Importance Densities for Particle Filtering using Iterated Conditional Expectations

Roland Hostettler, Filip Tronarp, Ángel F. García-Fernández, Simo Särkkä

DOI:

10.1109/LSP.2020.2964531

Importance Densities for Particle Filtering using Iterated Conditional Expectations

Roland Hostettler, *Member, IEEE,* Filip Tronarp, Ángel F. García-Fernández, and Simo Särkkä, *Senior Member, IEEE*

***Abstract*—In this letter, we consider Gaussian approximations of the optimal importance density in sequential importance sampling for nonlinear, non-Gaussian state-space models. The proposed method is based on generalized statistical linear regression and posterior linearization using conditional expectations. Simulation results show that the method outperforms the compared methods in terms of the effective sample size and provides a better local approximation of the optimal importance density.**

***Index Terms*—State estimation, particle filters, Monte Carlo methods, nonlinear systems, posterior linearization**

I. Introduction

Sequential Monte Carlo (SMC) methods are well-established sequential Bayesian inference methods. For example, particle filtering, which is an SMC method, is regularly employed for estimation in nonlinear, non-Gaussian dynamic systems in diverse applications [1]–[4]. In these methods, the state posterior density is approximated using a set of random samples (particles). Asymptotically (in the number of samples), this approximation converges to the true posterior, and hence, SMC methods provide an asymptotically exact solution [5], [6].

Since large numbers of samples induce high computational cost, it is preferable to use as few samples as possible. This requires sampling from an importance density which places the particles in the most likely parts of the state-space either using global or local importance densities [7], [8]. An intuitive, local approach is to sample from the dynamic model as in the bootstrap particle filter (BPF) [9]. However, this often requires a high number of samples. The optimal importance density (OID) that minimizes the particle weights' incremental variance uses the previous state as well as the latest measurement [10].

Unfortunately, it is often impossible to sample from the OID and a common approach is to use a Gaussian approximation instead [11]–[15]. These approaches use the same approximations as nonlinear Kalman filters such as the extended Kalman filter (Taylor series) or the unscented Kalman filter (unscented transform) to approximate the OID. However, they suffer from two important drawbacks. First, these approximations require that the measurement model can be written as a nonlinear transformation of the state and additive Gaussian noise. Second, the importance density may be far from the relevant area of the state-space, for example, for highly informative measurements [16], [17] or in likelihood-free inference problems [18].

The contribution of this letter is an importance sampling method that approximates the local OID based on posterior linearization and iterated conditional expectations [19], [20] with similarities to adaptive importance sampling [21], [22] and particle flow methods [23]–[25]. The proposed method finds a Gaussian OID approximation for nonlinear, non-Gaussian models without requiring the measurement to be a nonlinear transformation of the state with additive Gaussian noise and it is suitable for highly informative measurements. The properties of the proposed method are evaluated in two examples.

II. Problem Formulation

We consider inference in a latent Markovian state process $\{x_n \in \mathbb{R}^{d_x}\}_{n\geq 0}$ with noisy measurements $\{y_n \in \mathbb{R}^{d_y}\}_{n>0}$ that are conditionally independent given the states such that

$$x_0 \sim p(x_0), \tag{1a}$$
$$x_n \sim p(x_n \mid x_{n-1}), \tag{1b}$$
$$y_n \sim p(y_n \mid x_n). \tag{1c}$$

Here, n is the time step, $p(x_0)$ is the probability density function (pdf) of the initial state, $p(x_n \mid x_{n-1})$ is the dynamic model, and $p(y_n \mid x_n)$ is the likelihood. Also, we assume known and non-singular conditional means and covariances.

Based on model (1), we consider sequential importance sampling with resampling with the joint filtering posterior pdf

$$p(x_{1:n} \mid y_{1:n}) \propto p(y_n \mid x_n)p(x_n \mid x_{n-1})p(x_{1:n-1} \mid y_{1:n-1})$$

as the target distribution. This leads to a sample approximation with a set of J weighted particles $\{x^j_{1:n}, w^j_n\}^J_{j=1}$ as

$$p(x_{1:n} \mid y_{1:n}) \approx \sum_{j=1}^{J} w^j_n \delta(x_{1:n} - x^j_{1:n}), \tag{2}$$

where $\delta(\cdot)$ is the Kronecker delta function, $x^j_{1:n}$ is the jth trajectory sample, and w^j_n its importance weight.

