Piotr Kozierski*

POPULATION MONTE CARLO AND ADAPTIVE IMPORTANCE SAMPLING IN PARTICLE FILTER

Keywords: population Monte Carlo, adaptive importance sampling, particle filtering

1. INTRODUCTION

The main advantage of particle filters is their versatility, because they can be used even for cases, where all other methods fail. The three different methods have been presented in this paper to show, what is the impact of parameter values on estimation quality.

In the next Section the particle filter principle of operation has been described. In the third and fourth Sections the Population Monte Carlo and Adaptive Importance Sampling algorithms have been presented. The obtained results have been shown in fifth Section.

2. PARTICLE FILTER

Particle filter (PF) principle of operation is based on the Bayes filtration [2]

$$\overbrace{p\left(\mathbf{x}^{(k)}|\mathbf{Z}^{(k)}\right)}^{\text{posterior}} = \frac{\overbrace{p\left(\mathbf{z}^{(k)}|\mathbf{x}^{(k)}\right)}^{\text{likelihood}} \cdot \overbrace{p\left(\mathbf{x}^{(k)}|\mathbf{Z}^{(k-1)}\right)}^{\text{prior}}}{\underbrace{p\left(\mathbf{y}^{(k)}|\mathbf{Z}^{(k-1)}\right)}_{\text{evidence}}} , \tag{1}$$

where $\mathbf{x}^{(k)}$ is a state vector in k-th time step, $\mathbf{z}^{(k)}$ is a measurement vector in k-th time step, and $\mathbf{Z}^{(k)}$ is a set of all measurement vectors from first to k-th time step

$$\mathbf{Z}^{(k)} = \left\{\mathbf{z}^{(1)} \;\; \mathbf{z}^{(2)} \;\; \ldots \;\; \mathbf{z}^{(k)}\right\} . \tag{2}$$

Both vectors, $\mathbf{x}^{(k)}$ and $\mathbf{z}^{(k)}$, should be considered as the representations of the random variables X and Z, with specified probability density functions (PDFs) [7]:

$$\mathbf{x}^{(k)} \sim p(\mathbf{x}^{(k)}|\mathbf{x}^{(k-1)}) , \tag{3}$$

$$\mathbf{z}^{(k)} \sim p(\mathbf{z}^{(k)}|\mathbf{x}^{(k)}) , \tag{4}$$

where the first conditional PDF is the transition model, and the second – measurement model.

*Poznań University of Technology, Faculty of Electrical Engineering, Institute of Control and Information Engineering, Piotrowo 3a street, 60-965 Poznań, e-mail: piotr.kozierski@gmail.com

PF is one of the possible implementations of the Bayes filter, in which the posterior PDF is modelled not by the continuous function, but by a set of particles. It is assumed that each i-th particle contains some value $\mathbf{x}^{i,(k)}$ (state vector) and a weight $q^{i,(k)}$. Particles with higher weights have higher chance that their state vector (value) is close to the real state vector. Hence, the posterior PDF estimated by PF can be written by

$$\hat{p}(\mathbf{x}^{(k)}|\mathbf{Z}^{(k)}) = \sum_{i=1}^{N} q^{i,(k)} \delta_D \left(\mathbf{x}^{(k)} - \mathbf{x}^{i,(k)} \right) , \tag{5}$$

and, based on the strong law of large numbers, one can say that information contained in estimated posterior PDF, when a number of particles tends to infinity, is the same as in the continuous PDF

$$p(\mathbf{x}^{(k)}|\mathbf{Z}^{(k)}) \overset{N \to \infty}{=} \hat{p}(\mathbf{x}^{(k)}|\mathbf{Z}^{(k)}) . \tag{6}$$

The operation principle of a general particle filter has been presented in Algorithm 1.

Algorithm 1 – Generic Particle Filter

1. Initialization. Draw N particles from initial PDF $\mathbf{x}^{i,(0)} \sim p(\mathbf{x}^{(0)})$, set initial values of particle weights $q^{i,(0)} = \frac{1}{N}$, set the time step $k = 1$.

2. Prediction. Draw N new particles from the importance function $\mathbf{x}^{i,(k)} \sim g(\mathbf{x}^{(k)}|\mathbf{x}^{i,(k-1)}, \mathbf{z}^{(k)})$.

