A NEW SAMPLING METHOD IN PARTICLE FILTER

Qi Cheng and Pascal Bondon

CNRS UMR 8506, Université Paris-Sud
91192, Gif-sur-Yvette, France
phone: +33 1 69 85 17 31, fax: +33 1 69 85 17 65, email: bondon@lss.supelec.fr

ABSTRACT

This paper presents a new method to draw particles for the particle filter in the case of large state noise. The standard bootstrap filter draw particles randomly from the prior density which does not use the latest information of the observation. Some improvements consist in using extended Kalman filter or unscented Kalman filter to produce the importance distribution in order to move the particles from the domain of low likelihood to the domain of high likelihood by using the latest information of the observation. The performances of these methods vary with the structure of the models. We propose a modified bootstrap filter which uses a new method to draw the particles. Our method outperforms the bootstrap filter with the same computational complexity. The effectiveness of the proposed filter is demonstrated through numerical examples.

1. INTRODUCTION

Estimation of the hidden state from noisy observations is an important topic in control theory and signal processing. Many works have been devoted to this subject since the famous Kalman filter which is the optimal solution for a linear model with Gaussian noises. For a nonlinear model, the equations of the optimal (nonlinear) filter have been developed since the middle of the 1960s, but unfortunately the involved integrals are intractable. Many suboptimal nonlinear filters have been introduced in the literature, see e.g. [1], [16], [17], [20] and [4]. The extended Kalman filter (EKF) proposed by [10] and [6], the unscented Kalman filter (UKF) introduced by [11], and the particle filter (PF) are well-known nonlinear filters among others.

The simplest way to approximate a nonlinear state space model is to replace the state transition and the measurement equations by Taylor series expansions. The EKF uses first-order Taylor series expansions, and then uses the Kalman filter to estimate the state. Although the state estimation given by the EKF is biased, this method is still widely used because of its simpleness.

The UKF uses several so called sigma points to calculate recursively the means and covariances used in the Kalman filter, see [18]. At first glance, UKF and EKF are different. However, both filters use the Gaussian distribution to approximate the true posterior distribution. Essentially, EKF and UKF are two different implementations of the linear update equation in the Kalman filter. When the variance of the observation noise is small, the UKF provides generally a more accurate estimation of the state than the EKF. When the variance of the observation noise increases, the performances of the UKF and the EKF become similar.

Up to now, the most successful nonlinear filter is the PF which can be regarded as an approximation of a recursive Bayesian filter. The principle of the PF is to implement a recursive Bayesian filter by Monte Carlo simulations. The densities involved in the Bayesian filter are represented by a set of random samples with associated weights.

The PF has been used successfully in many domains such as guidance, signal and image processing and computer vision, but its performance depends heavily on the choice of the so called importance distribution (ID). Many works have been done to choose the ID, but no general rule seems to exist. The most popular choice is to use the transition prior function as the ID, see [8]. This method does not use the latest information of the observation. To overcome this problem, [5] and [19] suggested respectively to use the EKF and the UKF to produce the ID. These methods may provide a good approximation of the true posterior distribution, but they depend on the structure of the model. In this paper, a new method which uses likelihood to choose the particles from the ID is introduced and compared with the standard PF.

The remainder of this paper is organized as follows. The principle of the PF is introduced in Section 2. In Section 3, we present some parametric methods to produce the ID. In Section 4, a modified bootstrap filter (MBF) is proposed. Then, the effectiveness of our method is illustrated by numerical examples in Section 5. Finally, some conclusions are given in Section 6.