The samples $x^j_{1:n}$ are drawn from an importance density $q(x_{1:n})$ which is typically chosen such that

$$q(x_{1:n} \mid y_{1:n}) = q(x_n \mid x_{n-1}, y_n)q(x_{1:n-1} \mid y_{1:n-1}).$$

In this case, the weights are updated sequentially using [26]

$$w_n \propto w_{n-1}\frac{p(y_n \mid x_n)p(x_n \mid x_{n-1})}{q(x_n \mid x_{n-1}, y_n)}, \tag{3}$$

R. Hostettler is with the Department of Engineering Sciences, Uppsala University, Sweden. (e-mail: roland.hostettler@angstrom.uu.se).

F. Tronarp and S. Särkkä are with the Department of Electrical Engineering and Automation, Aalto University, Finland. (e-mail: firstname.lastname@aalto.fi).

Á. F. García-Fernández is with the Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, U.K. and the ARIES research center, Universidad Antonio de Nebrija, Madrid, Spain. (e-mail: angel.garcia-fernandez@liverpool.ac.uk).

where the fraction in (3) is called the incremental weight.

To avoid sample impoverishment, the samples $x_{1:n}^j$ are resampled regularly according to their importance weights w_n^j, which is done either at fixed intervals or according to a resampling criterion such as the effective sample size [9], [26], [27]. This yields the sequential importance sampling with resampling type of particle filter (see, e.g., [2] for more details).

In order to minimize the variance of the incremental weights, new samples for x_n should be drawn according to the OID [26]

$$\begin{aligned} q(x_n \mid x_{n-1}, y_n) &= p(x_n \mid x_{n-1}, y_n) \\ &\propto p(y_n \mid x_n) p(x_n \mid x_{n-1}). \end{aligned} \tag{4}$$

In practice, the OID can only be sampled from in a few special cases. Instead, a common approach is to use a Gaussian approximation of (4) [11], [12], [26], [28]. In this case, we can make a Gaussian approximation of the joint density

$$p(x_n, y_n \mid x_{n-1}) \approx \mathcal{N}\left(\begin{bmatrix} x_n \\ y_n \end{bmatrix}, \begin{bmatrix} m_n^x \\ m_n^y \end{bmatrix}, \begin{bmatrix} P_n^x & P_n^{xy} \\ P_n^{yx} & P_n^y \end{bmatrix} \right) \tag{5}$$

and by conditioning (5) on y_n, we obtain

$$p(x_n \mid x_{n-1}, y_n) \approx \mathcal{N}(x_n; m_n^{x|y}, P_n^{x|y}), \tag{6a}$$

$$m_n^{x|y} = m_n^x + P_n^{xy}(P_n^y)^{-1}(y_n - m_n^y), \tag{6b}$$

$$P_n^{x|y} = P_n^x - P_n^{xy}(P_n^y)^{-1}(P_n^{xy})^{\mathsf{T}}. \tag{6c}$$

This joint approximation is particularly suitable for unimodal dynamic models, but can be poor in multimodal cases.

The moments of the joint approximation (5) can be calculated using Taylor series or sigma-points [29]. However, these are local approximations around x_{n-1} that do not take the measurement y_n into account and more importantly, these approaches may not work for certain types of non-Gaussian likelihoods. The aim of this letter is to find an approximation of the form (5) to approximate the OID for arbitrary models of the form (1) that overcomes these limitations. This is achieved by using generalized statistical linear regression (SLR) and iterated conditional expectations [19], [20].

III. Importance Density

A. Generalized SLR

Generalized SLR is a statistical linearization technique to approximate a nonlinear, non-Gaussian relationship between two random variables as an affine transformation that minimizes the mean squared error and also provides a measure of the linearization error [20]. Definition 1 below reviews generalized SLR with respect to the linearization density $\pi(x)$ for random variables x and y [20], [30]. Note that in Definition 1, the subscript π indicates that the corresponding expectation (or covariance) is with respect to the linearization density $\pi(x)$.

