stackrel{\text{def}}{=} N - \mathcal{N}_N.$$

The  $Z_i$ 's are Bernoulli variables with parameter  $\Delta_N$ , so  $\mathcal{N}_N \sim \mathcal{B}(N, \Delta_N)$  is binomial.  $\mathcal{M}_N$  is the number of  $\bar{\theta}_k^i$ 's sampled from  $\gamma_1$ . Let  $\{\bar{\theta}_k^{i_1}, \dots, \bar{\theta}_k^{i_{\mathcal{M}_N}}\}$  be this subsample. Then let  $\bar{\theta}_k^1, \dots, \bar{\theta}_k^{\mathcal{M}_N} \stackrel{\text{iid}}{\sim} \gamma_3$  and

$$\{\tilde{\theta}_k^1, \dots, \tilde{\theta}_k^N\} \stackrel{\text{def}}{=} \{\bar{\theta}_k^{i_1}, \dots, \bar{\theta}_k^{i_{\mathcal{M}_N}}\} \cup \{\bar{\theta}_k^1, \dots, \bar{\theta}_k^{\mathcal{M}_N}\},$$

it is a  $N$ -sample from  $f_{\theta_k|Y_{1:k}}$  which holds  $\mathcal{M}_N$  common elements with  $S_k^N$  drawn from  $\pi_k^{2,N}$ .

Let  $\hat{\theta}_k^N \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^N \tilde{\theta}_k^i$  be the associated virtual estimate of  $\theta_k$ . Note that:

$$|\theta_k^N - \mathbb{E}[\theta|Y_{1:k}]| \leq |\theta_k^N - \hat{\theta}_k^N| + |\hat{\theta}_k^N - \mathbb{E}[\theta|Y_{1:k}]|. \quad (16)$$

As  $\{\tilde{\theta}_k^1, \dots, \tilde{\theta}_k^N\}$  are sampled from  $f_{\theta_k|Y_{1:k}}$ , the strong law of large numbers ensures that  $\lim_{N \rightarrow \infty} |\hat{\theta}_k^N - \mathbb{E}[\theta|Y_{1:k}]| = 0$  a.s. It remains to study the first term of the r.h.s. of (16). As  $\{\tilde{\theta}_k^1, \dots, \tilde{\theta}_k^N\}$  and  $\{\bar{\theta}_k^1, \dots, \bar{\theta}_k^N\}$  have  $\mathcal{M}_N$  common elements, we have:

$$|\theta_k^N - \hat{\theta}_k^N| = \frac{1}{N} \left| \sum_{j=1}^{\mathcal{M}_N} \bar{\theta}_k^{i_j} - \sum_{j=1}^{\mathcal{M}_N} \tilde{\theta}_k^{i_j} \right| \leq 2 \frac{\mathcal{M}_N}{N} \max_{\vartheta \in C} |\vartheta|.$$

However  $\frac{1}{N} \mathcal{M}_N$  is the empirical estimate of  $\Delta_N$ , by Hoeffding's inequality, for any  $\Delta_N$ ,

$$\mathbb{P}\left(\left|\frac{\mathcal{M}_N}{N} - \Delta_N\right| \geq \varepsilon\right) \leq 2 \exp\{-2N\varepsilon^2\}. \quad (17)$$

By Theorem 1 we have  $\Delta_N \rightarrow 0$  a.s., then (17) implies that  $\frac{1}{N} \mathcal{M}_N \rightarrow 0$  a.s. The proof is then completed.

## VI. BEARINGS-ONLY TARGET MOTION ANALYSIS: A SIMULATED CASE STUDY

We consider the classical problem of bearings-only target motion analysis in the plane. We adopt the modified polar coordinate system [1]. We address the simplest version of this problem: one target moving with a constant course and velocity. For more advanced problems like multiple targets tracking and association, or maneuvering targets, one can consult [13], [15], [2], [19], [3]. Our purpose is not to propose a sophisticated tracking algorithm but to compare:

- the convolution particle filter (CPF),
- the extended Kalman filter (EKF),
- the standard particle filter (PF)

applied to this well-known case study. The EKF and PF are based on an artificial dynamics for the unknown parameter with a constant noise variance.

