On Marginal Particle Filters with Linear Complexity

Fredrik Gustafsson, *Fellow, IEEE*

***Abstract*—The choice of proposal distribution affects critically the efficiency of a particle filter (PF). When it is not feasible to increase the number of particles, then more advanced proposals may avoid particle depletion. One further option is to compute the marginal posterior distribution. However, exact marginalization has quadratic complexity in the number of particles. This contribution discusses and compares some ideas for implementing a marginal PF with linear complexity, leading to an accept-reject algorithm. However, it turns out that the proposed solution can be interpreted as a way to implement the optimal proposal. Different proposals with and without marginalization are compared on a simple example.**

I. Introduction

Particle filtering (PF) as a research area started with the seminal paper [5], and the independent developments in [8], [7]. The key contribution was to introduce a resampling step to mitigate the depletion problem in simulation based approaches that have been around for a long time. Mitigating depletion has ever since the beginning been the most pressing issue in applied particle filtering. The state of the art is summarized in the article collection [4], the surveys [10], [1], [3], [2], [6], and the monograph [14].

The PF inherently estimates the state trajectory, not the filter density as is the output from the Kalman filter for instance. Although the marginal distribution can be trivially derived from the trajectory density, dedicated algorithms to compute the marginal density have been in focus for the last decade. Hereafter, this will be refered to as marginal particle filters (mPF). The key motivation is to decrease the weight variance, and thus decrease the sample depletion problem. A good indicator for this is the efficient number of particles. The mPF should have a larger efficient number of particles than the corresponding PF without marginalization.

The mPF has been proposed for system identification problems where some of the states are stationary [12], [13]. Stationary parameters are particular sensitive to depletion, since there is no inherent forgetting of the past. The idea in [12], [13] is to compute the gradient search of the marginal distribution with respect to the unknown parameters. The same parametric approach has been suggested for SLAM in [11] and optimal trajectory planning in [15].

II. The Particle Filter

The particle filter applies to a dynamic model, either given in state space form or as conditional probability functions,

$$x_{k+1} = f(x_k, v_k) \quad \text{or} \quad p(x_{k+1}|x_k), \tag{1a}$$

$$y_k = h(x_k, e_k) \quad \text{or} \quad p(y_k|x_k). \tag{1b}$$

The particle filter recursion at time k for N particles $x_{k-1}^{(i)}$ with associated weights $w_{k-1|k-1}^{(i)}$ can be summarized as follows.

Department of Electrical Engineering, Division of Automatic Control, Linköping University, SE-581 83 Linköping, Sweden, e-mail: {fredrik.gustafsson}@liu.se

1) *Weight update:* Time and measurement updates:

$$w_{k|k-1}^{(i)} \propto w_{k-1|k-1}^{(i)} \frac{p(x_k^{(i)}|x_{k-1}^{(i)})}{q(x_k^{(i)}|x_{k-1}^{(i)}, y_k)}, \tag{2a}$$

$$w_{k|k}^{(i)} \propto w_{k|k-1}^{(i)} p(y_k|x_k^{(i)}). \tag{2b}$$

2) *Resampling:* Optionally at each time, take N samples with replacement from the set $\{x_{1:k}^{(i)}\}_{i=1}^N$ where the probability to take sample i is $w_{k|k}^{(i)}$ and let $w_{k|k}^{(i)} = 1/N$.
3) *Prediction:* Generate samples from the proposal

$$x_{k+1}^{(i)} \sim q(x_{k+1}|x_k^{(i)}, y_{k+1}) \tag{2c}$$

The most natural proposal distributions are

- The prior $q(x_{k+1}|x_k^{(i)}, y_{k+1}) = p(x_{k+1}|x_k^{(i)})$.
- The likelihood $q(x_{k+1}|x_k^{(i)}, y_{k+1}) \propto p(y_{k+1}|x_{k+1})$.
- The optimal (minimizing weight variance) choice $q(x_{k+1}|x_k^{(i)}, y_{k+1}) \propto p(y_{k+1}|x_{k+1})p(x_{k+1}|x_k^{(i)})$.

The optimal proposal keeps the weights constant, and this would in theory avoid depletion, where depletion is interpreted as excessive weight variance. The practical limitation with the last two alternatives is that there need to be more measurements than states in the model.

