SAMPLE-EFFICIENCY-OPTIMIZED AUXILIARY PARTICLE FILTER

Feng Guo and Gang Qian

Arts, Media and Engineering Program and
Department of Electrical Engineering
Arizona State University
Tempe, AZ, 85281, USA

ABSTRACT

In this paper, we present a sample-efficiency-optimized auxiliary particle filter. We show that by computing the weights of the auxiliary variable through a simple optimization procedure, the variance of the resulting importance weights can be considerably reduced, in fact, minimized. In turn, the effective sample size and therefore the efficiency of the overall sampling both increase. Experiments have been extensively conducted using the bearings-only model. The experimental results fully support the theoretical conclusions that the proposed sample-efficiency-optimized auxiliary particle filter outperforms the nonoptimized auxiliary particle filtering, producing smaller tracking errors and larger effective sample sizes.

1. INTRODUCTION

Particle filters (or sequential Monte Carlo methods) [4] methods have now become very poplar while powerful tools for smoothing, filtering and prediction of the states of dynamic system involving non-linearity and non-Gaussianity. Particle filters utilize weighted samples, also called particles (both will be used interchangeably without further notice in this paper) drawn from the system state space to represent the posterior probability distributions of the states.

In particle filters, Bayesian sequential importance sampling (SIS)is often used to propagate weighted samples from current time instant to the next. In SIS, extra care needs to taken to combat against sample degeneracy, in which case, most of the samples, except one carry zero importance weight. Degeneracy prevent the weighted samples from presenting a complete picture of the solution space; moreover, it wastes a large portion of the computational resources. Degeneracy is often measured by the variance of importance weight. When all the samples are drawn from the target distribution, all the samples are equally weighted, there is no degeneracy and the importance weight variance will be zero. On the other extreme point, when all but one sample carry zero weight, there is severe degeneracy and the importance weight variance is large. To alleviate degeneracy, the importance-sampling-with-resampling (SIR) or similar approaches such as rejection methods and Hastings independence chain approach [9] has been introduced to draw samples for the current time instant based on the fitness of the samples to all available observations. Therefore, good samples which can interpret samples well with small prediction errors will be heavily sampled. Meanwhile, favorable proposal distributions, or also called importance functions, (both will be used interchangeably without further notice in this paper) are selected such that the latest measurements can be used to guide the sampling. Let x_t denote the state to be estimated at current time instant and $z_{1:t}$ be the measurements from time 1 to t. A favorable choice of the proposal distribution is given by $p(x_t|x_{1:t-1}, z_{1:t})$, since it minimizes the variance of the importance weight conditioned upon $x_{1:t-1}$ and $z_{1:t}$ [5].

A number of sampling strategies have been proposed to achieve SIR with the aforementioned proposal distribution. Liu and Chen [9] have extended existing methods and presented a generalized framework for simultaneously sample drawing and weight evaluation, referred as ”local Monte Carlo method” for SIS. The auxiliary particle filter (APF)[10] is one of the proposed frameworks and was considered as one way to perform local Monte Carlo in [9]. By introducing an auxiliary variable k, APF allows the sampling procedure to probe current solution space, considering the current measurement. Comparing to the SIR filter, APF can generate better estimates of the posterior whenever the likelihood is situated in one of the priors tails. In this paper, the variance of importance weights in APF is analyzed and expressed as a function of the auxiliary variable weights. Then the optimal auxiliary variable weights are selected to minimize the importance weights variance. The substantial improvement in bearings-only tracking application is shown compared with SIR and APF methods.

1.1. Existing Improvement upon APF

A number of methods have been proposed to improve APF. In [1], unscented approximation was used to compute auxiliary index variables so that high computation requirement is avoided. In [7], a hybrid importance function was proposed to employ the advantage of the posterior and the prior importance functions. The hybrid importance function can be combined with APF when the dynamic noise was small.

The auxiliary index variables are a linear combination of previous mean estimation and samples. In [3], local linearization techniques are adopted to approximate an optimal importance function. The auxiliary index variables are generated according to this optimal importance function. But all these methods can not guarantee that the variance of the weights has the minimum value.