Consider a mobile (the target) with a rectilinear uniform motion (i.e., with constant heading and velocity) in the plane. This mobile is tracked by an observer with a given trajectory. The state vector is

$$X_k = \left[ \dot{\beta}(t_k) \frac{\dot{r}(t_k)}{r(t_k)} \beta(t_k) \frac{1}{r(t_k)} \right]^*.$$

where  $r(t)$  is the relative range,  $v(t)$  the relative velocity,  $\beta(t)$  the azimuth. The state vector  $X_k$  is a solution of the following nonlinear noise-free system:

$$X_{k+1} = \Phi(\theta, t_{k+1}, t_k; X_k) \quad (18)$$

where  $\Phi$  is given, see [1] for details; here  $\theta$  is the initial position, velocity, and heading of the target.

The observations are a sequence of bearings corrupted by noise:

$$Y_k = \beta(t_k) + \sigma v_k$$

where  $v_k$  is a white Gaussian noise  $\mathcal{N}(0, 1)$ .

The parameters of the simulation scenario described in Fig. 1 are: total observation duration 1 hour, sampling interval 4 s, target speed 10 m/s, observer speed 10 m/s, observer initial position (0, 0), target initial position (30, 000 m, -5000 m). Hence, the initial true relative distance is 30, 414 m.

The initial pdf for the 3 filters (EKF, PF, CPF) is as follows: the initial relative distance pdf has a mean of 33, 000 m and a standard deviation of 5000 m; the initial pdf on  $\dot{\beta}(t_0)$  is  $\mathcal{N}(0, 0.001^2)$ ; the initial pdf on  $\dot{r}(t_0)/r(t_0)$  is  $\mathcal{N}(0, 0.0005^2)$ .

We perform 20 independent Monte Carlo runs of this scenario with  $\sigma = 1$  degree (Fig. 2) and  $\sigma = 1.5$  degree (Fig. 3). We plot the corresponding empirical positions (the empirical estimated trajectory) and the corresponding empirical uncertainty ellipses (every 10 minutes). For the PF and the CPF we use 10, 000 particles.

It is known that the EKF performs better in the modified polar coordinate system than in the cartesian coordinate system [1]. In Fig. 2, where  $\sigma = 1$  degree, the performance of the CPF is better than that of the EKF. In the Fig. 3, where  $\sigma = 1.5$  degree, the performances of the EKF deteriorate much more than those of the CPF. After the maneuver, the EKF is unable to find the track. In order to make use of the PF it is necessary to add an artificial dynamic to the parameter with constant variance noise: in doing so, the performance of the PF markedly deteriorates.

In Fig. 4, the CPF is presented in the case of the cartesian coordinate system. The EKF is extremely sensitive to the choice of coordinates and presents severe problems of divergence. The performances of the CPF are slightly better in the case of the cartesian coordinate system and are much less sensitive to the choice of coordinates.

Insert Figure 1 here

Insert Figure 2 here

Insert Figure 3 here

Insert Figure 4 here

## VII. CONCLUSION

The convolution particle filter (CPF) presents some interesting features. First, the CPF can be applied if the local likelihood function (3) is not known analytically, or even if the local likelihood function does not exist. The sole prerequisite is to be able to sample from the observation model (3). Second, the CPF can be used to estimate constant parameters or deterministic parts of the state dynamical system, whereas in such situations, the classical particle filters fail due to the particle degeneracy phenomenon. Finally, in terms of implementation, the CPF is as simple as the bootstrap particle filter. The CPF appears to be an efficient extension of the standard particle filter for the case of partially perturbed dynamical systems and online parameter tracking.