2. PARTICLE FILTER

Consider a dynamic nonlinear discrete time system described by a state-space model

$$x_t = f(x_{t-1}) + u_t, \tag{1}$$
$$y_t = h(x_t) + v_t, \tag{2}$$

where x_t is the hidden state, y_t is the observation, and u_t, v_t are the state and observation noises. Both noises are independent and identically distributed sequences and are mutually independent. When we write (1), we always assume implicitly that u_t is independent of $\{x_{t-k}, k \geq 1\}$. This condition is natural when the process (x_t) is generated from the model in the increasing time order. Then, x_t is a homogeneous Markov chain, i.e., the conditional probability density of x_t given the past states $x_{0:t-1} = (x_0, \ldots, x_{t-1})$ depends only on x_{t-1} through the transition density $p(x_t|x_{t-1})$, and the conditional probability density of y_t given the states $x_{0:t}$ and the past observations $y_{1:t-1}$ depends only on x_t through the conditional likelihood $p(y_t|x_t)$. We further assume that the initial state x_0 is distributed according to a density function $p(x_0)$.

The objective of filtering is to estimate the posterior density of the state given the past observations $p(x_t|y_{1:t})$. A recursive update of the posterior density as new observations arrive is given by the recursive Bayesian filter defined by

$$p(x_t|y_{1:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1}, \quad (3)$$

$$p(x_t|y_{1:t}) = \frac{p(y_t|x_t)p(x_t|y_{1:t-1})}{p(y_t|y_{1:t-1})}, \quad (4)$$

where the conditional density $p(y_t|y_{1:t-1})$ can be calculated by

$$p(y_t|y_{1:t-1}) = \int p(y_t|x_t)p(x_t|y_{1:t-1})dx_t.$$

The difficulty to implement the recursive Bayesian filter is that the integrals are intractable, except for a linear Gaussian system in which case the closed-form solution of the integral equations is the well known Kalman filter introduced by [12].

The PF uses Monte Carlo methods to calculate the integrals. The posterior density $p(x_{0:t}|y_{1:t})$ is represented by a set of N random samples $x^i_{0:t}$ (particles) drawn from $p(x_{0:t}|y_{1:t})$ with associated normalized positive weights ω^i_t ($\sum_i \omega^i_t = 1$). The posterior density is approximated by the discrete distribution, $\sum_{i=1}^N \omega^i_t \delta_{x^i_{0:t}}$, and the conditional expectation of any integrable function $g(x_{0:t})$ is approximated by the finite sum,

$$E[g(x_{1:t})|y_{1:t}] = \int g(x_{1:t})p(x_{1:t}|y_{1:t})dx_{1:t} \simeq \sum_{i=1}^N \omega^i_t g(x^i_{1:t}).$$

In general, it is difficult to sample directly from the full posterior density. To overcome this difficulty, the method of importance sampling is used, see e.g. [15]. The particles $x^i_{1:t}$ are drawn from an easy sampling ID $q(x_{1:t}|y_{1:t})$ and we define the non-normalized weights as

$$\omega^i_t = \frac{p(x^i_{1:t}|y_{1:t})}{q(x^i_{1:t}|y_{1:t})}.$$

The ID is chosen to factorize such that

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

in order that the weights can be updated sequentially as

$$\omega^i_t \propto \omega^i_{t-1} \frac{p(y_t|x^i_t)p(x^i_t|x^i_{t-1})}{q(x^i_t|x^i_{t-1}, y_t)}. \quad (5)$$

We can implement recursively a basic sequential importance sampling (SIS) PF in the following steps, see [2] :

1. Sample the particles $x^i_t \sim q(x_t|x^i_{t-1}, y_t)$;
2. Update the weights according to (5).

An important problem of the PF is the degeneracy problem. After a few iterations, only few particles have non negligible weights and the estimation may become unreliable. The sampling importance resampling filter has been developed by [8] to overcome this drawback. The objective of resampling is to eliminate samples with low importance weights and multiply samples with high importance weights. Several methods of resampling have been developed, such as the multinomial resampling, the residual resampling and the systematic resampling. In this paper, we use the residual resampling.