2. REVIEW OF AUXILIARY PARTICLE FILTER

Let x_t be the state to be estimated at the current time instant and $x_{1:t}, z_{1:t}$ be the state sequence and measurements sequence from the initial time instant to current time instant. Assume that a first order state space model of a dynamic system is used. The posterior density of the state sequence can be approximated by a set of properly weighted samples [9] and their weights $\{x_{1:t}^{(k)}, w_t^{(k)}\}_{k=1}^{M}$, where the normalized importance weight $w_t^{(k)}$ is recursively updated according to

$$w_t^{(k)} \propto w_{t-1}^{(k)} \frac{p(z_t|x_t)p(x_t|x_{t-1}^{(k)})}{q(x_t|x_{t-1}^{(k)}, z_t)} \tag{1}$$

$\{x_t^{(k)}\}_{k=1}^{N}$ is the new sample set, drawn from the proposal distribution $q(x_t|x_{t-1}, z_t)$.

The auxiliary particle filter (APF) with sampling/importance resampling is introduced by Pitt and Shephard [10] as a variant of the standard SIR filter to improve sample efficiency. The branch index K is introduced as an auxiliary variable. The target joint probability density function is thus given by

$$p(x_t, K = k|z_{1,:t}) \propto w_{t-1}^{(k)} p(x_t|x_{t-1}^{(k)}) p(z_t|x_t) \tag{2}$$

In practice, the joint samples are drawn from the following importance function:

$$q(x_t, K = k|z_{1:t}) \propto w_{t-1}^{(k)} p(x_t|x_{t-1}^{(k)}) p(z_t|\mu_t^{(k)}) \tag{3}$$

where $\mu_t^{(k)}, k = 1, \ldots, N$ could be the mean $\mu_t^{(k)} = E[x_t|x_{t-1}^{(k)}]$, the mode $\mu_t^{(k)} = \mathit{arg}_{x_t}\mathit{max}\{p(x_t|x_{t-1}^{(k)})\}$, a random probe $\mu_t^{(k)} \sim p(x_t|x_{t-1}^{(k)})$ or some other statistics of the the prior density.

The sampling can thus be performed in two steps. First, draw branch index K according to the marginal auxiliary variable weights

$$\hat{w}_k = q(K = k|z_{1:t}) \propto \lambda_k w_{t-1}^{(k)} \tag{4}$$

where $\lambda_k = p(z_t|\mu_t^{(k)})$ and $\sum_{k=1}^{M} \hat{w}_k = 1$. Then draw x_t from $p(x_t|x_{t-1}^{(k)})$, the prior density based on the branch indices k. Let $k_j = K^{(j)}$. The joint sample set is $\{x_t^{(j)}, K^{(j)}\}_{j=1}^{M}$ and the corresponding weights are computed using (1),

$$w_t^{(j)} \propto w_{t-1}^{(k_j)} \frac{p(z_t|x_t)p(x_t|x_{t-1}^{(k_j)})}{q(x_t, k_j|z_t)} = \frac{p(z_t|x_t^{(j)})}{\lambda_{k_j}} \tag{5}$$

where $j = 1, \ldots, M$ and $k_j = K^{(j)}$ denotes index of particle j's parent. Note that the joint sample set $\{x_t^{(j)}, K^{(j)}\}_{j=1}^{M}$ is equivalent to $\{x_{1:t}^{(j)}\}_{j=1}^{M}$ when the all of the branching information is stored.

3. SAMPLE-EFFICIENCY-OPTIMIZED AUXILIARY PARTICLE FILTER

APF naturally generates particles from the samples conditioned on the current measurement, associated with large predicative likelihoods. In [10], it was showed that the variance of the importance weights in APF is smaller than that of SIR in some situations. It is a common practice to evaluate the auxiliary variable weight using (4), by looking at the likelihood of $\mu_t^{(k)}$, which could be the mean, mode, a draw of any other statistics of prior density $p(x_t|x_{t-1}^{(k)})$. However, there is little attention paid to find out guidelines for computing λ_k, leading to the optimal performance, namely, minimized variances for the importance weights of $x_t^{(j)}$. In this section, we show that by careful selection of λ_k through a simple optimization procedure, the variances of the standardized importance weights of $x_t^{(j)}$ can be minimized.