3. Update. Calculate particle weights based on the measurement model, the transition model and the importance function

$$q^{i,(k)} \propto q^{i,(k-1)} \frac{p(\mathbf{z}^{(k)}|\mathbf{x}^{i,(k)}) p(\mathbf{x}^{i,(k)}|\mathbf{x}^{i,(k-1)})}{g(\mathbf{x}^{i,(k)}|\mathbf{x}^{i,(k-1)}, \mathbf{z}^{(k)})} . \tag{7}$$

4. Normalization. Normalize weights so that their sum is equal to 1.

5. Check the condition for resampling. If not met, go to the step 7.

6. Resampling. Draw N new particles based on the posterior PDF obtained in the previous steps, set new values of particle weights $q^{i,(k)} = \frac{1}{N}$.

7. Calculate the estimated value, increase the time step $k = k + 1$, go to the step 2.

The first PF was proposed by Gordon, Salmond and Smith [5] in 1993, and was called Bootstrap filter. It is the specific case of Algorithm 1, in which 2 assumptions have been made, i.e. the resampling is performed in each time step, and the importance function is equal to the transition model $g(\mathbf{x}^{(k)}|\mathbf{x}^{(k-1)}, \mathbf{z}^{(k)}) = p(\mathbf{x}^{(k)}|\mathbf{x}^{(k-1)})$. Thanks to this, equation (7) can be simplified to the form presented in Algorithm 2.

Algorithm 2 – Bootstrap Filter

1. Initialization. Draw N particles from the initial PDF $\mathbf{x}^{i,(0)} \sim p(\mathbf{x}^{(0)})$, set the time step $k = 1$.

2. Prediction. Draw N new particles from the transition model $\mathbf{x}^{i,(k)} \sim p(\mathbf{x}^{(k)}|\mathbf{x}^{i,(k-1)})$.

3. Update. Calculate particle weights based on the measurement model

$$q^{i,(k)} \propto p(\mathbf{z}^{(k)}|\mathbf{x}^{i,(k)}) \; . \tag{8}$$

4. Normalization. Normalize weights so that their sum is equal to 1.

5. Resampling. Draw N new particles based on the posterior PDF obtained in the previous steps.

6. Calculate the estimated value, increase the time step $k = k + 1$, go to the step 2.

The resampling step can be compared to the part of evolutionary algorithm, because particles with higher weights have higher chance for replication, whereas particles with lower weights probably will not be copied to the next time step [9].

The systematic resampling has been chosen by the author, and has been described below.

Algorithm 3 – Systematic Resampling

1. Initialization. Set $j = 1$ and $S = q^{1,(k)}$; draw one random value from uniform distribution $u \sim \mathcal{U}\left[0, \frac{1}{N}\right)$

2. For $i = 1..N$ do

    3. While $S < u$ do

        4. $j = j + 1$

        5. $S = S + q^{j,(k)}$

    6. Choose value $\mathbf{x}^{j,(k)}$ for replication

    7. $u = u + \frac{1}{N}$

8. Set chosen particles as a current set

9. Set new particle weights $q^{i,(k)} = \frac{1}{N}$

There are many others resampling methods, which have been described and compared in [6].

For more information about particle filtering, references [1, 2, 4] are recommended.

3. POPULATION MONTE CARLO

The method presented in Algorithm 2 is very easy to implement, however is not optimal, because the information about measurements is not taken into account. Hence the optimal choice of the importance function is [4]

$$g(\mathbf{x}^{(k)}|\mathbf{x}^{(k-1)},\mathbf{z}^{(k)}) \propto p(\mathbf{z}^{(k)}|\mathbf{x}^{(k)})p(\mathbf{x}^{(k)}|\mathbf{x}^{(k-1)}) \,. \tag{9}$$

However, presented PDF can be analytically calculated only for specific and relatively easy cases. In general, the importance function (9) can not be calculated. There are few methods to draw particles from PDF similar to the optimal one, and Population Monte Carlo (PMC) is one of them.

The main assumption of PMC is that different particles can be drawn from different importance functions [3]. This causes that there may be many different PMC algorithms. The method proposed by the author has been presented below.

