ROBUST AUXILIARY PARTICLE FILTERS USING MULTIPLE IMPORTANCE SAMPLING

Joel Kronander[1], *Thomas B. Schön*[2]

[1] Dept. of Science and Technology, Linköping University, joel.kronander@liu.se

[2] Dept. of Information Technology, Uppsala University, thomas.schon@it.uu.se

ABSTRACT

A poor choice of importance density can have detrimental effect on the efficiency of a particle filter. While a specific choice of proposal distribution might be close to optimal for certain models, it might fail miserably for other models, possibly even leading to infinite variance. In this paper we show how mixture sampling techniques can be used to derive robust and efficient particle filters, that in general performs on par with, or better than, the best of the standard importance densities. We derive several variants of the auxiliary particle filter using both random and deterministic mixture sampling via multiple importance sampling. The resulting robust particle filters are easy to implement and require little parameter tuning.

Index Terms— Sequential Monte Carlo, particle filter, mixture sampling, multiple importance sampling.

1. INTRODUCTION

Particle filters [1, 2] constitute a popular class of methods for inference in nonlinear/non-gaussian state space models (SSMs) with X-valued latent states $\{x_t\}_{t=1}^T$ and Y-valued observations $\{y_t\}_{t=1}^T$, where T denotes the number of time steps. The SSM is defined by a transition density $f(x_t|x_{t-1})$, an observation density $g(y_t|x_t)$ and an initial distribution $x_1 \sim \mu(x_1)$. These methods approximates the filtering distributions $\{p(x_t|y_{1:t})\}_{t=1}^T$ using a set of weighted particles that are propagated forward in time using resampling and sequential importance sampling. The choice of importance density is a crucial design choice for practical applications of particle filters. For general state space models the optimal proposal function is intractable. While approximations of the optimal proposal function have been proposed [3, 4], they can perform poorly for certain models and they often result in a higher computational cost per particle. In practice, a common choice is instead to sample proportional to either the transition density $f(\cdot|x_{t-1})$ or the observation density $g(y_t|\cdot)$ [2]. However, this choice is far from optimal. For heavy tailed noise distributions a poor choice of proposal distribution could even result in estimators with infinite variance [1]. In this paper we show how an auxiliary particle filter can be extended to sample from a mixture of several proposal densities. The simplest approach is to independently sample a proposal function for each particle from a random mixture of proposals. This has also been proposed previously in the literature [5, 6]. We also show how *multiple importance sampling* (MIS) [7] can be used to derive robust particle filters. These methods improve on random mixture sampling by sampling a fixed number of particles from each proposal function, effectively stratifying the random mixture. We discuss and compare several weighting strategies for combining the particles from the different proposal functions. Based on an experimental study we show that mixture sampling performs almost as good as sampling from the best choice of the transition density or the observation density, providing a safe and robust alternative to standard methods. We also show that for models where both the transition density and the observation density are represented by heavy tailed distributions, such as student's t-distributed noise, mixture sampling can significantly outperform the standard proposal distributions, rendering mixture sampling methods a much better choice in these scenarios.

This work was supported by: the Swedish Foundation for Strategic Research (SSF) through grant IIS11-0081, the project Probabilistic modeling of dynamical systems (Contract number: 621-2013-5524) funded by the Swedish Research Council and CADICS, a Linnaeus Center also funded by the Swedish Research Council.

2. AUXILIARY PARTICLE FILTER

The auxiliary particle filter (APF) [8, 9] sequentially compute weighted particle systems $\{x_t^i, w_t^i\}_{i=1}^N$ targeting $p(x_t|y_{1:t})$ with the empirical approximation $\widehat{p}(x_t|y_{1:t}) = \sum_{i=1}^N w_t^i \delta_{x_t^i}(x_t)$, where $\delta_{x_0}(x)$ denotes a Dirac delta mass located at x_0. Given a weighted particle system approximating $p(x_{t-1}|y_{1:t-1})$ the filtering distribution at time t is approximated by

$$\widehat{p}(x_t|y_{1:t}) \propto \sum_{i=1}^N w_{t-1}^i g(y_t|x_t) f(x_t|x_{t-1}^i) \tag{1}$$

The APF targets this distribution using importance sampling in the augmented space $\mathsf{X} \times \{1, \ldots, N\}$. Specifically, let a_t be a uniformly distributed random variable defined on the index space $\{1, \ldots, N\}$. Let $\pi(x_t, a_t)$ be a probability density defined on $\mathsf{X} \times \{1, \ldots, N\}$ according to

$$\pi(x_t, a_t) \propto w_{t-1}^{a_t} g(y_t|x_t) f(x_t|x_{t-1}^{a_t}) \quad (2)$$

