Statistics and Probability Letters 203 (2023) 109914

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Properties of marginal sequential Monte Carlo methods

Francesca R. Crucinio [a,*], Adam M. Johansen [a,b]

[a] Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK
[b] The Alan Turing Institute, British Library, 96 Euston Road, London, NW1 2DB, UK

ARTICLE INFO

Article history:
Received 6 March 2023
Received in revised form 9 June 2023
Accepted 28 July 2023
Available online 11 August 2023

MSC:
60F05
65C05
60F15
62F15

Keywords:
Asymptotic variance
Central limit theorem
Law of large numbers
Particle filters

ABSTRACT

We provide a framework which admits a number of "marginal" sequential Monte Carlo (SMC) algorithms as particular cases — including the marginal particle filter (Klaas et al., 2005), the independent particle filter (Lin et al., 2005) and linear-cost Approximate Bayesian Computation SMC (Sisson et al., 2007). We provide conditions under which such algorithms obey laws of large numbers and central limit theorems and provide some further asymptotic characterizations. Finally, it is shown that the asymptotic variance of a class of estimators associated with certain marginal SMC algorithms is never greater than that of the estimators provided by a standard SMC algorithm using the same proposal distributions.

1. Introduction

Sequential Monte Carlo (SMC) methods are a class of Monte Carlo methods which approximate a sequence of distributions and integrals with respect to those distributions using a population of weighted samples (or particles) which evolve according to a combination of mutation and selection dynamics. Such methods became popular in the context of filtering for state–space models in the engineering and statistics literature following the seminal work of Gordon et al. (1993), and have been extensively studied from a theoretical perspective as mean field approximations of a Feynman–Kac flow since (Del Moral, 1996); see Chopin and Papaspiliopoulos (2020) for a thorough treatment of these methods.

SMC methods approximate a sequence of distributions $(\hat{\eta}_n)_{n\geq 0}$ defined on Polish spaces $(E^n, \mathcal{E}^n)$, where $\mathcal{E}$ denotes the σ-field associated with E, of increasing dimension with the relationship

$$\hat{\eta}_n(\mathrm{d}x_{1:n}) \propto U_n(x_{n-1}, x_n) K_n(x_{n-1}, \mathrm{d}x_n) \hat{\eta}_{n-1}(\mathrm{d}x_{1:n-1}) \tag{1}$$

for some Markov kernels $K_n : E \times \mathcal{E} \to [0, 1]$ and non-negative functions $U_n : E \times E \to \mathbb{R}$.

In this work we focus on a particular class of SMC methods, in which the past evolution of the process is marginalized out. Marginal sequential Monte Carlo (MSMC) methods have been introduced in Klaas et al. (2005), Lin et al. (2005) where they were shown empirically to improve upon standard SMC in terms of the conditional variance of their unnormalized importance weights.

Despite being a popular class of algorithms which provides good results in practice (e.g. Sisson et al., 2007; Didelot et al., 2011; Everitt et al., 2017; Poyiadjis et al., 2011), the theoretical properties of MSMC are less studied. This paper

* Corresponding author.
E-mail address: francesca.crucinio@gmail.com (F.R. Crucinio).

https://doi.org/10.1016/j.spl.2023.109914

aims to fill this gap. Our main contribution is a comprehensive analysis of the limiting behaviour of MSMC that shows that MSMC shares many of the key properties of standard SMC methods (see, e.g. Del Moral, 2004). We establish that the estimates provided by MSMC methods obey laws of large numbers (Propositions 1 and 3), a central limit theorem (Proposition 6) and show that their $\mathbb{L}_p$ errors decay at the usual $N^{-1/2}$ rate (Proposition 2) and that the bias decays at rate N^{-1} (Proposition 5). We derive the asymptotic variance of some popular classes of MSMC methods and show that the asymptotic variance of the estimates obtained with marginal SMC is never larger than that one would obtain with standard SMC using the same proposals (Corollaries 1 and 2).

We introduce marginal sequential Monte Carlo methods in Section 2. Our main contributions are in Section 3 in which we present the theoretical properties of these algorithms and show that their asymptotic variance is never greater than that of standard SMC. Proofs of the results are postponed to Appendix D to Appendix H. We close the paper with a discussion in Section 4.

2. Marginal sequential Monte Carlo methods

Marginal sequential Monte Carlo methods (MSMC) are a class of sequential Monte Carlo algorithms introduced in Klaas et al. (2005) to approximate a sequence of distributions $(\hat{\eta}_n)_{n\geq 0}$ defined on measurable spaces $(E, \mathcal{E})$, where $\mathcal{E}$ denotes the σ-field associated with E. The sequence of distributions satisfies

$$\hat{\eta}_n(\mathrm{d}x_n) \propto \int U_n(x_{n-1}, x_n) K_n(x_{n-1}, \mathrm{d}x_n) \hat{\eta}_{n-1}(\mathrm{d}x_{n-1}), \tag{2}$$

with K_n, U_n as in (1) and $\hat{\eta}_0(\mathrm{d}x_0) \propto U_0(x_0)K_0(\mathrm{d}x_0)$.