3. THE STATE-OF-ART TO PRODUCE THE IMPORTANCE DISTRIBUTION

Choosing the ID is one of the key problems in the PF. The optimal ID satisfies $q(x_t|x_{0:t-1}, y_{1:t}) = p(x_t|x_{t-1}, y_t)$ and fully exploits the information in both x_{t-1} and y_t, see e.g. [5]. In practice, this distribution is unknown for a general nonlinear model and therefore, it is impossible to sample from it. The second choice of ID is the transition density for its easiness to sample and leads to the standard bootstrap filter (BF). This is the most popular choice, but since $p(x_t|x_{t-1})$ does not use the latest information of y_t, the performance of the BF varies with the variance of the observation noise. When this variance is small or the variance of the state noise is large, the BF performs badly.

Many methods have been developed to produce an ID which uses the latest information of y_t. [5] have proposed to use the method of local linearization to generate the ID. Then, EKF, UKF and Gaussian-Hermite filters have been used by [9] to produce the ID. Since these filters use the latest information of y_t, the choice of local linearization method may improve the performance of the PF when the variance of the observation noise is small. In the same vein, a Gaussian particle filter was proposed by [13] which uses a Gaussian distribution to approximate the true posterior distribution.

The auxiliary particle filter (APF) is a popular variant of the standard PF. The essential of the APF is to reserve the most possible survival particles in the simulation. The APF uses the expectation $E[x_t|x^i_{t-1}]$ to characterize $p(x_t|x^i_{t-1})$. When the variance of the state noise is small, this method performs very well. When the variance of the state noise is large, the distribution $p(x_t|x^i_{t-1})$ could not be characterized by $E[x_t|x^i_{t-1}]$, and it was pointed out by [2] that the use of the APF can degrade the performance.

Other existing nonlinear filtering techniques can be used to produce the ID. For instance, [21] have used the adaptive nonlinear filter proposed by [14], and [22] have considered the nonlinear projection filter introduced by [3]. Of course, the better the nonlinear filtering techniques approximate the true posterior distribution of the state, the better these methods perform.

These methods may perform better than the BF, but of course they depend on the structure of the models. In the next section, a new method is introduced to draw particles in the PF.

4. A NEW SAMPLING METHOD FOR PARTICLE FILTER

To improve the performance of the PF, we may either draw more particles or use the information in y_t and let the particles move toward the region of high likelihood. This is the principle used by [19], [9], [21], and [22]. Nevertheless, these techniques are valuable only for some specific models. Up to now, there is no algorithm can outperform the BF in the general nonlinear non Gaussian models with large state noise. The local linearization PF works worse than the BF, because the posterior density of the state is quite different from the Gussian density. This is clear shown in our numerical examples later. The APF is only valid in the condition of small state noise. When the state noise is large, the prior function contains few information to predict the future state. The shape of the likelihood function is closer to the posterior function of the state. In this case, as pointed out by [19]: " It is therefore of paramount importance to move the particles towards the regions of high likelihood."

We propose a new technique that may be used for any nonlinear system described by the state-space model (1)–(2) with large state noise. Our method uses a two-stage sampling technique :

1. For $j = 1, \ldots, M$, draw $x_t^{i,j} \sim p(x_t|x_{t-1}^i)$ and compute the conditional likelihood $p(y_t|x_t^{i,j})$.
2. Select the particle $x_t^{i,j^\star}$ whose conditional likelihood is maximum and set $x_t^i = x_t^{i,j^\star}$.

In the first step, the particles move randomly according to the prior information like in the BF, and in the second step, the information y_t is used to select the particle with the maximum conditional likelihood. This new sampling schema uses M as a compromise parameter between the prior function and the likelihood. When $M = 1$, only the prior information is used, this is the conventional BF; when M increases, more observation information is used. The resulting algorithm is the following :

Remark 1. In the MBF, we select particles from the prior density with high likelihood. Essentially, this idea is the same as the resampling which chooses particles according to their likelihood.