## REFERENCES

- [1] V. J. Aidala and S. E. Hammel. Utilization of modified polar coordinates for bearings-only tracking. *IEEE Transactions on Automatic Control*, AC-28(3):283–294, 1983.
- [2] M. S. Arulampalam, B. Ristic, N. Gordon, and T. Mansell. Bearings-only tracking of manoeuvring targets using particle filters. *EURASIP J. Appl. Signal Process.*, 2004(1):2351–2365, 2004.
- [3] T. Bréhard and J.-P. Le Cadre. Hierarchical particle filter for bearings-only tracking. *IEEE Transactions on Aerospace and Electronic Systems*, 43(4), 2007.

- [4] F. Campillo and L. Mevel. Recursive maximum likelihood estimation for structural health monitoring: tangent filter implementations. In *Proceedings of the 44th IEEE Conference on Decision and Control and European Control Conference, Sevilla (Spain)*, 2005.
- [5] O. Cappé, É. Moulines, and T. Ryden. *Inference in Hidden Markov Models*. Springer-Verlag, 2005.
- [6] D. Crişan and A. Doucet. A survey of convergence results on particle filtering for practitioners. *IEEE Transactions on Signal Processing*, 50(3):736–746, 2002.
- [7] P. Del Moral. *Feynman-Kac formulae – Genealogical and interacting particle approximations*. Springer-Verlag, New York, 2004.
- [8] P. Del Moral, J. Jacod, and P. Protter. The Monte Carlo method for filtering with discrete-time observations. *Probability Theory and Related Fields*, 120(3):346–368, 2001.
- [9] L. Devroye and G. Lugosi. *Combinatorial Methods in Density Estimation*. Springer-Verlag, 2001.
- [10] A. Doucet, N. de Freitas, and N. J. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York, 2001.
- [11] A. Doucet and V. B. Tadić. Parameter estimation in general state-space models using particle methods. *Annals of the Institute of Statistical Mathematics*, 55(2):409–422, 2003.
- [12] W. R. Gilks and C. Berzuini. Following a moving target — Monte Carlo inference for dynamic Bayesian models. *Journal of the Royal Statistical Society, Series B*, 63(1):127–146, 2001.
- [13] N. J. Gordon. A hybrid bootstrap filter for target tracking in clutter. *IEEE Transactions on Aerospace and Electronic Systems*, AES-33(1):353–358, 1997.
- [14] T. Higuchi. Monte Carlo filter using the genetic algorithm operators. *Journal of Statistical Computation and Simulation*, 59(1):1–23, 1997.
- [15] C. Hue, J.-P. Le Cadre, and P. Pérez. Tracking multiple objects with particle filtering. *IEEE Transactions on Aerospace and Electronic Systems*, 38(3):791–812, 2002.
- [16] G. Kitagawa. Non-Gaussian state space modeling of nonstationary time series (with discussion). *Journal of the American Statistical Association*, 82(400):1032–1063, 1987.
- [17] J. Liu and M. West. Combined parameter and state estimation in simulation-based filtering. In A. Doucet, N. de Freitas, and N. J. Gordon, editors, *Sequential Monte Carlo Methods in Practice*, pages 197–223. Springer-Verlag, 2001.
- [18] C. Musso, N. Oudjane, and F. Le Gland. Improving regularized particle filters. In A. Doucet, N. de Freitas, and N. J. Gordon, editors, *Sequential Monte Carlo Methods in Practice*, pages 247–271. Springer-Verlag, 2001.
- [19] B. Ristic, M. S. Arulampalam, and N. J. Gordon. *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House, 2004.
- [20] V. Rossi and J.-P. Vila. Nonlinear filtering in discrete time: A particle convolution approach. *Ann. I.S.U.P.*, 50(3):71–102, 2006.
- [21] G. Storvik. Particle filters in state space models with the presence of unknown static parameters. *IEEE Transactions on Signal Processing*, 50(2):281–289, 2002.