Contrary to standard SMC algorithms, in which a cloud of particles is used to approximate the measure-valued recursion (1) defined over $(E^n, \mathcal{E}^n)$ with approximations of time marginals obtained by discarding part of the sampled paths, MSMC algorithms deal directly with the marginal recursion and the presence of the integral w.r.t. $\hat{\eta}_{n-1}$ in (2) requires an additional approximation. Hence, as described below, in MSMC a sample approximation of the integral w.r.t. $\hat{\eta}_{n-1}$ is used to define an alternative sequence of targets.

Before describing MSMC algorithms, we describe one step of the idealized algorithms targeting (2) which MSMC approximates. At time n, we start with a cloud of equally weighted particles $\{\widetilde{X}^i_{n-1}, 1/N\}_{i=1}^N$ approximating $\hat{\eta}_{n-1}$. In the idealized algorithm, the particles are propagated forward in time to approximate what is sometimes known as the predictive or mutated distribution:

$$\eta_n(\mathrm{d}x_n) = \hat{\eta}_{n-1}M_n(\mathrm{d}x_n) := \int M_n(x_{n-1}, \mathrm{d}x_n)\hat{\eta}_{n-1}(\mathrm{d}x_{n-1}) \tag{3}$$

for $n \geq 1$ and some Markov kernels $M_n : E \times \mathcal{E} \to [0, 1]$ to obtain a new cloud of particles $\{X^i_n, 1/N\}_{i=1}^N$. The new cloud of particles is used as proposal in an importance sampling step targeting $\hat{\eta}_n$, with importance weights proportional to the Radon–Nikodym derivative $\mathrm{d}\hat{\eta}_n/\mathrm{d}\eta_n$ (which corresponds to a simple density ratio when both measures admit densities)

$$G_n(x_n) = \frac{\mathrm{d}\hat{\eta}_{n-1}(U_n \cdot K_n)}{\mathrm{d}\hat{\eta}_{n-1}M_n}(x_n) = \frac{\mathrm{d}\left(\int U_n(x_{n-1}, \cdot)K_n(x_{n-1}, \cdot)\hat{\eta}_{n-1}(\mathrm{d}x_{n-1})\right)}{\mathrm{d}\left(\int M_n(x_{n-1}, \cdot)\hat{\eta}_{n-1}(\mathrm{d}x_{n-1})\right)}(x_n), \tag{4}$$

where $(U_n \cdot K_n)(x_{n-1}, \mathrm{d}x_n) := U_n(x_{n-1}, x_n)K_n(x_{n-1}, \mathrm{d}x_n)$, to obtain a weighted cloud of particles $\{X^i_n, W^i_n\}_{i=1}^N$, with $W^i_n \propto G_n(X^i_n)$, which approximates $\hat{\eta}_n$. The equally weighted population approximating $\hat{\eta}_n$, $\{\widetilde{X}^i_n, \frac{1}{N}\}_{i=1}^N$, is obtained by resampling from $\{X^i_n, W^i_n\}_{i=1}^N$, see e.g. Gerber et al. (2019) for a recent review of popular resampling schemes.

For any probability measure, η, and non-negative integrable function G, we introduce the operator $\Psi_G(\eta)(\mathrm{d}x) := G(x)\eta(\mathrm{d}x)/\int G(x')\eta(\mathrm{d}x')$, which denotes the application of an importance sampling step with weights proportional to G to the measure η. With this notation we can write $\hat{\eta}_n = \Psi_{G_n}(\eta_n)$ and, for a cloud of weighted particles $\{X^i_n, W^i_n\}_{i=1}^N$ with $W^i_n \propto G_n(X^i_n)$, $\Psi_{G_n}(\eta^N_n)(\mathrm{d}x_n) = \sum_{i=1}^N W^i_n \delta_{X^i_n}(\mathrm{d}x_n)$.

The target distribution (2) involves an intractable integral w.r.t. $\hat{\eta}_{n-1}$. Using the notation above we define the approximate targets

$$\Psi_{G_{n-1}}(\eta^N_{n-1})(U_n \cdot K_n)(\mathrm{d}x_n) \propto \sum_{i=1}^N W^i_{n-1} U_n(X^i_{n-1}, x_n) K_n(X^i_{n-1}, \mathrm{d}x_n).$$

We then proceed as in the idealized algorithm, and, since the integrals w.r.t. $\hat{\eta}_{n-1}$ are intractable, replace $\hat{\eta}_{n-1}$ with its particle approximation obtained at time $n-1$. Thus, the particles are propagated forward in time using an approximation of the proposals in (3), $\eta_n(\mathrm{d}x_n) \approx \sum_{i=1}^N W^i_{n-1} M_n(X^i_{n-1}, \mathrm{d}x_n)$, and resampled using approximate weights

$$G^N_n(x_n) = \frac{\mathrm{d}\Psi_{G^N_{n-1}}(\eta^N_{n-1})(U_n \cdot K_n)}{\mathrm{d}\Psi_{G^N_{n-1}}(\eta^N_{n-1})M_n}(x_n) = \frac{\mathrm{d}\left(\sum_{i=1}^N W^i_{n-1} U_n(X^i_{n-1}, \cdot)K_n(X^i_{n-1}, \cdot)\right)}{\mathrm{d}\left(\sum_{i=1}^N W^i_{n-1} M_n(X^i_{n-1}, \cdot)\right)}(x_n). \tag{5}$$

The resulting MSMC method is summarized in Algorithm 1. For convenience, we identify the three fundamental steps of Algorithm 1 as a *mutation* step (Algorithm 7), *reweighting* steps (Algorithm) and a *resampling* step (Algorithm 10). To each step, we associate a measure and its corresponding particle approximation: the mutated measure η_n in (3) is approximated by $\eta_n^N := N^{-1}\sum_{i=1}^N \delta_{X_n^i}$ obtained after Algorithm 7, Algorithm provide a particle approximation of $\hat{\eta}_n$, after resampling we obtain another approximation of $\hat{\eta}_n$ in (2), $\hat{\eta}_n^N := N^{-1}\sum_{i=1}^N \delta_{\tilde{X}_n^i}$.