Based on (5), in APF, the true importance weights with respect to the posterior distribution $p(x_{1:t}^{(k)}|z_{1:t})$ is given by

$$w_a^*(x_t^{(k)}) = \frac{p(z_{1:t-1})}{p(z_{1:t})} \frac{p(z_t|x_t)}{\lambda_k} \tag{6}$$

Consider this importance weight as a random variable, which is a function of x_t. It is easy to verify that

$$E\{w_a^*(x_t)\} = 1 \tag{7}$$

which is required for the true importance weights, instead of a scaled one. The variance of the importance weight is given by

$$\begin{aligned} var\{w_a^*(x_t)\} &= E\{[w_a^*(x_t)]^2\} - E^2\{w_a^*(x_t)\} &\quad (8)\\ &= E\{[w_a^*(x_t)]^2\} - 1 &\quad (9) \end{aligned}$$

Let $\alpha^2 = \frac{p^2(z_{1:t-1})}{p^2(z_{1:t})}$, then

$$\begin{aligned} E\{[w_a^*(x_t)]^2\} &= \alpha^2 \sum_{k=1}^{M} \int \frac{p^2(z_t|x_t)}{\lambda_k^2} \hat{w}_k p(x_t|x_{t-1}^{(k)}) dx_t \\ &= \alpha^2 \sum_{k=1}^{M} \frac{w_{t-1}^{(k)}}{\lambda_k} \int p^2(z_t|x_t) p(x_t|x_{t-1}^{(k)}) dx_t \\ &= \alpha^2 \sum_{k=1}^{M} \frac{w_{t-1}^{(k)}}{\lambda_k} g_k &\quad (10) \end{aligned}$$

where $g_k \equiv \int p^2(z_t|x_t)p(x_t|x_{t-1}^{(k)})dx_t$. To optimize the sample efficiency of the auxiliary particle filter, one needs to find $\lambda_k, k = 1, \ldots, M$ such that

$$\frac{1}{\alpha^2}E\{[w_a^*(x_t)]^2\} = \sum_{k=1}^{M} \frac{w_{t-1}^{(k)}}{\lambda_k} g_k \tag{11}$$

is minimized subject to

$$\sum_{k=1}^{M} \hat{w}_k = \sum_{k=1}^{M} \lambda_k w_{t-1}^{(k)} = 1 \tag{12}$$

so that the auxiliary variable weights are a mass function of a discrete random variable. This minimization problem can be solved using Lagrangian multiplier. Let C be the augmented cost function,

$$C = \sum_{k=1}^{M} \frac{w_{t-1}^{(k)}}{\lambda_k} g_k + \beta(\sum_{k=1}^{M} \lambda_k w_{t-1}^{(k)} - 1) \tag{13}$$

Hence,

$$\frac{\partial C}{\partial \lambda_k} = -\frac{g_k w_{t-1}^{(k)}}{(\lambda_k)^2} + \beta w_{t-1}^{(k)} = 0 \tag{14}$$

Therefore,

$$\lambda_k = \sqrt{\frac{g_k}{\beta}} \tag{15}$$

Using the constraint (12),

$$\sum_{k=1}^{M} \sqrt{\frac{g_k}{\beta}} w_{t-1}^{(k)} = 1 \tag{16}$$

Hence

$$\sqrt{\beta} = \sum_{k=1}^{M} w_{t-1}^{(k)} \sqrt{g_k} \tag{17}$$

Substitute (17) into (18), it follows that

$$\lambda_k = \frac{\sqrt{g_k}}{\sum_{k=1}^{M} w_{t-1}^{(k)} \sqrt{g_k}} \tag{18}$$

Therefore, by selecting auxiliary variable weights $\hat{w}_k = \lambda_k w_{t-1}^{(k)}$ with λ_k obtained using (18), the variance of the importance weight can be minimized. In summary, the procedures is as follows.