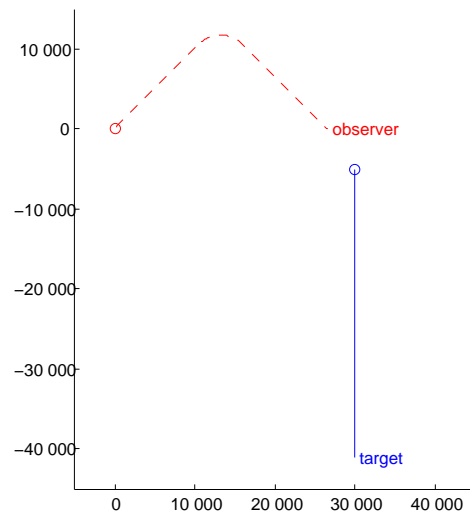


Fig. 1. Simulation scenario. Total observation duration 1 hour, sampling interval 4s. Initial relative distance 20025m, target speed 7m/s, observer speed 10m/s. Trajectories: target (plain line), maneuvering observer (dashed line), initial positions (o).

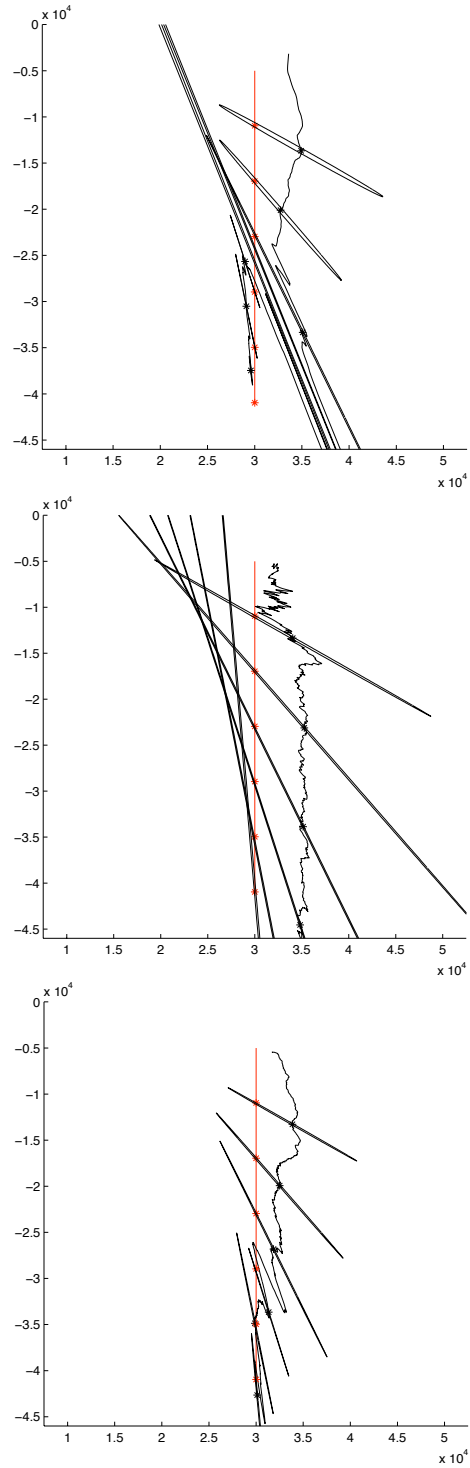


Fig. 2. Extended Kalman filter (top), particle filter (middle), and convolution particle filter (below) with  $\sigma = 1$  degree. The true trajectory (vertical line) with the empirical estimated trajectory after 20 Monte Carlo independent runs and the corresponding empirical uncertainty ellipses (every 10 minutes). The case of the modified polar coordinate system with low measurement noise is known to be favorable to the EKF (and unfavorable to the standard particle filter), nevertheless the CPF outperforms the EKF.