Algorithm 1 Marginal Sequential Monte Carlo (MSMC)

1: Set $n = 0$.
2: Sample $X_0^i \sim M_0$ for $i = 1, \dots, N$.
3: Compute $G_0(X_0^i) = \frac{\mathrm{d}\hat{\eta}_0}{\mathrm{d}M_0}(X_0^i)$ $i = 1, \dots, N$.
4: Compute $W_0^i = G_0(X_0^i)/\sum_{j=1}^N G_0(X_0^j)$.
5: Resample $\left\{X_0^i, W_0^i\right\}_{i=1}^N$ to obtain $\{\tilde{X}_0^i, \frac{1}{N}\}_{i=1}^N$.
6: Update $n \leftarrow n+1$.
7: Sample $X_n^i \sim M_n\left(\tilde{X}_{n-1}^i, \cdot\right)$ for $i = 1, \dots, N$.
8: Compute $G_n^N(X_n^i)$ in (5) for $i = 1, \dots, N$.
9: Compute $W_n^i = G_n^N(X_n^i)/\sum_{j=1}^N G_n^N(X_n^j)$.
10: Resample $\left\{X_n^i, W_n^i\right\}_{i=1}^N$ to obtain $\{\tilde{X}_n^i, \frac{1}{N}\}_{i=1}^N$.
11: Go to Line 6.

2.1. Examples of Marginal SMC

The class of marginal SMC methods encompasses a number of well-known algorithms, we briefly discuss some examples of interest below. Similar algorithms also appear in Poyiadjis et al. (2011), Everitt et al. (2017), Li et al. (2016).

2.1.1. Marginal particle filters

Marginal particle filters (MPF; Klaas et al., 2005) are a class of algorithms to perform inference on state space models (SSM), a family of time series models consisting of two discrete-time processes: a latent process $(X_n)_{n\geq 0}$ and conditionally independent observations $(Y_n)_{n\geq 1}$. Such a SSM $(X_n, Y_n)_{n\geq 0}$ is defined by the transition density $f_n(x_n|x_{n-1})$ of the latent process, with the convention that $f_0(x_0|x_{-1}) \equiv f_0(x_0)$, and the observation likelihood $g_n(y_n|x_n)$.

In this case, the target distribution is the filtering distribution

$$\hat{\eta}_n(\mathrm{d}x_n) \equiv p(x_n|y_{1:n})\mathrm{d}x_n \propto g_n(y_n|x_n)\int f_n(x_n|x_{n-1})\hat{\eta}_{n-1}(\mathrm{d}x_{n-1})\cdot \mathrm{d}x_n, \tag{6}$$

with $\hat{\eta}_0(\mathrm{d}x_0) = f_0(x_0)\mathrm{d}x_0$, $K_n(x_{n-1}, \mathrm{d}x_n) = f_n(x_n|x_{n-1})\mathrm{d}x_n$ and $U_n(x_{n-1}, x_n) \equiv U_n(x_n) = g_n(y_n|x_n)$, and proposal density which can incorporate the observation y_n, $M_n(x_{n-1}, \mathrm{d}x_n) = q_n(x_n|y_n, x_{n-1})\mathrm{d}x_n$. The corresponding weights are given by

$$G_n(x_n) = g_n(y_n|x_n)\frac{\int f_n(x_n|x_{n-1})\hat{\eta}_{n-1}(\mathrm{d}x_{n-1})}{\int q_n(x_n|y_n, x_{n-1})\hat{\eta}_{n-1}(\mathrm{d}x_{n-1})}; G_n^N(x_n) = g_n(y_n|x_n)\frac{\sum_{i=1}^N W_{n-1}^i f_n(x_n|X_{n-1}^i)}{\sum_{i=1}^N W_{n-1}^i q_n(x_n|y_n, X_{n-1}^i)}.$$

Independent particle filters (IPF; Lin et al., 2005) are a particular case of MPF with $M_n(x_{n-1}, \mathrm{d}x_n) = q_n(x_n|y_n)\mathrm{d}x_n$ (see also Lai et al., 2022, Appendix C); when $q_n \equiv f_n$ MPF collapse onto the bootstrap particle filter (Gordon et al., 1993)