Sample-Efficiency-Optimized Auxiliary Particle Filter

At time $t = 0$,Initialization

- For $k = 1, \ldots, N$, sample $x_0^{(i)} \sim p(x_0)$ and set $t = 1$.

At time $t \geq 1$, Auxiliary variable resampling step

- For each branch k,$k = 1, \ldots, M$,
  compute $g_k = \int p^2(z_t|x_t)p(x_t|x_{t-1}^{(k)})dx_t$ using the empirical expectation of $p^2(z_t|x_t)$,namely,
  $g_k \approx \frac{1}{L}\sum_{i=1}^{L} p^2(z_t|x_t^{(i)})$,
  where $x_t^{(i)}$ is drawn from $p(x_t|x_{t-1}^{(k)})$.
- Compute auxiliary variable weights for each branch k, $k = 1, \ldots, M$,
  - compute $\lambda_k = \frac{\sqrt{g_k}}{\sum_{k=1}^{M} w_{t-1}^{(k)} \sqrt{g_k}}$.
  - compute $\hat{w}_k = p(K = k|z_{1:t}) = \lambda_k w_{t-1}^{(k)}$.
- Draw new sample for time t. For $j = 1, \ldots, N$,
  - Draw branch index $K = k_j$ according to auxiliary variable weights $\{\hat{w}_k\}_{k=1}^{M}$.
  - Draw $x_t^{(j)} = p(x_t|x_{t-1}^{(k_j)})$
  - Assign weight

$$w_t^{(k)} = \frac{p(z_t|x_t^{(k_j)})}{\lambda_{k_j}} \tag{19}$$

4. EXPERIMENTAL RESULTS

Experiments are conducted based on the ship tracking example described in [6, 10], using a bearings-only model with only angular information as observations. Results obtained using the original SIR, APF and sample-efficiency-optimized APF (SEO-APF) are compared.

The dynamic model of the system is given by

$$\mathbf{X}_{t+1} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \mathbf{X}_t + \begin{pmatrix} 0.5 & 0 \\ 1 & 0 \\ 0 & 0.5 \\ 0 & 1 \end{pmatrix} \mathbf{W}_t \tag{20}$$

where $\mathbf{X}_t = [x_t, \dot{x}_t, y_t, \dot{y}_t]^T$ is the state vector: x_t and y_t denote horizonal and vertical coordinates of the ship at time t, with corresponding velocities $\dot{x}_t$ and $\dot{y}_t$. $\mathbf{W}_t = [w_x, w_y]^T \sim N(0, \sigma_a^2\mathbf{I})$ is a zero mean Gaussian white noise. The observation model is

$$z_t = \mu_t + v_t \tag{21}$$

with $\mu_t = tan^{-1}(y_t/x_t)$. v_t is the observation noise. Here noise distribution is a wrapped Cauchy given by

$$p(z_t|\mu_t) = \frac{1}{2\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho\cos(z_t - \mu_t)} \tag{22}$$
$$(0 \leq z_t \leq 2\pi, \quad 0 \leq \rho < 1)$$

The state model was simulated using the following parameters: $\sigma_a = 0.001$, $\rho = 1 - 0.02^2$. The initial starting vector is obtained by sampling a Gaussian with the mean $[-0.05, 0.001, 0.2, -0.055]^T$ and a diagonal covariance matrix Σ_1 with diagonal terms given by $0.01 \times (0.5^2, 0.005^2, 0.3^2, 0.01^2)$.

Figure 1 illustrates both the ground-truth trajectory (solid curve) and results obtained using SIR (circles), APF (squares) and the proposed SEO-APF (dots). In this simulation, 300 particles were used to track the ship over 20 frames.

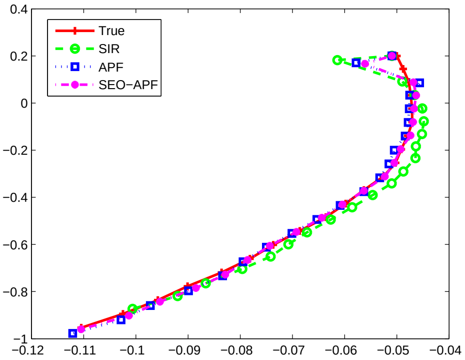


Fig. 1. Ground-truth trajectory (solid curve) and the tracking results using SIR (circles), APF (squares) and the proposed SEO-APF (dots).