III. The Marginal Particle Filter

The PF approximates the posterior distribution $p(x_{1:k}|y_{1:k})$ of the state trajectory $x_{1:k}$, while the purpose of the marginal particle filter (mPF) is to approximate the filter distribution $p(x_k|y_{1:k})$. The principal relation between these two distributions is given by the marginalization integral

$$p(x_k|y_{1:k}) = \int p(x_{1:k}|y_{1:k})\, dx_{1:k-1}. \tag{3}$$

To start an inductive derivation of the mPF, suppose that we at time k have an approximation of the marginal filter density of the form

$$p(x_k|y_{1:k}) = \sum_{i=1}^N w_{k|k}^{(i)} \delta(x_k - x_k^{(i)}). \tag{4}$$

There are a few different alternatives to proceed described in the following sections.

A. Sampling and Marginalization

The standard (trajectory) PF samples a new state from the proposal distribution and modifies the weights according to the importance sampling principle. The new sample is then appended to the previous one $x_{k:k+1}^{(i)} = \{x_k^{(i)}, x_{k+1}^{(i)}\}$. We then have an approximation of the form

$$p(x_{k:k+1}|y_{1:k}) = \sum_{i=1}^N w_{k+1|k}^{(i)} \delta(x_k - x_k^{(i)}) \delta(x_{k+1} - x_{k+1}^{(i)}), \tag{5}$$

978-1-4673-3146-3/13/$31.00 ©2013IEEE

which is trivially marginalized to

$$p(x_{k+1}|y_{1:k}) = \int p(x^{(i)}_{k:k+1}|y_{1:k})dx_k \tag{6}$$

$$= \sum_{i=1}^{N} w^{(i)}_{k+1|k}\delta(x_{k+1} - x^{(i)}_{k+1}) \tag{7}$$

That is, a straightforward application of the marginalization principle does not affect the result, and will not improve on the depletion problem.

B. Marginalization before Sampling

A more promising approach is based on applying the marginalization principle in the prediction step

$$p(x_{k+1}|y_{1:k}) = \int p(x_{k:k+1}|y_{1:k})dx_k \tag{8}$$

$$= \int p(x_{k+1}|x_k)p(x_k|y_{1:k})dx_k. \tag{9}$$

Using the filter approximation (4) gives

$$p(x_{k+1}|y_{1:k}) = \sum_{i=1}^{N} w^{(i)}_{k|k}p(x_{k+1}|x^{(i)}_k). \tag{10}$$

This is an analytic form of the prediction density where no further approximations besides (4) have been done. There are now several ways to sample from (10).

C. Marginalization and Multinomial Sampling

The predictive density (10) is a mixture distribution, which can be sampled from directly, using the following algorithm:

1) Select N indeces i_j, $j = 1, \ldots, N$ from the multinomial distribution specified by the weights $\{w^{(j)}_{k|k}\}_{j=1}^{N}$.
2) For each index, sample $x^{(j)}_{k+1} \sim p(x_{k+1}|x^{i_j}_k)$, $j = 1, \ldots, N$.

This has linear complexity. However, this is exactly what is done in the SIR (or bootstrap) PF, which performs a resampling step corresponding to the multinomial sampling, followed by a prediction.

D. Marginalization and Systematic Sampling

The standard approach in literature appears to be to re-use the set of particles from the sampling exactly as it is. This gives

$$p(x_{k+1}|y_{1:k}) = \sum_{i=1}^{N} w^{(i)}_{k+1|k}\delta(x_{k+1} - x^{(i)}_{k+1}), \tag{11a}$$

$$w^{(i)}_{k+1|k} = \sum_{j=1}^{N} w^{(j)}_{k|k}p(x^{(i)}_{k+1}|x^{(j)}_k). \tag{11b}$$

This can be seen as systematic sampling in the available set of particles, where all samples are selected in turn, and the weight is updated accordingly. Note that the complexity increases from $\mathcal{O}(N)$ in the PF to $\mathcal{O}(N^2)$ in the marginal PF, due to the computation of the new importance weights. However, a sound approximation with $\mathcal{O}(N\log(N))$ complexity is suggested in [9].

E. Marginalization and Importance Sampling

The idea is here to use the standard PF in (2) with simple marginalization to get the proposal $p(x_{k+1}|y_{1:k+1})$ in an importance sampling scheme. That is, generate samples $x^{(j)}_{k+1}$ from the set $\{\bar{x}^{(i)}_{k+1}\}_{i=1}^{N}$ according to the weights $\{\bar{w}^{(i)}_{k+1|k}\}_{i=1}^{N}$, where the bar indicates that this is a preliminary set of samples. The weights for the marginal distribution are then modified by the importance sampling principle,

$$w^{(i)}_{k+1|k} = \frac{\sum_{j=1}^{N} w^{(j)}_{k|k}p(x^{(i)}_{k+1}|x^{(j)}_k).}{\bar{w}^{(i)}_{k+1|k}}. \tag{12}$$

Again, the complexity is quadratic due to the sum.