2.1.2. Marginal auxiliary particle filters

Marginal auxiliary particle filters (MAPFs) are a variant of MPF introduced in Klaas et al. (2005) (see also Fearnhead (1998, page 59)), as a marginalized version of standard auxiliary particle filters (Pitt and Shephard, 1999; Carpenter et al., 1999; see also Elvira et al., 2019 for a different perspective). An MAPF can be described as a standard MPF applied to $\hat{\eta}_0(\mathrm{d}x_0) \propto f_0(x_0)\tilde{p}(y_1|x_0)\mathrm{d}x_0$ and

$$\begin{aligned}\hat{\eta}_n(\mathrm{d}x_n) \propto \tilde{p}(y_{n+1}|x_n)p(x_n|y_{1:n})\mathrm{d}x_n &\propto \tilde{p}(y_{n+1}|x_n)g(y_n|x_n)\left(\int f_n(x_n|x_{n-1})p(x_{n-1}|y_{1:n-1})\mathrm{d}x_{n-1}\right)\mathrm{d}x_n \\ &\propto \tilde{p}(y_{n+1}|x_n)g(y_n|x_n)\left(\int \frac{f_n(x_n|x_{n-1})}{\tilde{p}(y_n|x_{n-1})}\hat{\eta}_{n-1}(\mathrm{d}x_{n-1})\right)\mathrm{d}x_n,\end{aligned} \tag{7}$$

where $\tilde{p}(y_{n+1}|x_n)$ is an approximation of $p(y_{n+1}|x_n) := \int g_{n+1}(y_{n+1}|x_{n+1})f_{n+1}(x_{n+1}|x_n)\mathrm{d}x_{n+1}$, to which an importance sampling step is added to guarantee that we are targeting the correct distribution $p(x_n|y_{1:n})$ (Johansen and Doucet, 2008).

Setting $K_n(x_{n-1}, \mathrm{d}x_n) = f_n(x_n|x_{n-1})\mathrm{d}x_n$ and $U_n(x_{n-1}, x_n) = g_n(y_n|x_n)\tilde{p}(y_{n+1}|x_n)/\tilde{p}(y_n|x_{n-1})$, one can apply Algorithm 1 with proposal kernel $M_n(x_{n-1}, \mathrm{d}x_n) = q_n(x_n|x_{n-1}, y_n)\mathrm{d}x_n$ which incorporates the current observation so that the weights are

$$\begin{aligned}G_n(x_n) &= g_n(y_n|x_n)\frac{\tilde{p}(y_{n+1}|x_n)\int f_n(x_n|x_{n-1})/\tilde{p}(y_n|x_{n-1})\hat{\eta}_{n-1}(\mathrm{d}x_{n-1})}{\int q_n(x_n|x_{n-1}, y_n)\hat{\eta}_{n-1}(\mathrm{d}x_{n-1})}, \\ G_n^N(x_n) &= g_n(y_n|x_n)\frac{\tilde{p}(y_{n+1}|x_n)\sum_{i=1}^N W_{n-1}^i f_n(x_n|X_{n-1}^i)/\tilde{p}(y_n|X_{n-1}^i)}{\sum_{i=1}^N W_{n-1}^i q_n(x_n|X_{n-1}^i, y_n)}.\end{aligned}$$

To obtain an algorithm targeting $\pi_n(\mathrm{d}x_n) := p(x_n|y_{1:n})\mathrm{d}x_n$, one additional importance sampling step is applied, using as proposal the approximation of $\hat{\eta}_n$ before resampling, $\Psi_{G_n^N}(\eta_n^N)$, and importance weights

$$\widetilde{w}_n(x_n) = \frac{\mathrm{d}\pi_n}{\mathrm{d}\hat{\eta}_n}(x_n) \propto \frac{1}{\tilde{p}(y_{n+1}|x_n)}. \tag{8}$$

2.1.3. SMC for approximate Bayesian computation

ABC-SMC is an instance of SMC samplers (Del Moral et al., 2006) studied in Sisson et al. (2007), Didelot et al. (2011) which approximates the posterior distribution of a parameter θ when the likelihood function $p(y_{obs}|\theta)$ is intractable but can be sampled from. Given a prior on the parameter θ, $p(\mathrm{d}\theta)$, and the intractable likelihood $p(\cdot|\theta)$, ABC-SMC considers an extended target defined over the space of parameter and data, $\hat{\eta}_n(\mathrm{d}(\theta_n, y_n)) \propto p(\mathrm{d}\theta_n)p(\mathrm{d}y_n|\theta_n)\pi_{\epsilon_n}(y_n|y_{obs})$, where $(\pi_{\epsilon_n})_{n\geq 0}$ is the density of a normalized kernel with a degree of concentration determined by ϵ_n which measures how close y_n is to the observed data y_{obs}. As $\epsilon_n \to 0$, $\hat{\eta}_n$ converges to the true posterior for θ (Sisson et al., 2007).

Didelot et al. (2011, Section 2.1) note that ABC-SMC is an instance of MSMC with $K_n((\theta_{n-1}, y_{n-1}), \mathrm{d}(\theta_n, y_n)) = p(\mathrm{d}\theta_n)p(\mathrm{d}y_n|\theta_n)$, $U_n((\theta_{n-1}, y_{n-1}), (\theta_n, y_n)) = U_n((\theta_n, y_n)) = \pi_{\epsilon_n}(y_n|y_{obs})$, $M_n((\theta_{n-1}, y_{n-1}), \mathrm{d}(\theta_n, y_n)) = q_n(\mathrm{d}\theta_n|\theta_{n-1})p(\mathrm{d}y_n|\theta_n)$, for some proposal q_n (we assume for brevity that the observations themselves are used, rather than some summary statistic but the use of such statistics does not present any difficulties).