Definition 1 (Generalized SLR)**.** *Given the random variables $x \sim \pi(x)$ with moments $m_\pi^x \triangleq \mathrm{E}_\pi\{x\}$ and $P_\pi^x \triangleq \mathrm{Cov}_\pi\{x\}$, and $y \sim p(y \mid x)$ with conditional moments $\mathrm{E}\{y \mid x\}$ and $\mathrm{Cov}\{y \mid x\}$, the generalized SLR of y with respect to $\pi(x)$ is*

$$y \approx Ax + b + v, \tag{7a}$$

$$v \sim \mathcal{N}(0, \Omega), \tag{7b}$$

with

$$A = P_\pi^{yx}(P_\pi^x)^{-1}, \tag{8a}$$

$$b = m_\pi^y - A m_\pi^x, \tag{8b}$$

$$\Omega = P_\pi^y - A P_\pi^x A^{\mathsf{T}}, \tag{8c}$$

and where the moments are

$$m_\pi^y = \int \mathrm{E}\{y \mid x\} \pi(x) \mathrm{d}x, \tag{9a}$$

$$\begin{aligned} P_\pi^y = &\int (\mathrm{E}\{y \mid x\} - m_\pi^y)(\mathrm{E}\{y \mid x\} - m_\pi^y)^{\mathsf{T}} \pi(x) \mathrm{d}x \\ &+ \int \mathrm{Cov}\{y \mid x\} \pi(x) \mathrm{d}x, \end{aligned} \tag{9b}$$

$$P_\pi^{yx} = \int (\mathrm{E}\{y \mid x\} - m_\pi^y)(x - m_\pi^x)^{\mathsf{T}} \pi(x) \mathrm{d}x. \tag{9c}$$

In practice, the integrals in (9) can be solved analytically in a few special cases only. Instead, sigma-point or Taylor series approximations as in Gaussian filtering for nonlinear systems can be used [12], [29], [31].

B. OID Approximation

Generalized SLR in Definition 1 can now be used to find the OID approximation (5)–(6) for (nonlinear and non-Gaussian) models of the form (1) for each particle as follows. Given the particle x_{n-1}^j, the mean $m_n^{x,j} \triangleq \mathrm{E}\{x_n \mid x_{n-1}^j\}$ and covariance $P_n^{x,j} \triangleq \mathrm{Cov}\{x_n \mid x_{n-1}^j\}$ of the dynamic model $p(x_n \mid x_{n-1}^j)$, and using generalized SLR with respect to $\pi(x_n)$, the moments $m_n^{y,j}$, $P_n^{y,j}$, and $P_n^{xy,j}$ of (5) for the jth particle become

$$m_n^{y,j} = A^j m_n^{x,j} + b^j, \tag{10a}$$

$$P_n^{y,j} = A^j P_n^{x,j} (A^j)^{\mathsf{T}} + \Omega^j, \tag{10b}$$

$$P_n^{xy,j} = P_n^{x,j} (A^j)^{\mathsf{T}}. \tag{10c}$$

The linearization density $\pi(x_n)$ controls where the affine approximation is most accurate. Hence, a suitable linearization density is the pdf $p(x_n \mid x_{n-1}^j, y_n)$ such that the linearization is accurate around the OID for the jth particle. This results in the linearization of $\mathrm{E}\{y_n \mid x_n\}$ that minimizes the mean square error given x_{n-1}^j and y_n, also accounting for the linearization error [32]. However, since we are trying to approximate the OID to start with, this is not feasible. Instead, one can use the following iterative procedure [19], [20].

Assume that at the lth iteration, the mean and covariance of the posterior approximation from the $l-1$th iteration are given by $m_\pi^{x,l-1}$ and $P_\pi^{x,l-1}$, respectively. Then, the lth iteration's approximation is given by A^l, b^l, and Ω^l which are calculated using SLR with respect to the $l-1$th posterior approximation, see (8). Next, using A^l, b^l, Ω^l, and (10), $p(x_n \mid x_{n-1}^j, y_n)$ is approximated as in (6), which yields the lth iteration's conditional mean $m_n^{x|y,l}$ and covariance $P_n^{x|y,l}$. This posterior approximation is used as the linearization density in the $l+1$th iteration with $m_\pi^{x,l} = m_n^{x|y,l}$ and $P_\pi^{x,l} = P_n^{x|y,l}$.