Algorithm 4 – Population Monte Carlo

1. For $p = 1..P_n$ perform steps 2–13
   2. If $p = 1$ then
      3. $\boldsymbol{\mu}_{[0]} = \hat{\mathbf{x}}^{(k-1)}, \quad \boldsymbol{\sigma}_{[0]} = \boldsymbol{\sigma}_{\mathbf{x}}$
   4. Else
      5. $\boldsymbol{\mu}_{[p-1]} = \sum_{i=1}^{N/P_n} q_{[p-1]}^{i,(k)} \cdot \mathbf{x}_{[p-1]}^{i,(k)}$
      6. $\boldsymbol{\Sigma}_{[p-1]} = \sum_{i=1}^{N/P_n} q_{[p-1]}^{i,(k)} \cdot \left(\mathbf{x}_{[p-1]}^{i,(k)} - \boldsymbol{\mu}_{[p-1]}\right)^2$
      7. $\boldsymbol{\sigma}_{[p-1]} = \sqrt{\boldsymbol{\Sigma}_{[p-1]}}$
      8. If $\boldsymbol{\sigma}_{[p-1]} < \frac{\boldsymbol{\sigma}_{[p-2]}}{T_d}$
         9. $\boldsymbol{\sigma}_{[p-1]} = \frac{\boldsymbol{\sigma}_{[p-2]}}{T_d}$
   10. For $i = 1..\frac{N}{P_n}$ perform steps 11–13
      11. $\mathbf{x}_{[p]}^{i,(k)} \sim g\left(\mathbf{x}_{[p]}^{(k)}|\boldsymbol{\mu}_{[p-1]},\boldsymbol{\sigma}_{[p-1]}^2\right) = \mathcal{N}\left(\boldsymbol{\mu}_{[p-1]},\boldsymbol{\sigma}_{[p-1]}^2\right)$
      12. $q_{[p]}^{i,(k)} = \frac{p(\mathbf{z}^{(k)}|\mathbf{x}_{[p]}^{i,(k)})p(\mathbf{x}_{[p]}^{i,(k)}|\mathbf{x}^{i,(k-1)})}{g\left(\mathbf{x}_{[p]}^{i,(k)}|\boldsymbol{\mu}_{[p-1]},\boldsymbol{\sigma}_{[p-1]}^2\right)}$
      13. Remember particle for further calculations $\mathbf{x}^{N(p-1)/P_n+i,(k)} = \mathbf{x}_{[p]}^{i,(k)}$

All particles have been divided into P_n parts (populations) and the particles in each population have been drawn from another importance function. Additional assumption has been made that standard deviation in p-th population can be maximum T_d times smaller than in the previous population (see steps 8–9).

4. ADAPTIVE IMPORTANCE SAMPLING

Some another approach to PMC method has been presented in [8] (Example 1), where each particle has been treated as an another population (number of populations equal to the number of particles N). In addition, the forgetting factor f_{rg} has been proposed (Proposition 2), which causes that old samples have lower weights in comparison to the newest one.

In this article the use of N_f index has been proposed – this is the number of additional particles, which have been drawn from the initial PDF. Without this the calculations with one (first) particle causes that the variance is equal to 0. The solution of this problem may be to set non-zero (but very small) initial values of W, $\mathbf{W}_X$ and $\mathbf{W}_{XX}$, but it does not work, when the first drawn particle has weight smaller than the initial value.

The Algorithm of Adaptive Importance Sampling (AIS) used in simulations, has been presented below.