F. Marginalization and Accept-Reject Sampling

The key idea in the marginalization is to break up the particle pair associations,

$$x^{(i)}_k \leftrightarrow x^{(i)}_{k+1}, \tag{13}$$

from the prediction step. The systematic and importance sampling principles above both lead to that all possible paths from all $x^{(j)}_k$ to each $x^{(i)}_{k+1}$ are considered, which is the source for the quadratic complexity. Now, we randomize the associations

$$x^{(j)}_k \leftrightarrow x^{(i)}_{k+1}, \tag{14}$$

by randomly choosing pairs i, j with the following accept-reject algorithm:

1) Generate a random index i_k from the multinomial distribution specified by $\{w^{(i)}_{k|k}\}_{i=1}^{N}$.
2) Generate a random index j_k from the multinomial distribution specified by $\{w^{(j)}_{k+1|k+1}\}_{j=1}^{N}$.
3) Generate a uniform random number u.
4) Accept $x^{j_k}_{k+1}$ if

$$\frac{p(x^{j_k}_{k+1}|x^{i_k}_k)}{w^{j_k}_{k+1|k+1}M} > u. \tag{15}$$

5) Repeat until N samples $x^{j_k}_{k+1}$ are obtained.

The accept-reject algorithm requires that the left hand side of (15) is always less than one. That is, we have to choose the normalization factor M such that

$$M \geq \frac{p(x^{j_k}_{k+1}|x^{i_k}_k)}{w^{j_k}_{k+1|k+1}} \tag{16}$$

If we use the particles after the resampling step in the basic PF (2), we have $w^{j_k}_{k+1|k+1} = 1/N$, so the denominator is bounded from below. The Markov model assures that $p(x_{k+1}|x_k)$ is bounded, since all non-degenerate distributions are bounded.

For instance, a typical model with additive Gaussian noise

$$x_{k+1} = f(x_k) + v_k, \tag{17a}$$

$$v_k \in N(0, Q), \tag{17b}$$

implies an upper bound

$$p(x_{k+1}|x_k) \leq (\det 2\pi Q)^{-1/2}. \tag{18}$$

Thus, we can take

$$M = N(\det 2\pi Q)^{-1/2} \tag{19}$$

in the case of additive Gaussian noise.

The interpretation is that this leads to a two step procedure:

1) A non-SIR PF is used to provide a set of candidate particles. Both the measurement update and resampling steps are required. There is as usual in the PF fixed association of particle pairs $x_k^{(i)}$ and $x_{k+1}^{(i)}$. Note that the marginal distribution of a SIR PF is identical to the trajectory distribution, so there is no point to approximate the marginal further.
2) The second step reorders these associations randomly using accept-reject sampling.

IV. Illustration

Various proposal distributions will be compared on a fundamentally simple example that enables closed form simple expressions. Consider a first order linear Gaussian system

$$x_{k+1} = fx_k + v_k, \qquad v_k \sim \mathrm{N}(0, Q). \tag{20}$$
$$y_k = hx_k + e_k, \qquad e_k \sim \mathrm{N}(0, R), \tag{21}$$
$$x_0 \sim \mathrm{N}(\bar{x}_0, P_0). \tag{22}$$

Let, again for simplicity, $f = h = 1$, $\bar{x}_0 = 1$, $P_0 = 0.1$, $R = 0.1$. The process noise variance $Q \in [10^{-4}, 1]$ will be varied to examine different signal to noise ratios (SNR), here defined as Q/R.

To make the simulation study as simple and transparent as possible, only two time steps $k = 1, 2$ will be simulated. The purpose of the first time step is to normalize information to the current SNR, and the second one is for computing the root mean square error (RMSE) for the following different proposals:

1) Prior proposal

$$x_2^{(i)} \sim \mathrm{N}(x_2 - hx_1^{(i)}, Q), \tag{23}$$
$$\omega_2^{(i)} = \omega_1^{(i)} \mathrm{N}(y_2 - hx_2^{(i)}, R). \tag{24}$$

2) Likelihood proposal

$$x_2^{(i)} \sim \mathrm{N}(y_2/h, R/h^2), \tag{25}$$
$$\omega_2^{(i)} = \omega_1^{(i)} \mathrm{N}(x_2^{(i)} - hx_1^{(i)}, Q). \tag{26}$$

The proposal follows by solving the measurement equation for $x_2 = (y_2 - e_2)/h$.