The iterative procedure is initialized using SLR with respect to the dynamic model $p(x_n \mid x_{n-1}^j)$, and it is terminated either after a fixed number of iterations L or upon convergence. Suitable convergence criteria include, for example, the change in OID approximation mean or the step size falling below a

Algorithm 1 Iterative OID Approximation for Particle j

Set $m_\pi^{x,0} = \mathrm{E}\{x_n \mid x_{n-1}^j\}$, $P_\pi^{x,0} = \mathrm{Cov}\{x_n \mid x_{n-1}^j\}$, and $l \leftarrow 0$
do
  Set $l \leftarrow l + 1$
  Calculate $m_\pi^{y,l}$, $P_\pi^{y,l}$, and $P_\pi^{xy,l}$ using (9) and $m_\pi^{x,l-1}$, $P_\pi^{x,l-1}$
  Calculate A^l, b^l, and Ω^l using (8)
  Calculate $m_n^{x|y,l}$ and $P_n^{x|y,l}$ using (6) and (10)
  Set $m_\pi^{x,l} = m_n^{x|y,l}$ and $P_\pi^{x,l} = P_n^{x|y,l}$
while $l \leq L$ and not converged
Set $m_n^{x|y,j} = m_n^{x|y,L}$, $P_n^{x|y,j} = P_n^{x|y,L}$, and $q(x_n \mid x_{n-1}^j, y_n) \triangleq \mathcal{N}(x_n; m_n^{x|y,j}, P_n^{x|y,j})$

certain threshold, or, as proposed in [19], the Kullback–Leibler (KL) divergence between two consecutive OID approximations

$$D_{\mathrm{KL}}^l = \frac{1}{2}\left[\mathrm{tr}((P_n^{x|y,l})^{-1} P_n^{x|y,l-1}) - d_x - \log\left(\frac{|P_n^{x|y,l-1}|}{|P_n^{x|y,l}|}\right) + \|m_n^{x|y,l} - m_n^{x|y,l-1}\|^2_{\left(P_n^{x|y,l}\right)^{-1}}\right], \tag{11}$$

falling below a threshold $\epsilon > 0$ which is typically chosen between 1×10^{-5} and 0.1, see [19].

This yields the iterative conditional expectation importance density approximation method in Algorithm 1, where we may need to use Taylor series or sigma-points to approximate (9). Once the algorithm terminates, the OID approximation

$$p(x_n \mid x_{n-1}^j, y_n) \approx \mathcal{N}(x_n; m_n^{x|y,j}, P_n^{x|y,j}) \tag{12}$$

is obtained, which is then used for sampling.

Without iterations ($L = 1$), this method is a generalization of the Taylor series or sigma-point OID approximations to arbitrary likelihoods [10]–[12]. However, some advantages specifically come from the iterative re-linearization using the currently best posterior approximation [19]. Algorithm 1 is run for each particle and for at most L iterations, which yields an upper limit on the computational cost of $\mathcal{O}(LJ)$. In contrast, non-iterative importance densities scale according to $\mathcal{O}(J)$.

C. Convergence

1) Convergence to the OID: SLR with respect to the OID minimizes the mean squared error of the affine approximation to the conditional mean, also accounting for the linearization error. Local convergence proofs for Gaussian filters based on iterated SLR have been provided in [19], [20] and for additive Gaussian noise models, these can be seen as an approximate Kullback–Leibler minimization [19].

In particular, it has been shown that the iterative linearization approach converges to the local mode of the linearization density if it is initialized sufficiently close to that mode (see [20] for the specific conditions). Thus, if a particle is sufficiently close to a mode of the OID, the mean of the OID approximation for the corresponding particle will converge to this mode. This implies that multimodal posteriors do not pose a problem for the proposed algorithm as long as the posterior at the previous time step includes particles sufficiently close to the current time step's modes. However, due to this mode-seeking nature, heavy-tailed OIDs (and posteriors) may cause problems and generally can not be handled well by the proposed approach.

2) Filter Convergence: A sufficient (but not necessary) condition for mean squared error convergence of the filter is that the importance weights are upper bounded [6]. Thus provided that this condition holds for the model at hand, the proposed filter is guaranteed to converge. However, it has been shown that for Gaussian importance densities, the boundedness is not always guaranteed [33], which might or might not lead to non-convergence of the filter.

D. Practical Considerations

An important aspect to note is that using a Gaussian approximation of the form (6) may yield a poor importance density if the measurement covariance P_n^y becomes very small. This typically happens when the gradient of the measurement model approaches zero, together with small measurement noise (informative observations). Consequently, the linearization and thus also the importance density approximation become poor.