Algorithm 5 – Adaptive Importance Sampling

1. $\boldsymbol{\mu}_{AIS} = \hat{\mathbf{x}}^{(k-1)}, \;\; \boldsymbol{\sigma}_{AIS} = \boldsymbol{\sigma}_{\text{start}}$
2. $W = 0, \;\; \mathbf{W}_X = 0, \;\; \mathbf{W}_{XX} = 0$
3. For $i = 1..N$ perform steps 4–11
   4. $\mathbf{x}^{i,(k)} \sim g\left(\mathbf{x}^{(k)}|\boldsymbol{\mu}_{AIS}, \boldsymbol{\sigma}^2_{AIS}\right) = \mathcal{N}\left(\boldsymbol{\mu}_{AIS}, \boldsymbol{\sigma}^2_{AIS}\right)$
   5. $q^{i,(k)} = \frac{p(\mathbf{z}^{(k)}|\mathbf{x}^{i,(k)})p(\mathbf{x}^{i,(k)}|\mathbf{x}^{i,(k-1)})}{g\left(\mathbf{x}^{i,(k)}|\boldsymbol{\mu}_{AIS}, \boldsymbol{\sigma}^2_{AIS}\right)}$
   6. $W = f_{rg} \cdot W + q^{i,(k)}$
   7. $\mathbf{W}_X = f_{rg} \cdot \mathbf{W}_X + q^{i,(k)} \cdot \mathbf{x}^{i,(k)}$
   8. $\mathbf{W}_{XX} = f_{rg} \cdot \mathbf{W}_{XX} + q^{i,(k)} \cdot \mathbf{x}^{i,(k)} \left(\mathbf{x}^{i,(k)}\right)^T$
   9. If $i > N_f$ then
      10. $\boldsymbol{\mu}_{AIS} = \frac{\mathbf{W}_X}{W}$
      11. $\boldsymbol{\sigma}_{AIS} = \sqrt{\frac{\mathbf{W}_{XX}}{W} - \boldsymbol{\mu}_{AIS} \cdot (\boldsymbol{\mu}_{AIS})^T}$

The square root in the step 11 means the Cholesky decomposition in a general case.

5. SIMULATION RESULTS

The average root mean square error (aRMSE) has been chosen as the estimation quality

$$\text{aRMSE} = \frac{1}{B}\sum_{i=1}^{B} \text{RMSE}_i \;, \tag{10}$$

$$\text{RMSE}_i = \sqrt{\text{MSE}_i} \;, \tag{11}$$

$$\mathrm{MSE}_i = \frac{1}{M}\sum_{k=1}^{M}\left(\hat{x}_i^{(k)} - x_i^{(k)+}\right)^2 , \tag{12}$$

where B is the number of state variables, M is the simulation length, and the state variables with hat and plus are the estimated and the real values respectively.

The object used in simulations has been presented below

$$x^{(k)} = x^{(k-1)} + v^{(k-1)} , \tag{13}$$
$$z^{(k)} = x^{(k)} + n^{(k)} , \tag{14}$$

where v and n are the Gaussian noises ($v \sim \mathcal{N}(0,1)$, $n \sim \mathcal{N}(0,1)$). The obtained results have been presented in Figure 1.

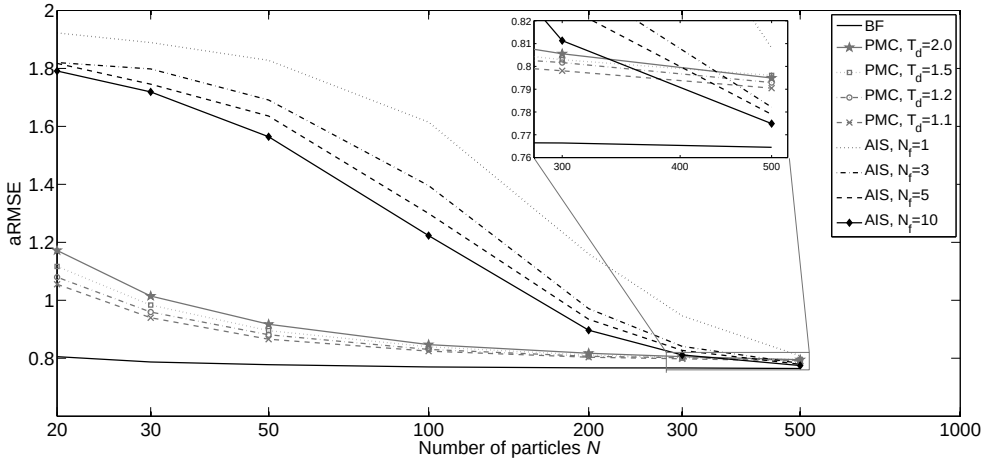

Fig. 1. Results for object described in (13–14); parameter used in PMC method: $P_n = 5$; parameters used in AIS method: $f_{rg} = 0.98$, $\sigma_{\mathrm{start}} = 0.1$

The impact of the threshold T_d in PMC method and the number of particles N_f in AIS method have been checked. As one can see, lower value of T_d provides better estimation quality. The same effect has a greater value of N_f, which have been used in initial calculations of mean and standard deviation in AIS method.