In order to evaluate the proposed method quantitatively, the mean square error(MSE) was computed over a number of runs. By a run, we mean the process of simulating a ground-truth trajectories over 20 frames using the dynamic model and parameters described above and then applying the particle filtering a number of times, say, 20, to obtain the tracking results. Hence, a total of 400 simulations using 20 sets of ground-truth trajectories were conducted. In this part of experiment, 2000 particles were deployed.

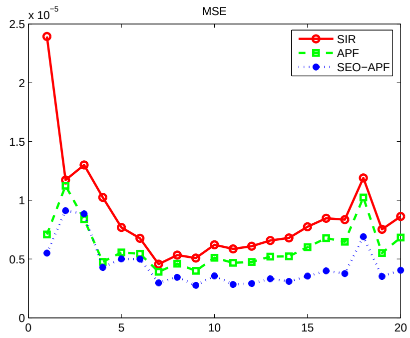
The MSE is then computed as follows:

$$MSE_{t,m} = \frac{1}{N_{mc}} \sum_{s=1}^{N_{mc}} \frac{1}{K} \sum_{k=1}^{K} (\mathbf{X}_{t,m,s}^{(k)} - \mathbf{X}_{t,m,s}^{(true)})^2 \quad (23)$$

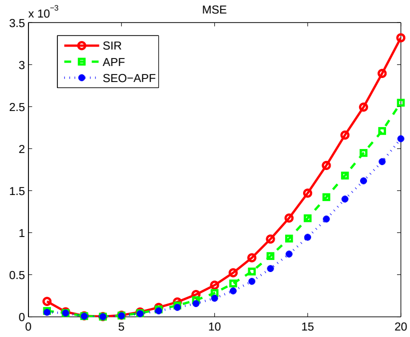
where $\mathbf{X}_{t,m,s}^{(k)}$ is the particle mean for run s,simulation k, state component m, at time t. $\mathbf{X}_{t,m,s}^{(true)}$ is the true value for run s, state component m, at time t.

All the three versions of particle filters were applied and the corresponding MSE were computed. Figure 2 shows the MSE of different state component using the same legend aforementioned. It can be seen that the proposed SEO-APF considerably outperformed the both the SIR and the nonoptimized APF.

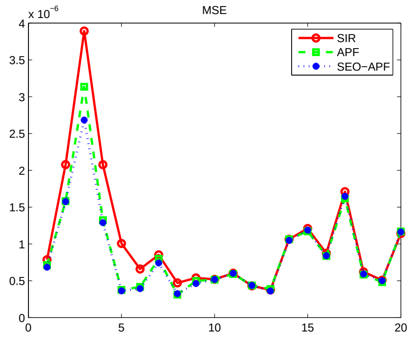
The effective sample size(ESS)[8] is used to measure and



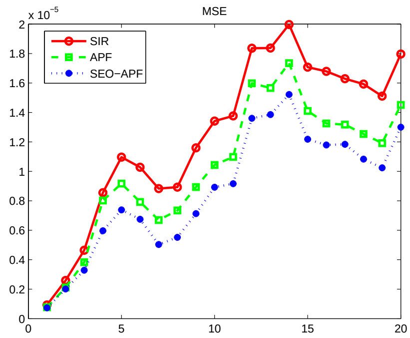
(a) x coordinate



(b) y coordinate



(c) x velocity



(d) y velocity

Fig. 2. MSE obtained using SIR (circles), APF (squares) and the proposed SEO-APF (dots).

compare the sample efficiency:

$$ESS = \frac{N}{1 + var(w_a^*(x_t))} \tag{24}$$

where N and $w_a^*(x_t)$ are sample numbers and true importance weight. An estimate $\widehat{ESS}$ of ESS is given by [5]:

$$\widehat{ESS} = \frac{1}{\sum_{k=1}^{N} (\tilde{w}_t^{(k)})^2} \tag{25}$$

where $\tilde{w}_t^{(k)}$ are the normalized importance weights. Figure 3 shows the ESS related to the three approaches using results from one of the 400 simulations. It can be seen that the SEO-APF has a much larger effective sample size than SIR and the nonoptimized APF. Similar ESS plots are obtained using other simulation results. Roughly speaking, in over 95of the time through out the entire simulations (20 frames × 20 simulations × 20 runs), AEO-APF has larger ESS than the others.