3) Optimal proposal

$$x_2^{(i)} \sim \mathrm{N}(x_1^{(i)} + K(y_2 - hx_1^{(i)}), P), \tag{27}$$
$$\omega_2^{(i)} = \omega_1^{(i)}. \tag{28}$$

where

$$K = \frac{Qh}{h^2Q + R}, \tag{29}$$
$$P = Q - \frac{h^2Q^2}{h^2Q + R} = \frac{QR}{h^2Q + R}. \tag{30}$$

The optimal proposal follows from the Kalman filter, starting with a given particle $x_1^{(i)}$, which has zero variance.

4) Likelihood proposal with full marginalization.

$$x_2^{(i)} \sim \mathrm{N}(y_2/h, R/h^2), \tag{31}$$
$$\omega_2^{(i)} = \sum_{j=1}^{N} \omega_1^{(j)} \mathrm{N}(x_2^{(i)} - hx_1^{(j)}, Q). \tag{32}$$

5) Likelihood proposal with accept/reject marginalization, where a sample $x_2^{(i)} \sim \mathrm{N}(y_2/h, R/h^2)$ is accepted if

$$\frac{\left(x_2^{(i)} - fx_1^{(i)}\right)}{Q} < -\log(u) \tag{33}$$

where u is a random number from the standard uniform distribution.

6) The Kalman filter solution, which provides the Cramer-Rao lower bound (CRLB) for this problem. The variance updates can be simplified to

$$P_{1|1} = \frac{P_0 R}{P_0 + R}, \tag{34}$$
$$P_{2|1} = P_{1|1} + Q, \tag{35}$$
$$P_{2|2} = \frac{P_{2|1} R}{P_{2|1} + R}. \tag{36}$$

The results from 5000 Monte Carlo simulations with $N = 100$ are summarized in Figure 1 and below:

- For low SNR, the prior proposal reaches the CRLB, while the likelihood proposal reaches the CRLB for high SNR. This is according to intuition, the most informative proposal should give the best result.
- The full marginalization improves the RMSE slightly for the likelihood proposal.
- The accept/reject marginalization improves the RMSE for high SNR's. To understand why, the acceptance rate is plotted against SNR in Figure 2. As seen, the acceptance rate is very small for low SNR.
- Figure 3 shows that the computation time of the likelihood proposal with accept-reject sampling has linear complexity over the whole range, though the constant term is quite high (similar to full marginalization).
- The accept/reject algorithm simplifies a lot for this concrete example, and there are probably many other interpretations of the algorithm than the one presented here: to randomize the connections between particle i and j in the full marginalization step.

The accept-reject algorithm was derived to avoid the quadratic complexity in marginalization. However, for this concrete example where the likelihood is used as proposal of samples, another interpretation is that it implements an accept-reject version of the optimal proposal.

V. Acknowledgments

The author would like to thank Fredrik Lindsten and Edson Aoki for valuable discussions. Funding from the Swedish Research Council (VR) for the Linnaeus center CADICS is gratefully acknowledged.

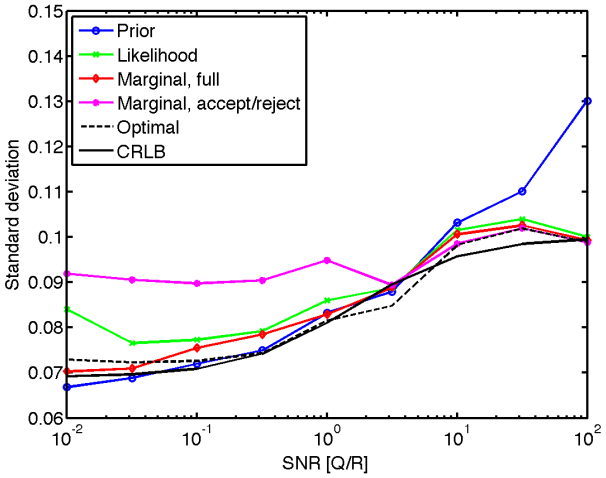

Fig. 1. RMSE for different proposal distributions.

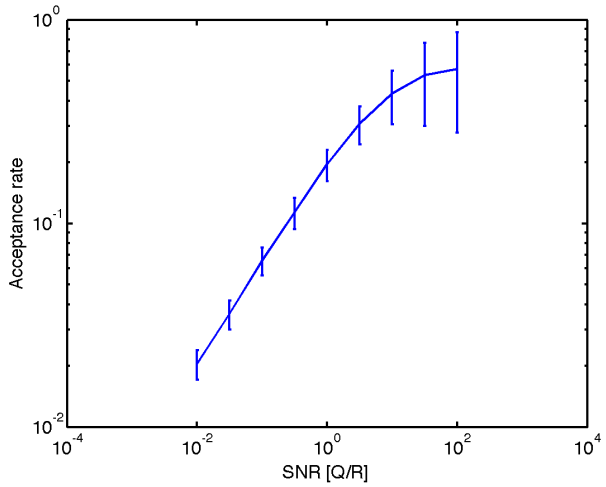

Fig. 2. Acceptance probability (with confidence interval) for likelihood proposal as a function of SNR.