The iterations of the proposed method are able to correct this problem gradually and eventually, a good OID approximation is obtained again. However, this process typically takes many iterations and thus, in practice, it is preferable to prevent this from happening. In this case, either damped posterior linearization [34] or the following diagnostic can be employed.

Given the joint approximation (5), the test statistic

$$\gamma_n \triangleq \|y_n - m_n^{y,l}\|^2_{\left(P_n^{y,l}\right)^{-1}}, \tag{13}$$

where $m_n^{y,l}$ and $P_n^{y,l}$ are the predicted measurement mean and covariance after the lth iteration, respectively, can be used to evaluate the goodness of the linearization. Under the Gaussian assumption, γ_n follows a χ^2 distribution with d_y degrees of freedom. Thus, if γ_n exceeds a threshold γ_T, the iterations can be stopped and the last valid approximation can be used instead. A suitable threshold γ_T is obtained by choosing a tail probability (e.g., $\kappa = 0.05$) followed by evaluating the inverse cumulative distribution function of the χ^2 distribution at $1 - \kappa$.

IV. Numerical Illustrations

We evaluate the method in two examples and compare it to the BPF, one-step OID approximation [11], [12], Gaussian flow (GFPF) OID approximation [24], and the particle flow particle filter (PFPF) with local particle flow [25]. We run 100 Monte Carlo simulations with 100 time steps. Details of the models and parameters are given in the Supplementary Material.

A. Univariate Nonlinear Growth Model

First, we consider the univariate nonlinear growth model [9], [35], a benchmark problem for nonlinear filtering algorithms. We use the standard parametrization from the literature, but assume the measurement variance to be 1×10^{-2}, which is two magnitudes lower than commonly considered. Despite being an univariate problem, there are two particular challenges: First, due to the square in the observation model, no information about the state's sign is measured, which gives rise to a bi-modal posterior for large values of the state. Second, the low measurement noise causes the posterior to be peaky, which is well-known to be challenging for sampling-based inference.

TABLE I
MEAN AND STANDARD DEVIATION OF THE PERFORMANCE INDICATORS FOR THE COMPARED METHODS (MULTIVARIATE RICKER POPULATION MODEL).

Method	J	RMSE	Time / s	Resampling / %	ESS	ESS / %
BPF	2 500	3.5 (±0.5)	0.99 (±0.1)	98.1 (±0.96)	101 (±16)	4.0 (±0.6)
BPF	10 000	3.4 (±0.8)	3.2 (±0.3)	98.0 (±1.0)	410 (±73)	4.1 (±0.7)
BPF	50 000	3.3 (±0.5)	11.5 (±0.7)	98.2 (±0.89)	1 975 (±331)	4.0 (±0.7)
One-step	250	3.6 (±0.6)	4.9 (±0.4)	93.9 (±0.1)	27.3 (±2.8)	10.9 (±1.1)
One-step	500	3.5 (±0.5)	9.8 (±0.8)	93.7 (±0.1)	52.5 (±5.2)	10.5 (±1.0)
GFPF	100	3.3 (±0.5)	80.7 (±0.9)	98.8 (±0.4)	15.7 (±0.7)	15.7 (±0.7)
GFPF	250	3.5 (±0.6)	204.6 (±7.8)	99 (±0.1)	34.5 (±1.45)	13.8 (±0.6)
PFPF	100	3.3 (±0.5)	11.3 (±0.7)	53.8 (±3.0)	34.7 (±1.4)	34.7 (±1.4)
PFPF	250	3.3 (±0.5)	27.8 (±2.5)	59.6 (±2.8)	78.8 (±3.4)	31.5 (±1.4)
Proposed	100	3.7 (±0.6)	4.7 (±0.5)	49 (±3.0)	37 (±1.4)	37 (±1.4)
Proposed	250	3.6 (±0.5)	10.5 (±0.5)	51.8 (±2.7)	88.1 (±3.4)	35.2 (±1.4)
Proposed	500	3.4 (±0.4)	21.7 (±1.2)	53.7 (±2.9)	170 (±8.4)	34.1 (±1.8)