Marginalizing this density over a_t gives (1). This distribution is targeted using self-normalized importance sampling by first sampling the (auxiliary) index variable $a_t \sim Cat(\{v_{t-1}^i)\}_{i=1}^N)$, where v_{t-1}^i denotes the probability of sampling index i, and $Cat(\{p^i\}_{i=1}^n)$ denotes the categorical distribution on $\{1, \ldots, N\}$ with probabilities $\{p^i\}_{i=1}^N$. A new particle location is then sampled according to $x_t \sim q(x_t|x_{t-1}^{a_t^i}, y_t)$, resulting in the importance weights

$$\widetilde{w}_t^i = \frac{w_{t-1}^{a_t^i} g(y_t|x_t^i) f(x_t^i|x_{t-1}^{a_t^i})}{v_{t-1}^{a_t^i} q(x_t^i|x_{t-1}^{a_t^i}, y_t)}, \quad w_t^i = \frac{\widetilde{w}_t^i}{\sum_{i=1}^N \widetilde{w}_t^i} \quad (3)$$

The weighted particle system $\{x_t^i, w_t^i\}_{i=1}^N$ can then be used to approximate $p(x_{t+1}|y_{1:t+1})$ analogously to the above development. The algorithm is initialized at time $t = 1$ by sampling from the prior $\mu(x_1)$.

3. RANDOM MIXTURE PROPOSALS

A well established criteria to design efficient proposal functions is to minimize the variance of the unnormalized importance weights $\widetilde{w}_t$ [1]. The *optimal proposal distributions* minimizing the variance of the importance weights are given by $v_{t-1}^i = w_{t-1}^i p(y_t|x_{t-1}^i)$ and $q(x_t|x_{t-1}^{a_t}, y_t) = p(x_t|x_{t-1}^{a_t}, y_t)$. However, for most models these proposal distributions are not analytically tractable. A common choice for the index proposal is instead $v_{t-1}^i = w_{t-1}^i$, which corresponds to resampling the particles at time $t-1$. Assuming that $v_{t-1}^i = w_{t-1}^i$, let us consider the choice of proposal for x_t, $q(x_t|x_{t-1}^{a_t}, y_t)$. In practice, this proposal distribution is often constrained to either the transition density $f(\cdot|x_{t-1}^{a_t})$ or some distribution proportional to the observation density $q_g(x_t|y_t) \propto g(y_t|\cdot)$. However, when the proposal function is a poor match for the target distribution, the resulting approximation can be extremely poor. Specifically the importance weights $\widetilde{w}_t$ only have a finite asymptotical variance if

$$\int \frac{(g(y_t|x_t) f(x_t|x_{t-1}))^2}{q(x_t|x_{t-1}, y_t)} \mathrm{d}x_t < \infty \quad (4)$$

Thus if $q(x_t|x_{t-1}, y_t)$ decrease towards zero faster than $(g(y_t|x_t) f(x_t|x_{t-1}))^2$ as x_t moves away from its mode(s), the asymptotical variance of the importance weights will be infinite.

Instead of sampling exclusively from either the transition or observation density, consider sampling from the random mixture proposal

$$q(x_t|x_{t-1}^{a_t}, y_t) = \alpha f(x_t|x_{t-1}^{a_t^i}) + (1-\alpha) q_g(x_t|y_t) \quad (5)$$

where $\alpha \in [0, 1]$ is the mixture coefficient. This proposal function results in a finite variance of the importance weights if any of the mixture components provide a finite variance as

$$\int \frac{(g(y_t|x_t) f(x_t|x_{t-1}^{a^t}))^2}{\alpha_1 q_1(x_t) + \alpha_2 q_2(x_t)} \mathrm{d}x_t \leq \frac{1}{\alpha_p} \int \frac{(g(y_t|x_t) f(x_t|x_{t-1}^{a^t}))^2}{q_p(x_t)} \mathrm{d}x_t < \infty \quad (6)$$

for $p = 1$ and/or $p = 2$. The mixture distribution (5) thus represents a robust choice when the optimal proposal function is intractable. A similar argument can be made also for other choices of the index proposal weights v_{t-1}^i.