Algorithm 1 Modified bootstrap filter(MBF)

```
Initialization, t = 0
for i = 0 to N do
  Draw particle x_0^i ~ p(x_0) and set t = 1
end for
for t = 1 to T do
  for i = 1 to N do
    for j = 1 to M do
      Draw particle x_t^{i,j} ~ p(x_t|x_{t-1}^i)
      Compute the conditional likelihood p(y_t|x_t^{i,j})
    end for
    Select x_t^{i,j*} such that p(y_t|x_t^{i,j*}) is maximum
    Set x_t^i = x_t^{i,j*}
  end for
  Resample particle from the x_t^i according to the
  weights ω_t^i
end for
```

Remark 2. In the MBF, the ID depends on M. It is difficult to get ID's analytical expression. It can be approximated by the prior function. The weight can be approximated by

$$\omega_t^i \approx \frac{p(y_t|x_t^i)p(x_t^i|x_{t-1}^i)}{p(x_t^i|x_{t-1}^i)} = p(y_t|x_t^i).$$

Remark 3. M is an integer and $M < \infty$. How to choose M is the key point in the MBF. M is chosen accordingly to the ratio of the state noise variance and the observation noise variance. M increases when this ratio increases.

Remark 4. The MBF is different from the APF which chooses the particles whose conditional expectation has high conditional likelihood. The MBF is used in the scenario of a large state noise, while the APF is used in the scenario of small state noise.

5. NUMERICAL EXAMPLES

It is difficult to compare the performances of nonlinear filters by a theoretical analysis since in general an explicit expression of the estimation error is not available. We compare the MBF with the BF, the PFEKF (PF with the EKF to produce the ID), the UPF (PF with the UKF to produce the ID), PFMCMC (PF with the Markov chain Monte Carlo resampling, see [7]) through numerical simulations. We show that the MBF using N particles is more efficient than the BF using N particles, but the implementation time is slightly longer. Now, for a same implementation time (which means using more than N particles in the BF), we show that the MBF still outperforms the BF.

We consider a nonlinear model which was used in

Filter (Sample size)	μ	σ^2	Time
BF (600)	3.51	0.98	2.94
BF (2000)	3.41	0.54	9.96
PFMCMC(600)	3.45	0.53	6.63
PFMCMC(2000)	3.35	0.39	21.91
PFEKF(600)	21.67	1.06	11.02
UPF(600)	21.72	1.06	19.68
MBF (600)	3.34	0.33	9.01

Table 1: Comparison of estimation result by different filters

[19] and is given by

$$x_t = -40 + \sin(w\pi(t-1)) + \frac{x_{t-1}}{2} + u_t,$$

$$y_t = \begin{cases} \frac{x_t^2}{5} + v_t, & t \leq 30, \\ \frac{x_t}{2} - 2 + v_t, & t > 30, \end{cases}$$

where $w = 4e-2$, u_t follows a $\Gamma(80, 0.5)$ distribution, and v_t follows a $N(0,1)$ distribution. We want to estimate the hidden states x_t for $t = 1, \ldots, T$. Let $\hat{x}_t = \frac{1}{N}\sum_{i=1}^{N} x_t^i$ be the estimation of x_t obtained by the N particles x_t^i after resampling. To measure the performance of estimation of the states x_t for $t = 1, \ldots, T$, we introduce the root mean-squared error

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(x_t - \hat{x}_t)^2}.$$

We calculate RMSE with $T = 60$ for both algorithms using different numbers of particles. We take $M = 3$ in the MBF. The experiment is repeated 100 times independently.

In Table 1, μ and σ^2 denote the estimated mean and variance of RMSE calculated over the 100 realisations. Time denotes the computation time of our Matlab code expressed in seconds and measured on a PC with a Pentium D at 3.40 GHz. We see that for a same number of particles (600), the MBF outperforms the BF in terms of mean and variance of the RMSE but the computation time is multiplied by 3. Now, when the BF uses 2000 particles and the MBF uses 600 particles, the MBF still outperforms the BF but using less time. This is the advantage of the MBF with respect to the BF. The PFEKF and the UPF provide worse performances than the basic BF. The PFMCMC could improve the performance of the basic BF but it uses more time. When using 600 particles in the PFMCMC, its performance is better than the BF but worse than the MBF, while it uses more time than the BF but less time than the MBF. When 2000 particles is used in the PFMCMC, it works still worse than the MBF using 600 particles, but it uses much more time than the MBF. We conclude that the MBF is the best filter when the same computing time is used.