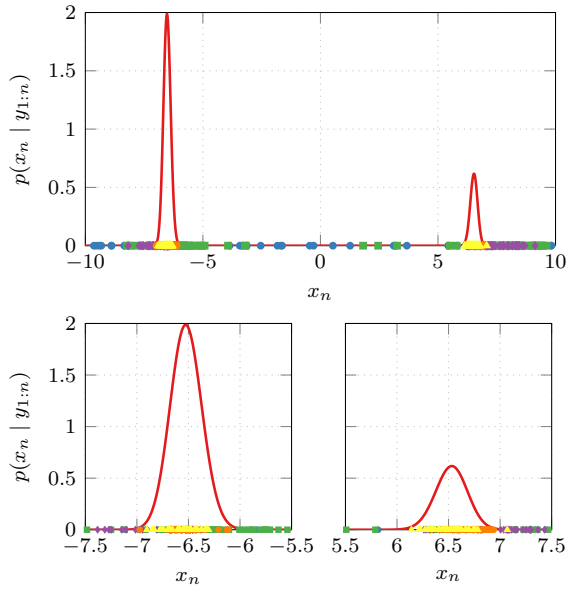

Fig. 1. True posterior pdf (——) and particle locations for the BPF (●), one-step OID approximation (■), GFPF (♦), PFPF (●), and the proposed method (▲). Top: Full posterior, bottom: detailed view of the two modes.

For the proposed method, we use $\epsilon = 1 \times 10^{-2}$ and $L = 5$. For the GFPF, we use 5 equally spaced integration steps and for the PFPF, we use the parameters from [25].

Fig. 1 shows the true posterior (calculated using a dense grid filter) together with the particle locations of the filters for time step $n = 1$ and $J = 100$. Fig. 1 (top) shows the complete posterior that exhibits two narrow modes which are difficult to approximate using particles. Most of the BPF's and one-step OID approximation's particles are where the posterior is close to zero, whereas the GFPF is somewhat better but still has a significant number of particles far from the modes. In contrast, the detailed views of the modes (Fig. 1, bottom) show that only the PFPF and the proposed method find both modes and concentrate the particles in the corresponding areas. Also, the BPF, as well as one-step approximation and GFPF only have few particles close to the smaller mode, which causes these filters to loose track of that mode after a few more time steps. Further results are provided in the Supplementary Material.

B. Multivariate Ricker Population Model

In this example, we consider a multivariate Ricker model for interacting populations [36]. The model includes migration between different localized populations of a species as well as population- and species-level stochastic effects. We consider a 10-dimensional version of the model together with a left-skewed, heavy-tailed, generalized Poisson likelihood [37], [38]. The aim of this example is to compare the sampling effectiveness of the filters. Hence, we choose the number of particles such that approximately the same (asymptotic) root mean squared error (RMSE) is achieved. Furthermore, we chose different particle numbers to evaluate the effect on the computational time. In this example, we use 10 integration steps for the GFPF and the parameters as in [25] for the PFPF.

The results are shown in Table I. The proposed method is able to keep the highest relative effective sample size (ESS) and the lowest resampling rate, followed by the PFPF, the GFPF and one-step OID approximations. Hence, the proposed method is able to sample in the high probability areas of the state-space and produce a high fraction of high weight samples. Note that the GFPF is considerably slower as it requires solving a Sylvester equation for each particle and integration step.

Despite the better placement of the particles of the proposed method, in terms of computational complexity and absolute ESS, the BPF performs best in this example, mainly due to the possibility of efficiently implementing the weight calculations for this model. However, advantages of using a low number of particles in a particle filter include lower memory requirements for path storage [39] and lower computational requirements when a smoothing pass follows [40], [41]. In this case, the proposed method is preferable as it achieves similar performance at considerably lower sample size.

V. CONCLUSIONS

In this paper, a Gaussian approximation to the OID for SMC methods for nonlinear, non-Gaussian state-space models has been presented. The method is able to efficiently sample in the high-probability regions of the state-space, as shown by the high relative ESS and low number of resampling steps, and is particularly suitable for models with unimodal dynamic models. Possible extensions of the method include its use in auxiliary particle filtering [42] and SMC for static models [43], [44], or to make use of Rao–Blackwellized SLR [45], [46].

REFERENCES

[1] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P.-J. Nordlund, "Particle filters for positioning, navigation, and tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 425–437, February 2002.

[2] M. 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*, vol. 50, no. 2, pp. 174–188, February 2002.

[3] P. M. Djurić, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Miguez, "Particle filtering," *IEEE Signal Processing Magazine*, vol. 20, no. 5, pp. 19–38, September 2003.