4. MULTIPLE IMPORTANCE SAMPLING

Instead of sampling a random number of particles from each proposal distribution, consider sampling a fixed, deterministic, number of samples from each proposal. There exist many ways of combining these samples still leading to unbiased estimators. Veach and Guibas [7] first studied this family of sampling techniques for solving integration problems in computer graphics under the name *multiple importance sampling* (MIS).

Let $\{q_p(x)\}_{p=1}^M$ denote a set of proposal distributions. By proposing $n_p > 0$ samples from $q_p(x)$ for $p = 1, \ldots, M$ the MIS estimator of a target density $\pi(x)$ is given by the empirical approximation

$$\widehat{\pi}(x) = \sum_{p=1}^P \frac{1}{n_p} \sum_{i=1}^{n_p} \beta_p(x^{p,i}) \frac{\pi(x^{p,i})}{q_p(x^{p,i})} \delta_{x^{p,i}}(x), \quad x^{p,i} \sim q_p(x) \quad (7)$$

where $x^{p,i}$ denotes the i^{th} sample drawn from proposal q_p and $\beta_p(x) \geq 0$ are the mixture weights satisfying $\sum_{p=1}^M \beta_p(x) = 1$ for all x. Let $\phi(x)$ be an arbitrary test function. It can be shown that the resulting estimator of $E_\pi[\phi(x)]$ is unbiased [10]. A well studied proposal for the mixture weights is the so called *balance heuristic* setting $\beta_p(x) = \frac{n_p q_p(x)}{\sum_{p=1}^M n_p q_p(x)}$. While there is no optimal choice of the mixture weights in the general case, one can show that the weights given by the balance heuristic is never much worse than any other weighting strategy [11, 10].

An APF proposing particles from both the transition density and observation density can be derived by using a self-normalized MIS estimator targeting the auxiliary distribution $\pi(x_t, a_t)$ given by (2). Choosing the mixture weights according to the balance heuristic results in the estimator

$$\widehat{p}(x_t|y_{1:t}) = \sum_{i=1}^{N} w_t^i(x_t^i)\delta_{x_t^i}(x_t) \tag{8a}$$

$$\widetilde{w}_t^i = \frac{w_{t-1}^{a_t^i} g(y_t|x_t^i) f(x_t^i|x_{t-1}^{a_t^i})}{v^i(N_f f(x_t^i|x_{t-1}^{a_t^i}) + N_g q_g(x_t^i|y_t))} \tag{8b}$$

$$w_t^i = \frac{\widetilde{w}_t^i}{\sum_{i=1}^N \widetilde{w}_t^i} \tag{8c}$$

where N_f and N_g are the number of samples drawn from $f(x_t^i|x_{t-1}^{a_t^i})$ and $q_g(x_t^{p,i}|y_t)$, respectively. This is the same estimator as obtained by stratifying the number of samples drawn from each mixture component in the random mixture proposal (5), i.e. drawing $N_f = \alpha N$ particles from $f(x_t|x_{t-1})$ and $N_g = (\alpha - 1)N$ particles from $q_g(x_t|y_t)$. It can be shown that the MIS estimator (7) using the balance heuristic always results in an estimator with lower variance than a standard importance sampling estimator with the corresponding random mixture proposal [11, 10]. This estimator is thus expected to lead to lower variance of the importance weights, and is always to be preferred over sampling from the random mixture proposal (5). An APF based on the balance heuristic is summarized in Algorithm 1.

Compared to a standard APF designed to sample new particles proportional to either the the transition density or the observation density, Algorithm 1 come at the cost of a slightly more computationally demanding weight calculation. The reason is that, for each particle weight, both the transition density and the observation density have to be evaluated and do not cancel with the proposal in the expression for the importance weights. When these densities are computationally demanding to evaluate this can be a major concern. However, a MIS estimator with the same computational requirements as a standard APF, can be derived using the mixture weights, $\beta_p(x) = \frac{1}{M}$. The resulting algorithm is summarized in Algorithm 2.