When the same number of particles (600) is used, since the MBF exploits the observation information of y_t, it outperforms the BF. This is easy to understand. When 2000 particles are used in the BF, the performance of the MBF is still better then the BF. Because the MBF chooses the most possible survival particles which are different from the BF. Furthermore, these 600 particles of the MBF are better than the best 600 particles of the PF because they include the observation information.

In Figure 1, we show for one realisation the result of the state estimation x_t for $t = 1, \ldots, 60$ obtained after resampling, using 2000 particles in the BF and PFMCMC, 600 particles in the MBF. We see that the results are more precise with the MBF than with the BF and the PFMCMC . This is confirmed by Table 1.

In Figure 2, we show for the same realisation the result of the state estimation x_t for $t = 1, \ldots, 60$ obtained after resampling, using 600 particles in the MBF, the PFEKF and the UPF. It is clear that the PFEKF and the UPF provide very bad results of the estimation. It is because that the posterior density of the state is quite different from the Gaussian distribution. The MBF is better than both the PFEKF and UPF.

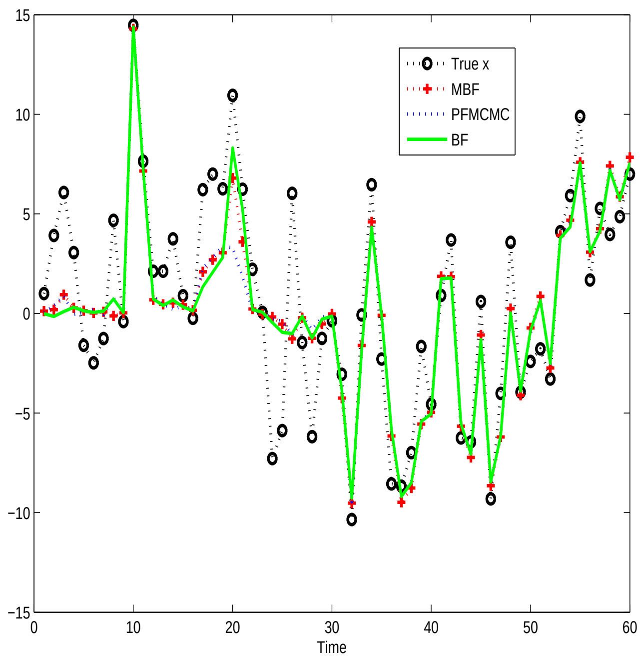

Figure 1: State estimation using 2000 particles in the BF and the PFMCMC, using 600 particles in the MBF

6. CONCLUSION

We have proposed a modification of the standard PF for nonlinear filtering. The idea is to select particles with high conditional likelihood. Our algorithm outperforms the PF with the same computational complexity when the state noise is large. In the future, it will be