3. Convergence results

We now state our main results, which show that marginal SMC methods have qualitatively the same convergence properties as standard SMC methods. We give an overview of the proofs for these results in Section 3.1, full details are given in Appendix D to Appendix G. For simplicity, we focus on the case of measurable bounded test functions $\varphi : E \to \mathbb{R}$ with $\|\varphi\|_\infty := \sup_{x\in E} |\varphi(x)| < \infty$, a set we denote by $\mathcal{B}_b(E)$. For any distribution η and any $\varphi \in \mathcal{B}_b(E)$ we denote $\eta(\varphi) := \int \varphi(x)\eta(\mathrm{d}x)$, similarly for all empirical distributions $\eta^N := N^{-1}\sum_{i=1}^N \delta_{X^i}$ we denote the corresponding average by $\eta^N(\varphi) := N^{-1}\sum_{i=1}^N \varphi(X^i)$. The probability space under which all probabilities and expectations are taken is that given by the law of all random variables generated within the SMC algorithm as given in Andrieu et al. (2010, Eq. (22)).

These results are presented under fairly strong assumptions, which are somewhat standard in this literature, in the interests of brevity. The techniques which allow these to be relaxed in the standard case would also apply here, but their use would substantially complicate the presentation. Similarly, in our arguments we only consider multinomial resampling (Gordon et al., 1993). Lower variance resampling schemes can be employed but considerably complicate the theoretical analysis (Gerber et al., 2019).

Assumption 1. The potentials G_n are positive everywhere, $G_n(x_n) > 0$ for every $x_n \in E$.

Assumption 2. For all $n \geq 0$, the functions U_n are bounded above, i.e. $U_n(x_{n-1}, x_n) \leq \|U_n\|_\infty < \infty$, and the Radon–Nikodym derivative $\mathrm{d}K_n(x_{n-1}, \cdot)/\mathrm{d}M_n(x_{n-1}, \cdot)$ is bounded above for all x_{n-1}, i.e. there exists some $\alpha > 0$ such that, for every $x_{n-1} \in E$, $\|\mathrm{d}K_n(x_{n-1}, \cdot)/\mathrm{d}M_n(x_{n-1}, \cdot)\|_\infty \leq \alpha < \infty$.

Assumption 3. For all $n \geq 0$, the functions U_n are bounded below, and the Radon–Nikodym derivative $\mathrm{d}K_n(x_{n-1}, \cdot)/\mathrm{d}M_n(x_{n-1}, \cdot)$ is bounded below for all x_{n-1}, i.e. there exist $\beta > 0$ such that $0 < \beta \leq U_n(x_{n-1}, x_n)$ and $0 < \beta \leq \mathrm{d}K_n(x_{n-1}, \cdot)/\mathrm{d}M_n(x_{n-1}, \cdot)$ uniformly in x_{n-1}.

Assumption 1 ensures that the system does not become extinct (i.e. the weights have never all simultaneously taken the value zero), and can be relaxed in various ways, including introducing stopping times Del Moral (2004, Theorem 7.4.3) or considering local boundedness (Whiteley, 2013). Assumption 2 ensures that, uniformly in $x \in E$, $G_n(x) \leq \|U_n\|_\infty\alpha$, $G_n^N(x) \leq \|U_n\|_\infty\alpha$, this and the boundedness assumption on the test functions φ can be avoided by considering appropriate integrability conditions, e.g. (Chopin, 2004; Douc and Moulines, 2008; Agapiou et al., 2017). Assumption 3 strengthens 1 and further guarantees $G_n(x) \geq \beta^{-2}$, $G_n^N(x) \geq \beta^{-2}$.

Assumptions 1 and 2 allow us to obtain the following weak law of large numbers (WLLN) whose proof is provided in Appendix E:

Proposition 1 (*Weak Law of Large Numbers*)**.** *Under Assumptions 1 and 2, for all $n \geq 0$ and for every $\varphi \in \mathcal{B}_b(E)$, we have $\Psi_{G_n^N}(\eta_n^N)(\varphi) \xrightarrow{p} \Psi_{G_n}(\eta_n)(\varphi)$ and $\hat{\eta}_n^N(\varphi) \xrightarrow{p} \hat{\eta}_n(\varphi)$.*

Assumption 3 is used to obtain stronger results like finite-N error bounds, whose proof is given in Appendix D:

Proposition 2 ($\mathbb{L}_p$-*Inequality*). *Under Assumptions 1–3, for every time $n \geq 0$, every $p \geq 1$ and $N \geq 1$ there exist finite constants $C_{p,n}, \bar{C}_{p,n}$ such that for every measurable bounded function $\varphi \in \mathcal{B}_b(E)$*

$$(a)\ \mathbb{E}\left[\left|\Psi_{G_n^N}(\eta_n^N)(\varphi) - \Psi_{G_n}(\eta_n)(\varphi)\right|^p\right]^{1/p} \leq \bar{C}_{p,n}\frac{\|\varphi\|_\infty}{\sqrt{N}},$$
$$(b)\ \mathbb{E}\left[\left|\hat{\eta}_n^N(\varphi) - \hat{\eta}_n(\varphi)\right|^p\right]^{1/p} \leq C_{p,n}\frac{\|\varphi\|_\infty}{\sqrt{N}}.$$

The strong law of large numbers requires stronger assumptions than the WLLN in Proposition 1 and can be obtained from the $\mathbb{L}_p$ inequality obtained in Proposition 2 using Markov's inequality within a Borel–Cantelli argument as shown in e.g. Boustati et al. (2020, Appendix D).