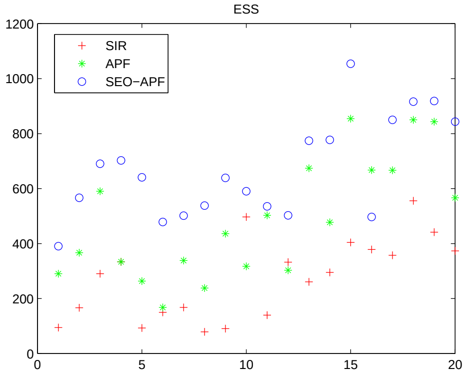


Fig. 3. Effective sample size corresponding to SIR (cross), APF (star) and SEO-APF (circle)

Table 1. Computation time (second)

SIR	APF (One Draw)	APF (Mean)	APF (Mode)	Optimized APF
0.1648	0.1535	1.3461	1.1617	1.1438

Regarding computational cost, the SEO-APF is much higher than the SIR, which is expected. The computational cost of SEO-APF is similar to that of APF, except the case when a single probe is used. Table 1 shows average time used to process one frame using 2000 particles for different approaches. For APF, we implemented a draw, mean and mode as μ_k respectively. Generally speaking, the proposed SEO-APF is about 10 times slower than SIR. If either mean or mode is selected as μ_k in APF, the computation time is almost the same for APF and SEO-APF.

5. CONCLUSIONS

In this paper, we presented a sample-efficiency-optimized auxiliary particle filter. Different from the original auxiliary particle filter, we proposed a optimal way to compute the auxiliary variable weights to minimize the true importance weight variance. Simulation results shows that the proposed SEO-APF considerably improves the tracking results and the efficiency of the particle filters.

6. REFERENCES

[1] C.Andrieu, M.Davy and A.Doucet,“Improved Auxiliary Particle Filtering: Applications to Time-varying Spectral Analysis,“ *Proceedings of IEEE SSP*,2001

[2] M.S.Arulampalam, S.Maskell,N.Gordon and T.Clapp,“A Tutorial on Particle Filters for Online Nonlinear/Non-Guassian Bayesian Tracking,“*IEEE Trans.Signal Processing*,Vol.50,No.2,2002

[3] M.G.S.Bruno and A.Pavlov,“Improved Particle Filters for Ballistic Target Tracking,“*Proceedings of IEEE ICASSP*,2004

[4] A.Douct, J.de.Freitas, and N.Gordon, Eds.,“Sequential Monte-Carlo Methods in Practice,“ New York: Springer-Verlag, 2001

[5] A.Doucet, S.Godsill and C.Andrieu,“On Sequential Monte Carlo Sampling Methods for Bayesian Filtering,“*Statistics and Computing*,Vol.10,page 197-208,2000

[6] N.J.Gordon, D.J.Salmond and A.F.M.Smith,“Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation,“ *IEE-Procedings-F*, Vol.140,no.2,pp.107-113,1993

[7] Y.Huang and P.M.Djuric,“A Hybird Importance Function for Particle Filtering,“ *IEEE Signal Processing Letter*,Vol.11,no.3,2004

[8] A.Kong, J.S.Liu and W.H.Wong,“Sequential Imputations and Bayesian Missing Data Problems,“ *JJournal of the American Statistical Association*,Vol.89, page 278-288, 1994

[9] J.S.Liu and R.Chen, “Sequential Monte Carlo Methods for Dynamic Systems,“ *Journal of the American Statistical Association*,Vol.93,Page 1032-1044, 1998

[10] M.K.Pitt and N.Shephard,“Filtering via Simulation: Auxiliary Particle Filters,“ *Journal of the American Statistical Association*, Vol.94, pp.590-599, 1999