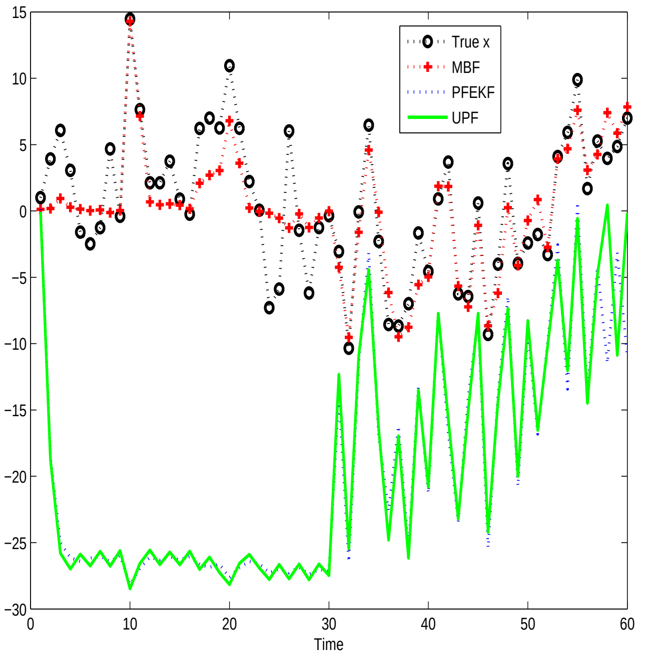

Figure 2: State estimation using using 600 particles in the MBF, the PFEKF and the UPF.

interesting to investigate how to choose the number M adaptively.

REFERENCES

[1] B. Anderson and J. Moore. *Optimal Filtering.* Prentice-Hall, Englewood Cliffs, NJ, 1979.

[2] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filter for on-line nonlinear/non-gaussian Bayesian tracking. *IEEE Trans. Signal Process.*, 50(2):174–188, 2002.

[3] R. Beard, J. Gunther, J. Lawton, and W. Stirling. Nonlinear projection filter based on galerkin approximation. *AIAA J. of Guidance, Control and Dynamics*, 22(2):258–266, 1999.

[4] A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo methods in practice.* Springer-Verlag, New York, 2001.

[5] A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10(3):197–208, 2000.

[6] A. Gelb, editor. *Applied optimal estimation.* The MIT Press, Cambridge, 1974.

[7] W. Gilks and C. Berzuini. Following a moving target - bayesian inference for dynamic bayesian models. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 63(1):127–146, 2001.

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

[9] D. Guo, X. Wang, and R. Chen. New sequential Monte Carlo methods for nonlinear dynamic systems. *Statistics and Computing*, 15(2):135–147, 2005.

[10] A. Jazwinski. *Stochastic Processes and Filtering Theory.* Academic Press, New York, 1970.

[11] S. Julier and J. Uhlmann. A new extension of the Kalman filter to nonlinear systems. In *Int. Symp. Aerospace/Defense Sensing, Simul. and Controls, Orlando, FL*, 1997.

[12] R. E. Kalman. A new approach to linear filtering and prediction problems. *Trans. ASME, Series D, J. Basic Eng.*, 82:35–45, 1960.

[13] J. H. Kotecha and Djurić. Gaussian particle filter. *IEEE Trans. Signal Process.*, 51(10):2592–2601, 2003.

[14] D. Lainiotis and P. Papaparaskeva. A new class of efficient adaptive nonlinear filters (anlf). *IEEE Trans. Signal Process.*, 46(6):1730–1737, 1998.

[15] C. P. Robert and G. Casella. *Monte Carlo statistical methods.* Springer-Verlag, New York, second edition, 2004.

[16] H. W. Sorenson, editor. *Kalman filtering: theory and application.* IEEE Press, New York, 1985.

[17] J. C. Spall, editor. *Bayesian Analysis of Time Series and Dynamic Models.* Marcel Dekker, New York, 1988.

[18] R. Van der Merwe. *Sigma-Point Kalman Filters for Probabilistic Inference in Dynamic State-Space Models.* PhD thesis, Oregon Health and Science University, 2004.

[19] R. Van der Merwe, N. de Freitas, A. Doucet, and E. Wan. The unscented particle filter. In *Advances in Neural Information Processing Systems*, Nov 2000.

[20] M. West and J. Harrison. *Bayesian Forecasting and Dynamic Models.* Springer-Verlag, New York, second edition, 1997.

[21] Y. Zhai and M. Yeary. A novel nonlinear state estimation technique based on sequential importance sampling and parallel filter banks. In *Proceedings of the Conference on Control Applications*, Toronto, Canada, 2005. IEEE.

[22] Y. Zhai, M. Yeary, and D. Zhou. Target tracking using a particle filter based on the projection method. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Honolulu, Hawaï, 2007. IEEE.