[4] C. Schäfer and N. Chopin, "Sequential Monte Carlo on large binary sampling spaces," *Statistics and Computing*, vol. 23, no. 2, pp. 163–184, March 2013.

[5] R. Douc and E. Moulines, "Limit theorems for weighted samples with applications to sequential Monte Carlo methods," *The Annals of Statistics*, vol. 36, no. 5, pp. 2344–2376, October 2008.

[6] D. Crisan and A. Doucet, "A survey of convergence results on particle filtering methods for practitioners," *IEEE Transactions on Signal Processing*, vol. 50, no. 3, pp. 736–746, March 2002.

[7] J. Cornebise, É. Moulines, and J. Olsson, "Adaptive methods for sequential importance sampling with application to state space models," *Statistics and Computing*, vol. 18, no. 4, pp. 461–480, December 2008.

[8] ——, "Adaptive sequential Monte Carlo by means of mixture of experts," *Statistics and Computing*, vol. 24, no. 3, pp. 317–337, May 2014.

[9] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *Radar and Signal Processing, IEE Proceedings F*, vol. 140, no. 2, pp. 107–113, April 1993.

[10] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.

[11] R. van der Merwe, A. Doucet, N. de Freitas, and E. A. Wan, "The unscented particle filter," in *Advances in Neural Information Processing Systems*, 2001, pp. 584–590.

[12] S. Saha, P. K. Mandal, Y. Boers, H. Driessen, and A. Bagchi, "Gaussian proposal density using moment matching in SMC methods," *Statistics and Computing*, vol. 19, no. 2, pp. 203–208, June 2009.

[13] N. Oudjane and C. Musso, "Progressive correction for regularized particle filters," in *3rd International Conference on Information Fusion (FUSION)*, vol. 2, July 2000.

[14] J. Q. Smith and A. A. F. Santos, "Second-order filter distribution approximations for financial time series with extreme outliers," *Journal of Business & Economic Statistics*, vol. 24, no. 3, pp. 329–337, 2006.

[15] P. Bunch and S. Godsill, "Particle filtering with progressive Gaussian approximations to the optimal importance density," in *5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, December 2013, pp. 360–363.

[16] P. Del Moral and L. M. Murray, "Sequential Monte Carlo with highly informative observations," *SIAM/ASA Journal on Uncertainty Quantification*, vol. 3, no. 1, pp. 969–997, 2015.

[17] A. Svensson, T. B. Schön, and F. Lindsten, "Learning of state-space models with highly informative observations: A tempered sequential Monte Carlo solution," *Mechanical Systems and Signal Processing*, vol. 104, pp. 915–928, 2018.

[18] J.-M. Marin, P. Pudlo, C. P. Robert, and R. J. Ryder, "Approximate Bayesian computational methods," *Statistics and Computing*, vol. 22, no. 6, pp. 1167–1180, November 2012.

[19] A. F. García-Fernández, L. Svensson, M. R. Morelande, and S. Särkkä, "Posterior linearization filter: Principles and implementation using sigma points," *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5561–5573, October 2015.

[20] F. Tronarp, A. F. García-Fernández, and S. Särkkä, "Iterative filtering and smoothing in nonlinear and non-Gaussian systems using conditional moments," *IEEE Signal Processing Letters*, vol. 25, no. 3, pp. 408–412, March 2018.

[21] M. F. Bugallo, V. Elvira, L. Martino, D. Luengo, J. Míguez, and P. M. Djurić, "Adaptive importance sampling: The past, the present, and the future," *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 60–79, July 2017.

[22] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, "Generalized multiple importance sampling," *Statistical Science*, vol. 34, no. 1, pp. 129–155, February 2019.

[23] S. Reich, "A guided sequential Monte Carlo method for the assimilation of data into stochastic dynamical systems," in *Recent Trends in Dynamical Systems*, A. Johann, H.-P. Kruse, F. Rupp, and S. Schmitz, Eds. Basel, Switzerland: Springer, 2013, pp. 205–220.

[24] P. Bunch and S. J. Godsill, "Approximations of the optimal importance density using Gaussian particle flow importance sampling," *Journal of the American Statistical Association*, vol. 111, no. 514, pp. 748–762, 2016.

[25] Y. Li and M. Coates, "Particle filtering with invertible particle flow," *IEEE Transactions on Signal Processing*, vol. 65, no. 15, pp. 4102–4116, August 2017.