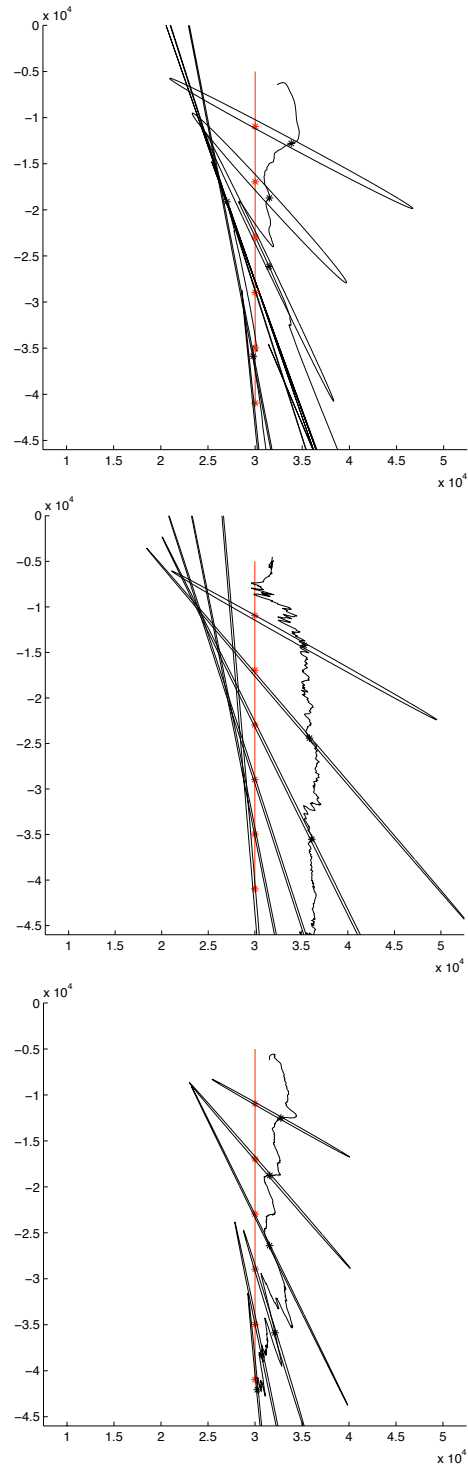


Fig. 3. Extended Kalman filter (top), particle filter (middle), and convolution particle filter (below) with  $\sigma = 1.5$  degree. The true trajectory (vertical line) with the empirical estimated trajectory after 20 Monte Carlo independent runs and the corresponding empirical uncertainty ellipses (every 10 minutes). Compared to Fig. 2, the performances of the EKF deteriorate much more than those of the CPF. After the maneuver, the EKF is unable to find the track.

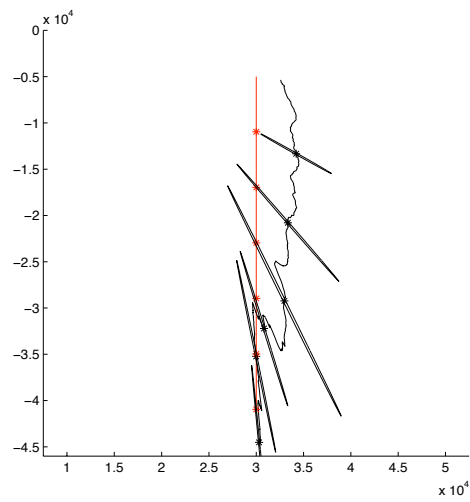


Fig. 4. Convolution particle filter in the cartesian coordinate system with  $\sigma = 1.5$  degree. The true trajectory (vertical line) with the empirical estimated trajectory after 20 Monte Carlo independent runs and the corresponding empirical uncertainty ellipses (every 10 minutes). Compared to Fig. 3, the performance of the CPF are slightly better than in the case of modified polar coordinates. The EKF (not plotted here) is known to be very sensitive to the choice of coordinates and to present severe problems of divergence. The CPF is therefore much less sensitive to the choice of coordinates.