Algorithm 1: MIS APF using the balance heuristic

1 Sample $x_1^i \sim \mu(x_1)$ for $i = 1, \ldots, N = N_f + N_g$
2 Set $w_t^i = \frac{1}{N}$ for $i = 1, \ldots, N$
3 **for** $t = 2, \ldots, T$ **do**
4 Sample $a_t^i \sim Cat(\{v_{t-1}^j\}_{j=1}^N)$ for $i = 1, \ldots, N$
5 Sample $x_t^i \sim f(x_t^i|x_{t-1}^{a_t^i})$ for $i = 1, \ldots, N_f$
6 Sample $x_t^i \sim q_g(x_t^i|y_t)$ for $i = N_f + 1, \ldots, N$
7 Set $\widetilde{w}_t^i = \frac{w_{t-1}^{a_t^i} f(x_t^i|x_{t-1}^{a_t^i}) g(y_t|x_t^i)}{v_{t-1}^i(N_f f(x_t^i|x_{t-1}^{a_t^i}) + N_g q_g(x_t^i|y_t))}$ for $i = 1, \ldots, N$
8 Set $w_t^i = \frac{\widetilde{w}_t^i}{\sum_{i=1}^N \widetilde{w}_t^i}$ for $i = 1, \ldots, N$
9 **end**

Algorithm 2: MIS APF using $\beta_p(x) = \frac{1}{M}$

Same as Algorithm 1, but with line 7 replaced by

Set $\widetilde{w}_t^i = \frac{w_{t-1}^{a_t^i} g(y_t|x_t^i)}{v^i}$ for $i = 1, \ldots, N_f$

Set $\widetilde{w}_t^i = \frac{w_{t-1}^{a_t^i} f(x_t^i|x_{t-1}^{a_t^i})}{v^i}$ for $i = N_f + 1, \ldots, N$

5. NUMERICAL ILLUSTRATION

To evaluate the proposed methods we have compared them to a standard particle filter proposing samples from a) the transition density and b) proportional to the observation density. For all filters resampling is performed at each time step.

5.1. Linear model

Consider the following state space model model

$$x_t = x_{t-1} + v_t, \tag{9a}$$

$$y_t = x_t + e_t \tag{9b}$$

where $x_1 \sim \mathcal{N}(0, 0.1)$. We consider two cases of this model, with Gaussian and Student's t-distributed noise, respectively.

Gaussian noise - Let $v_t \sim \mathcal{N}(0, \sigma_v^2)$ and $e_t \sim \mathcal{N}(0, \sigma_e^2)$. For this type of model 100 data realizations were considered, each with $T = 200$. Table 1 show mean squared error (MSE) compared to the Kalman filter solution for $N = 100$ ($n_f = 50, n_g = 50$) particles for varied settings of σ_v^2 and σ_e^2. Sampling from the transition density results in better estimators when the process noise (σ_v) is low and the observation noise (σ_e) is high. Sampling proportional to the observation density exhibits opposite properties. The MIS particle filter given in Algorithm 1 performs on par with the best of the other proposals in both cases with higher noise, and performs better than all other methods when considering equal process and observation noise. The MIS particle filter given in Algorithm 2 performs a bit worse than the balance heuristic although it exhibits good results overall at a slightly lower computational cost.

σ_v	σ_e	Transition density	Observation density	Alg. 1	Alg. 2
1	1	0.016	0.016	**0.008**	0.014
0.5	2	**0.019**	0.055	**0.019**	0.031
0.2	5	**0.028**	0.327	0.041	0.124
2	0.5	0.018	0.006	**0.005**	0.008
5	0.2	0.033	**0.002**	0.003	0.005

Table 1. Average MSE relative to Kalman filter solution for the linear model (9) with Gaussian noise.

Student's t-distributed noise - Let $v_t \sim \text{St}(\nu_v)$ and $e_t \sim \text{St}(\nu_e)$, where $\text{St}(\nu)$ denotes the Student's t-distribution with

ν degrees of freedom. For this model we considered 200 data realizations, each of length $T = 400$. For all filters we considered $N = 200$ ($N_f = 100, N_g = 100$) particles. Table 2 lists the MSE relative to the true latent states. The APF given in Algorithm 1 performs best for all considered parameter settings. The APF given in Algorithm 2 performs almost as good, and is for all parameter settings better than using an APF proposing from either the transition or proportional to the observation density for an equal computational cost.

σ_v	σ_e	Transition density	Observation density	Alg. 1	Alg. 2
2	2	56.1	9.1	**5.7**	5.9
2	3	135.4	2.3	**2.1**	**2.1**
3	2	4.8	8.8	**2.4**	2.5

Table 2. Average MSE for the linear model (9) with student's t-distributed noise.