Proposition 3 (*Strong Law of Large Numbers*). *Under Assumptions 1–3, for all $n \geq 0$ and for every $\varphi \in \mathcal{B}_b(E)$, we have $\Psi_{G_n^N}(\eta_n^N)(\varphi) \xrightarrow{a.s.} \Psi_{G_n}(\eta_n)(\varphi)$ and $\hat{\eta}_n^N(\varphi) \xrightarrow{a.s.} \hat{\eta}_n(\varphi)$.*

Using standard techniques (e.g. Berti et al., 2006) given in detail for the context of interest in Schmon et al. (2021, Supplementary Material, Theorem 1), the result of Proposition 3 can be strengthened to the convergence of the measures in the weak topology:

Proposition 4. *Under Assumptions 1–3, for all $n \geq 0$, $\Psi_{G_n^N}(\eta_n^N)$ converges almost surely in the weak topology to $\Psi_{G_n}(\eta_n)$, $\Psi_{G_n^N}(\eta_n^N) \rightharpoonup \Psi_{G_n}(\eta_n)$, and $\hat{\eta}_n^N$ converges similarly to $\hat{\eta}_n$, $\hat{\eta}_n^N \rightharpoonup \hat{\eta}_n$.*

As it is the case for standard SMC algorithms, the reweighting step introduces a bias into estimates of normalized quantities, however, this decays at rate N^{-1} as established in Appendix F:

Proposition 5 (*Bias Estimate*). *Under Assumptions 1–3, for all $n \geq 0$ and any $\varphi \in \mathcal{B}_b(E)$ we have*

$$(a)\ \left|\mathbb{E}\left[\Psi_{G_n^N}(\eta_n^N)(\varphi)\right] - \Psi_{G_n}(\eta_n)(\varphi)\right| \leq \bar{C}_n\frac{\|\varphi\|_\infty}{N},$$
$$(b)\ \left|\mathbb{E}\left[\hat{\eta}_n^N(\varphi)\right] - \hat{\eta}_n(\varphi)\right| \leq C_n\frac{\|\varphi\|_\infty}{N},$$

for some finite $\bar{C}_n, C_n$.

The following result, proved in Appendix G, quantifies the asymptotic variance of the estimates provided by Algorithm 1 using multinomial resampling. We focus on this resampling scheme because of its simplicity and because, as shown in Gerber et al. (2019, Theorem 7), it provides an upper bound on the asymptotic variance obtained with more sophisticated resampling schemes.

Proposition 6 (*Central Limit Theorem*). *Under Assumptions 1–3, for every $n \geq 1$ and $\varphi \in \mathcal{B}_b(E)$:*

$$(a)\ \sqrt{N}\left[\Psi_{G_n^N}(\eta_n^N)(\varphi) - \Psi_{G_n}(\eta_n)(\varphi)\right] \xrightarrow{d} \mathcal{N}\left(0, \bar{V}_n(\varphi)\right), \qquad (b)\ \sqrt{N}\left[\hat{\eta}_n^N(\varphi) - \hat{\eta}_n(\varphi)\right] \xrightarrow{d} \mathcal{N}\left(0, V_n(\varphi)\right),$$

where $\bar{V}_n(\varphi), V_n(\varphi)$ are operators acting on φ given by the following recursion: $\bar{V}_0(\varphi) = \mathrm{var}_{M_0}\left(\frac{G_0(\varphi - \hat{\eta}_0(\varphi))}{\eta_0(G_0)}\right)$,

$$\widehat{V}_n(\varphi) = \mathrm{var}_{\eta_n}(G_n\varphi) + \bar{V}_{n-1}(K_n(U_n\varphi)); \bar{V}_n(\varphi) = \frac{1}{\eta_n(G_n)^2}\widehat{V}_n\left(\varphi - \Psi_{G_n}(\eta_n)(\varphi)\right); V_n(\varphi) = \mathrm{var}_{\hat{\eta}_n}(\varphi) + \bar{V}_n(\varphi),$$

where we denote $\mathrm{var}_\eta(\varphi) = \eta(\varphi^2) - \eta(\varphi)^2$ for any distribution η.

3.1. Method of proof

We emphasize that the proof techniques used are small extensions of those used in the standard SMC setting; the primary interest of these results is that they demonstrate that the marginal version of the algorithm inherits many of the good properties of the standard algorithm and allowing comparison between the two algorithms via their asymptotic variances. The details of the proofs are postponed to Appendix D to Appendix H where we give self-contained arguments incorporating the novel elements discussed in this section with well-known techniques used to obtain similar results for standard SMC.