PMC and AIS methods are particularly useful in cases where measurement model is relatively thin or the measurement redundancy occurs. Therefore another object has been used, with one state variable and 4 measurements

$$x^{(k)} = x^{(k-1)} + v^{(k-1)} , \tag{15}$$
$$\begin{aligned} z_1^{(k)} &= x^{(k)} + n_1^{(k)} , \\ z_2^{(k)} &= x^{(k)} + n_2^{(k)} , \\ z_3^{(k)} &= x^{(k)} + n_3^{(k)} , \\ z_4^{(k)} &= x^{(k)} + n_4^{(k)} . \end{aligned} \tag{16}$$

Gaussian noises have been used the same as in (13–14), and the results have been presented in the Figure 2.

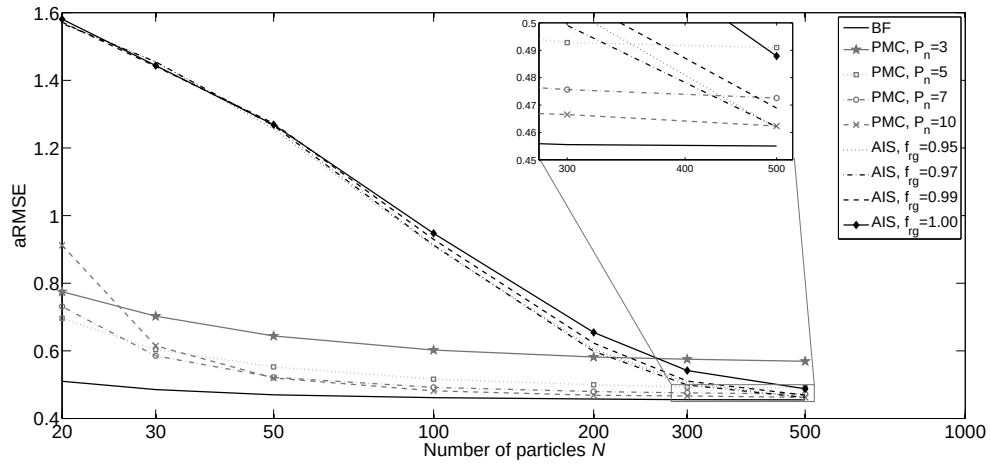

Fig. 2. Results for object described in (15–16); parameter used in PMC method: $T_d = 1.1$; parameters used in AIS method: $N_f = 10$, $\sigma_{\mathrm{start}} = 0.1$

As one can see, the higher number of populations in PMC method results in the lower final aRMSE value. Hence the AIS method should provide the best quality of estimation, what can be met for higher number of particles N. The best results for AIS method have been obtained for forgetting factor equal to $f_{rg} = 0.97$.

If the PMC and AIS methods "look for" the extremes, methods should work well also in cases, where the information about transition or measurement model is wrong. Therefore the same model as in the (13)–(14) has been used, but with another noises: $v \sim \mathcal{N}(0, 1^2)$ and $n_m \sim \mathcal{N}(0, 3^2)$, but the values in algorithms remained unchanged. Additionally, the AIS algorithm has been modified, i.e. the initial value of standard deviation σ_{AIS} has been assigned only in the first time step ("old-std"), and not in every time step ("new-std") as in the previous simulations. Thanks to this, the σ_{AIS} value should be close to the optimal for most of the particles. The results have been shown in Figure 3.

Based on the obtained results it can be said that for cases, in which measurement model is unknown, the both PMC and AIS methods could provide better estimation quality than the simple BF method. It can be seen that proposed modification in AIS algorithm does not provide better results.

It should be also noted that in this case the best results of PMC method have been obtained for the highest value of threshold T_d (otherwise than in the first simulations).