[26] A. Doucet and A. M. Johansen, "A tutorial on particle filtering and smoothing: Fifteen years later," in *Handbook of Nonlinear Filtering*, ser. Oxford Handbooks, D. Crisan and B. Rozovskii, Eds. Oxford, UK: Oxford University Press, 2011, vol. 12, pp. 656–704.

[27] T. Li, M. Bolić, and P. M. Djurić, "Resampling methods for particle filtering: Classification, implementation, and strategies," *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 70–86, May 2015.

[28] R. Hostettler and T. B. Schön, "Auxiliary-particle-filter-based two-filter smoothing for Wiener state-space models," in *21th International Conference on Information Fusion (FUSION)*, Cambridge, UK, July 2018.

[29] S. Särkkä, *Bayesian Filtering and Smoothing*. Cambridge, UK: Cambridge University Press, 2013.

[30] I. Arasaratnam, S. Haykin, and R. J. Elliott, "Discrete-time nonlinear filtering algorithms using Gauss-Hermite quadrature," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 953–977, May 2007.

[31] M. Roth, G. Hendeby, and F. Gustafsson, "Nonlinear Kalman filters explained: A tutorial on moment computations and sigma point methods," *Journal of Advances in Information Fusion*, vol. 11, no. 1, pp. 47–70, 2016.

[32] A. F. García-Fernández, F. Tronarp, and S. Särkkä, "Gaussian target tracking with direction-of-arrival von Mises–Fisher measurements," *IEEE Transactions on Signal Processing*, vol. 67, no. 11, pp. 2960–2972, June 2019.

[33] J. Kokkala and S. Särkkä, "On the (non-)convergence of particle filters with gaussian importance distributions," in *17th IFAC Symposium on System Identification SYSID 2015*, vol. 48, no. 28, 2015, pp. 793–798.

[34] M. Raitoharju, L. Svensson, A. F. García-Fernández, and R. Piché, "Damped posterior linearization filter," *IEEE Signal Processing Letters*, vol. 25, no. 4, pp. 536–540, April 2018.

[35] G. Kitagawa, "Non-Gaussian state-space modeling of nonstationary time series," *Journal of the American Statistical Association*, vol. 82, no. 400, pp. 1032–1041, 1987.

[36] M. Heino, V. Kaitala, E. Ranta, and J. Lindström, "Synchronous dynamics and rates of extinction in spatially structured populations," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 264, no. 1381, pp. 481–486, 1997.

[37] P. C. Consul and G. C. Jain, "A generalization of the Poisson distribution," *Technometrics*, vol. 15, no. 4, pp. 791–799, 1973.

[38] P. C. Consul and F. Famoye, "Generalized Poisson regression model," *Communications in Statistics - Theory and Methods*, vol. 21, no. 1, pp. 89–109, 1992.

[39] P. E. Jacob, L. M. Murray, and S. Rubenthaler, "Path storage in the particle filter," *Statistics and Computing*, vol. 25, no. 2, pp. 487–496, March 2015.

[40] S. J. Godsill, A. Doucet, and M. West, "Monte Carlo smoothing for nonlinear time series," *Journal of the American Statistical Association*, vol. 99, no. 465, pp. 156–168, 2004.

[41] F. Lindsten and T. B. Schön, "Backward simulation methods for Monte Carlo statistical inference," *Foundations and Trends in Machine Learning*, vol. 6, pp. 1–143, 2013.

[42] M. K. Pitt and N. Shephard, "Filtering via simulation: Auxiliary particle filters," *Journal of the American Statistical Association*, vol. 94, no. 446, pp. 590–599, 1999.

[43] N. Chopin, "A sequential particle filter method for static models," *Biometrika*, vol. 89, no. 3, pp. 539–552, 08 2002.

[44] P. Del Moral, A. Doucet, and A. Jasra, "An adaptive sequential Monte Carlo method for approximate Bayesian computation," *Statistics and Computing*, vol. 22, no. 5, pp. 1009–1020, September 2012.

[45] T. Schön, F. Gustafsson, and P.-J. Nordlund, "Marginalized particle filters for mixed linear/nonlinear state-space models," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2279–2289, July 2005.

[46] R. Hostettler and S. Särkkä, "Rao–Blackwellized Gaussian smoothing," *IEEE Transactions on Automatic Control*, vol. 64, no. 1, pp. 302–309, January 2019.