5.2. Nonlinear model

Consider the nonlinear model

$$x_t = \frac{x_{t-1}}{2} + \frac{bx_{t-1}}{1 + x_{t-1}^2} + 8\cos 1.2t + v_t \tag{10a}$$

$$y_t = \frac{x_t^2}{20} + e_t \tag{10b}$$

where $x_1 \sim \mathcal{N}(0, 0.1)$, $v_t \sim \mathcal{N}(0, \sigma_v^2)$ and $e_t \sim \mathcal{N}(0, \sigma_e^2)$. To sample proportional to the observation density for this model, we use the proposal density derived in [12]. Table 3 shows MSE relative to the true latent states for different settings of b for $N = 100$ ($N_f = 50, N_g = 50$). For each value for b we averaged the MSE over 200 datasets each of length $T = 100$. For $b = 1$ and $b = 10$ sampling from the transition density performs better than sampling proportional to the observation density. For $b = 20$ and $b = 30$ sampling proportional to the observation density gives better results than sampling from the transition density. The APF given in Algorithm 1 performs best for all the different parameter settings. The performance of the APF given in Algorithm 2 performs on par with or almost as good as the best choice of sampling from the transition density or proportional to the observation density.

b	Transition density	Observation density	Alg. 1	Alg. 2
1	**4.3**	4.6	**4.3**	**4.3**
10	**7.3**	7.7	**7.3**	7.4
20	19.0	18.8	**17.7**	18.5
30	36.6	34.4	**32.6**	36.5

Table 3. Average MSE for the nonliner model (10).

6. DISCUSSION

Apart from the balance heuristic, several other MIS weighting strategies have been considered with success in e.g. computer graphics applications. The *power heuristic*, sets $\beta_p(x) = \frac{(N_p q_p(x))^c}{\sum_{p=1}^M (N_p q_p(x))^c}$, where $c \in [0, \infty]$. These weights should intuitively provide better estimators when one of the proposals $q_p(x)$ is close to $\pi(x)$, but not the mixture $\sum_{p=1}^M q_p(x)$ [11]. However, in our experiments the resulting APFs did not perform better than the proposed APF using the balance heuristic. Other MIS weight heuristics could however prove to be useful in specific applications.

Previous work have considered approximations to the optimal proposal function using linearization [3] or the unscented transform to design a Gaussian importance density for each particle [4]. Such proposal distributions might also be included in the pool of considered proposal distributions, and the APFs proposed in this work could easily be extended to this case. Similar ideas has previously been explored for random mixture sampling [6].

7. REFERENCES

[1] Arnaud Doucet and Adam M Johansen, "A tutorial on particle filtering and smoothing: 15 years later," in *Oxford Handbook of Nonlinear Filtering*. Oxford University Press, 2011.

[2] Fredrik Gustafsson, "Particle filter theory and practice with positioning applications," *IEEE Aerospace and Electronic Systems Magazine*, vol. 25, no. 7, pp. 53–82, 2010.

[3] Arnaud Doucet, Simon Godsill, and Christophe Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and computing*, vol. 10, no. 3, pp. 197–208, 2000.

[4] Rudolph Van Der Merwe, Arnaud Doucet, Nando De Freitas, and Eric Wan, "The unscented particle filter," in *Neural Information Processing Systems (NIPS)*, 2000, pp. 584–590.

[5] Sebastian Thrun, Dieter Fox, Wolfram Burgard, et al., "Monte Carlo localization with mixture proposal distribution," in *Proceedings of the AAAI National Conference on Artificial Intelligence*, Austin, TX, USA, 2000, pp. 859–865.

[6] Junyi Zuo, Yan Liang, Yizhe Zhang, and Quan Pan, "Particle filter with multimode sampling strategy," *Signal Processing*, vol. 93, no. 11, pp. 3192–3201, 2013.

[7] Eric Veach and Leonidas J Guibas, "Optimally combining sampling techniques for Monte Carlo rendering," in *Proceedings of SIGGRAPH*, New York, NY, USA, 1995, ACM, pp. 419–428.

[8] Michael K Pitt and Neil Shephard, "Filtering via simulation: Auxiliary particle filters," *Journal of the American statistical association*, vol. 94, no. 446, pp. 590–599, 1999.

[9] James Carpenter, Peter Clifford, and Paul Fearnhead, "Improved particle filter for nonlinear problems," *IEE Proceedings-Radar, Sonar and Navigation*, vol. 146, no. 1, pp. 2–7, 1999.

[10] Art Owen and Yi Zhou, "Safe and effective importance sampling," *Journal of the American Statistical Association*, vol. 95, no. 449, pp. 135–143, 2000.

[11] Eric Veach, *Robust Monte Carlo methods for light transport simulation*, Ph.D. thesis, Stanford University, 1997.

[12] M. Arulampalam, Sanjeev, Simon Maskell, Neil Gordon, and Tim Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.