6. SUMMARY

The simple particle filter, i.e. Bootstrap Filter method, and two another algorithms have been presented in this article. Based on the performed simulations one can see that in most cases the BF method provides the best estimation quality, even for very small number of

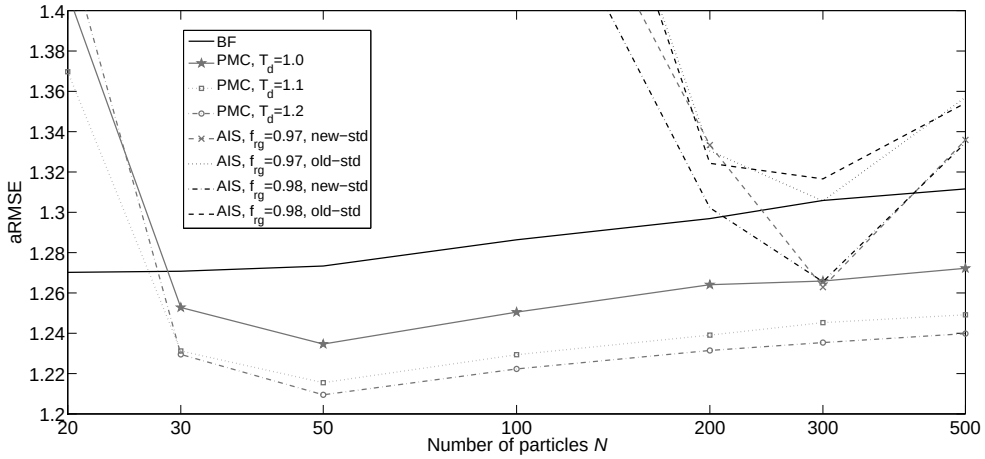

Fig. 3. Results for object described in (15–16), with increased variance of measurement model; parameter used in PMC method: $P_n = 10$; parameters used in AIS method: $N_f = 10$, $\sigma_{\text{start}} = 0.1$

particles. So why one should use the other methods? It should be noted that presented results are related to the object with one state variable. In such case the BF algorithm is sufficient even for very small N. However, the number of needed particles grows exponentially with the object dimension. Therefore, for object with a few or more state variables, one should use a method different than BF.

In the future, studies will be extended by the multidimensional cases, where the utility of the presented methods will be especially visible.

REFERENCES

[1] Arulampalam S., Maskell S., Gordon N., Clapp T., *A Tutorial on Particle Filters for On-line Nonlinear/Non-Gaussian Bayesian Tracking*, IEEE Proceedings on Signal Processing, Vol. 50(2), 2002, pp. 174–188.

[2] Candy J. V., *Bayesian signal processing*, WILEY, New Jersey 2009, pp. 36–44, 237–298.

[3] Cappe O., Guillin A., Marin J. M., Robert C. P., *Population Monte Carlo*, Journal of Computational and Graphical Statistics, Vol. 13(4), 2004, pp. 907–929.

[4] Doucet A., Freitas N., Gordon N., *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, New York 2001, pp. 225–246.

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

[6] Kozierski P., Lis M., Ziętkiewicz J., *Resampling in Particle Filtering – Comparison*, Studia z Automatyki i Informatyki, Vol. 38, 2013, pp. 35–64.

[7] Schon T. B., Wills A., Ninness B., *System Identification of Nonlinear State-space Models*, Automatica, Vol. 47(1), 2011, pp. 39–49.

[8] Šmídl V., Hofman R., *Adaptive Importance Sampling in Particle Filtering*, In Information Fusion, 16th International Conference on (FUSION), 2013, pp. 9–16.

[9] Thrun S., Burgard W., Fox D., *Probabilistic robotics*, MIT Press, Cambridge, MA, 2005, pp. 67–90.

ABSTRACT

Population Monte Carlo and Adaptive Importance Sampling methods have been presented and compared in the paper. The impact of parameters on the estimation quality of the plant also has been studied.

POPULACJA MONTE CARLO I ADAPTACYJNA FUNKCJA WAŻNOŚCI W FILTRZE CZĄSTECZKOWYM

Piotr Kozierski

W artykule przedstawiono i porównano metody Populacja Monte Carlo oraz Adaptacyjna Funkcja Ważności. Sprawdzono również wpływ parametrów tych metod na jakość estymacji stanu obiektu.

ACKNOWLEDGEMENTS

Huge thanks to Vaclav Šmídl, who helped me understand both presented algorithms.

Received: 2014-12-19
Accepted: 2015-03-18