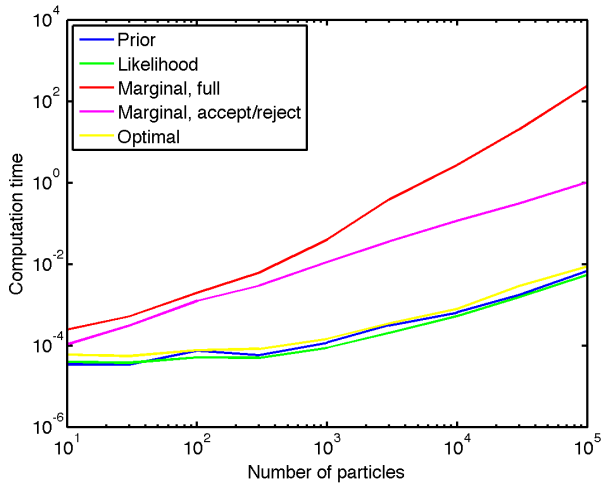

Fig. 3. Comparison of how the computation time depends on the number of particles for the different algorithms.

VI. Conclusions and Discussion

Marginalization of the particle filter posterior distribution has been proposed as one way to mitigate depletion problems. Straightforward marginalization has quadratic complexity. One approximation suggested in literature has $N\log(N)$ complexity. The purpose of this contribution was to find implementations with linear complexity, leading to an accept-reject algorithm.

There are several natural alternatives for the choice of proposal distribution (prior, likelihood, optimal), where one can try to marginalize (or smooth) the posterior distribution. However, there are several pitfalls for the user.

- If the most natural prior is used as proposal, then marginalization has no effect at all.
- If the optimal proposal is used, there is no need for marginalization, since the variance of the weights is constant. Thus, there is no depletion problem at all in this case (in theory).
- If the likelihood is used as proposal, then the simple example showed that marginalization gives a significant increase in performance for high SNR's. However, the algorithm presented here can be seen as a way to implement the optimal proposal.

A fundamental question remains to be answered: Does there exist a simple example where mPF outperforms PF?

References

[1] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, 2002.

[2] O. Cappé, S.J. Godsill, and E. Moulines. An overview of existing methods and recent advances in sequential Monte Carlo. *IEEE Proceedings*, 95:899, 2007.

[3] P.M. Djuric, J.H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M.F. Bugallo, and J. Miguez. Particle filtering. *IEEE Signal Processing Magazine*, 20:19, 2003.

[4] A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer Verlag, 2001.

[5] N.J. Gordon, D.J. Salmond, and A.F.M. Smith. A novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *IEE Proceedings on Radar and Signal Processing*, volume 140, pages 107–113, 1993.

[6] F. Gustafsson. Particle filter theory and practice with positioning applications. *IEEE Transactions on Aerospace and Electronics Magazine Part II: Tutorials*, 7(July):53–82, 2010.

[7] M. Isard and A. Blake. Condensation - conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28, 1998.

[8] G. Kitagawa. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1):1–25, 1996.

[9] M. Klaas. Toward practical n^2 Monte Carlo: The marginal particle filter. *Uncertainty in Artificial Intelligence*, 2005.

[10] J.S. Liu and R. Chen. Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93, 1998.

[11] R. Martinez-Cantin, N. de Freitas, and J.A. Castellanos. Analysis of particle methods for simultaneous robot localization and mapping and a new algorithm: Marginal-SLAM. In *Proceedings of IEEE International Conference on Robotics and Automation*, Roma, Italy, 2007.

[12] G. Poyiadjis, A. Doucet, and S.S. Singh. Maximum likelihood parameter estimation in general state-space models using particle methods. In *Proceedings of Joint Statistical Meeting*, Minneapolis, Minnesota, 2005.

[13] G. Poyiadjis, A. Doucet, and S.S. Singh. Maximum likelihood parameter estimation using particle methods. In *IEEE Conference on Acoustics, Speech and Signal Processing*, 2006.

[14] B. Ristic, S. Arulampalam, and N. Gordon. *Beyond the Kalman filter: Particle filters for tracking applications*. Artech House, London, 2004.

[15] S. Sing, N. Kantas, B. Vo, A. Doucet, and R.J. Evans. Simulation-based optimal sensor scheduling with application to observer trajectory planning. *Automatica*, 43:817–830, 2007.