The main difference between standard SMC methods and marginal SMC methods is the presence of an additional approximations in the weights (5); if we could use the idealized algorithm in which G_n in (4) can be computed exactly, then we could apply the theoretical results for standard SMC (e.g., Del Moral, 2013). Hence, to obtain convergence results for marginal SMC we need to control the behaviour of the non-standard weights G_n^N. We point out that since the weights G_n^N are biased approximations of G_n, we cannot use the arguments based on extensions of the state space (as in particle filters using unbiased estimates of the potentials; Fearnhead et al., 2008) to provide theoretical guarantees for MSMC.

To control the effect of G_n^N, we identify the conditional expectation of $\eta_n^N(G_n^N\varphi)$ in Proposition 7 extending the result of Branchini and Elvira (2021, Lemma 1). Proposition 7, whose simple proof is provided in Appendix C, combined with a number of results presented in the appendix which employ this conditional expectation within expansions of various aspects of the sampling error allow us to obtain results for MSMC algorithms in some generality.

Proposition 7. *Let $\mathcal{F}_{n-1}^N$ denote the σ-field generated by the weighted samples up to (and including) time $n-1$. We have $\mathbb{E}\left[\eta_n^N(G_n^N\varphi) \mid \mathcal{F}_{n-1}^N\right] = \Psi_{G_{n-1}^N}(\eta_{n-1}^N)(K_n(\varphi U_n))$, for all $n \geq 1$ and all $\varphi \in \mathcal{B}_b(E)$.*

For the proof of the $\mathbb{L}_p$-inequality we combine the results of Crisan and Doucet (2002), Míguez et al. (2013) which control the approximation introduced by the mutation step, the reweighting step and the resampling step with Lemma 2 in Appendix D which controls the error induced by using the approximated weights (5). The proof of Lemma 2 is based on a comparison between $\eta_n^N(G_n^N\varphi)$ and its conditional expectation in Proposition 7 combined with the inductive hypothesis. Similarly, the proof of the bias estimates in Proposition 5 combines the inductive approach of Olsson and Rydén (2004) with Proposition 7, showing that the approximate weights do not worsen the rate of decay of bias with N.

For the proof of the weak law of large numbers we again adopt an inductive strategy similar to that of Douc and Moulines (2008, Theorem 1) and Cappé et al. (2005, Theorem 9.4.5), we combine standard arguments establishing a WLLN for the mutation, the reweighting and the resampling step with Lemma 5 in Appendix E which establishes a WLLN for $\eta_n^N(G_n^N\varphi)$.

The central limit theorem follows using the inductive approach of Chopin (2004), with an additional result (Lemma 9) which shows that $\eta_n^N(G_n^N\varphi)$ satisfies a central limit theorem.

3.2. Variance comparison

Proposition 6 gives a recursive formula for the asymptotic variance similar to that in Chopin (2004). Comparing Proposition 6 and Chopin (2004, Section 2.3), we find that the main difference in the variance expression appears in $\widehat{V}_n(\varphi)$: in the case of standard SMC we have $\widehat{V}_n^{\text{SMC}}(\varphi) = \text{var}_{\eta_n}(G_n\varphi) + \bar{V}_{n-1}(M_n(G_n\varphi))$. This is not surprising since the main difference between SMC and MSMC is in the importance weights (4), whose expression is taken into account in $\widehat{V}_n(\varphi)$.

Using an inductive argument whose details are given in Appendix H.1 (see also Chopin, 2004, Eq. 9 and Johansen and Doucet, 2007, Appendix A), we obtain a closed form for $\bar{V}_n(\varphi)$. To this end, let us define the following operator akin that of Chopin (2004, Eq. 10) and Del Moral (2004, Section 2.7.2):

$$\forall q \in \mathbb{N}: \qquad \Gamma_q(\varphi)(x_{q-1}) := \int K_q(x_{q-1}, \mathrm{d}x_q) U_q(x_{q-1}, x_q)\varphi(x_q),$$

and, for all $p < q$, the two-parameter semigroup $\Gamma_{p:q}(\varphi) = \Gamma_{p+1} \circ \cdots \circ \Gamma_q(\varphi)$, where $\circ$ denotes the composition operator, such that $\hat{\eta}_q \propto \hat{\eta}_p \Gamma_{p:q}$, with the convention $\Gamma_{q:q} = \mathrm{Id}$.

The following representation is established in Appendix H.1:

Proposition 8. *For all $n \geq 0$ and all $\varphi \in \mathcal{B}_b(E)$, the variance $\bar{V}_n(\varphi)$ in Proposition 6 can be equivalently written as*

$$\bar{V}_n(\varphi) = \sum_{k=0}^{n} \mathbb{E}_{\eta_k}\left[\left(G_k\left[\Gamma_{k:n}(\varphi) - \hat{\eta}_n(\varphi)\Gamma_{k:n}(1)\right]\right)^2\right] \prod_{j=k}^{n} \frac{1}{\eta_j(G_j)^2} \tag{9}$$

The variances of MPF and MAPF in Section 2.1 can be obtained from (9) by simple algebraic manipulations which we postpone to Appendix H. See Appendix H.2 for the calculations underlying the following corollary:

Corollary 1 (*Variance of MPF*)**.** *For all $n \geq 0$ and all $\varphi \in \mathcal{B}_b(E)$ we have*

$$\begin{aligned} V_n^{MPF}(\varphi) = &\int \frac{p(x_0|y_{1:n})^2}{q_0(x_0)} \left(\int p(x_n|y_{1:n}, x_0)\left[\varphi(x_n) - \bar{\varphi}_n\right] \mathrm{d}x_n\right)^2 \mathrm{d}x_0 \\ &+ \int \frac{p(x_k|y_{1:n})^2}{\int q_k(x_k|x_{k-1}, y_k) p(x_{k-1}|y_{1:k-1}) \mathrm{d}x_{k-1}} \left(\int p(x_n|y_{k+1:n}, x_k)\left[\varphi(x_n) - \bar{\varphi}_n\right] \mathrm{d}x_n\right)^2 \mathrm{d}x_k \\ &+ \int \frac{p(x_n|y_{1:n})^2}{\int q_n(x_n|x_{n-1}, y_n) p(x_{n-1}|y_{1:n-1}) \mathrm{d}x_{n-1}} (\varphi(x_n) - \bar{\varphi}_n)^2 \mathrm{d}x_n, \end{aligned}$$

with $\bar{\varphi}_n := \int \varphi(x_n) p(x_n|y_{1:n}) \mathrm{d}x_n$. In addition, $V_n^{MPF}(\varphi) \leq V_n^{PF}(\varphi)$, where $V_n^{PF}(\varphi)$ denotes the asymptotic variance of a particle filter with the same proposals (Johansen and Doucet, 2008, Section 2.4).

The variance of the MAPF can be obtained from $\bar{V}_n(\varphi)$ in Proposition 8 via an additional importance sampling step with weights $\tilde{w}_n$ in (8), as shown in Appendix H.3.

Corollary 2 (*Variance of MAPF*). *For all $n \geq 0$ and all $\varphi \in \mathcal{B}_b(E)$ we have*

$$V_n^{MAPF}(\varphi) = \int \frac{p(x_0|y_{1:n})^2}{q_0(x_0)} \left(\int p(x_n|x_0, y_{1:n}) [\varphi(x_n) - \bar{\varphi}_n] \, \mathrm{d}x_n \right)^2 \mathrm{d}x_0 + \int \frac{p(x_k|y_{1:n})^2}{\int q_k(x_k|x'_{k-1}, y_k) \hat{\eta}_{k-1}(\mathrm{d}x'_{k-1})} \left(\int p(x_n|x_k, y_{1:n}) [\varphi(x_n) - \bar{\varphi}_n] \, \mathrm{d}x_n \right)^2 \mathrm{d}x_k + \int \frac{p(x_n|y_{1:n})^2 [\varphi(x_n) - \bar{\varphi}_n]^2}{\int q_n(x_n|x'_{n-1}, y_n) \hat{\eta}_{n-1}(\mathrm{d}x'_{n-1})} \mathrm{d}x_n.$$

with $\bar{\varphi}_n := \int \varphi(x_n) p(x_n|y_{1:n}) \mathrm{d}x_n$. In addition, $V_n^{MAPF}(\varphi) \leq V_n^{APF}(\varphi)$, where $V_n^{APF}(\varphi)$ denotes the asymptotic variance of an APF with the same proposals (Johansen and Doucet, 2008, Section 2.4).

Using the results in Walker (2014), we can show that equality occurs in the MPF case only when $q_k \equiv f_k$ and the variance expression in Corollary 1 coincides with that of the bootstrap particle filter (see Johansen and Doucet (2008, Section 2.4) and Appendix H.2.1). Whereas, in the MAPF case equality occurs only when $q_k \propto C_k f_k$ where C_k is a positive function only depending on x_k and $\tilde{p}(y_k|x_{k-1}) = \int C_k(x_k) f_k(x_k|x_{k-1}) \mathrm{d}x_k$—a special case is $C_k(x_k) = g_k(y_k|x_k)$ for which the MAPF collapses onto the fully adapted APF (FA-APF) and hence has the same asymptotic variance (see Johansen and Doucet (2008, Corollary); Appendix H.3.1). In all other cases, the variance reduction can be quantified using, e.g., Walker (2014, Theorem 3.1).

4. Discussion

In this work we established that a class of marginal sequential Monte Carlo (MSMC) algorithms, which encompasses marginal particle filters and other popular algorithms in the literature, satisfies many of the key properties that standard sequential Monte Carlo methods have. The results in Section 3 guarantee that the estimates provided by MSMC are consistent, asymptotically normal, with a bias decaying at rate N^{-1} and $\mathbb{L}_p$ error decaying at rate $N^{-1/2}$.

Comparing the asymptotic variances in Proposition 6 with those for standard SMC obtained in, e.g., Chopin (2004), Del Moral (2004), we find that marginal particle filters have no larger asymptotic variance than the corresponding non-marginal particle filter (Corollaries 1 and 2) and are therefore likely to provide better estimates then the corresponding non-marginal algorithms, as already observed empirically (e.g. Klaas et al., 2005; Xu and Jasra, 2019). Corollary 2 complements Klaas et al. (2005, Proposition 1) showing that the importance sampling weights of MAPF have lower (conditional) variance.

Unbiasedness of the normalizing constant estimates has been shown in Branchini and Elvira (2021, Theorem 2). Combining Proposition 7 with the approach of Branchini and Elvira (2021, Lemma 2) one can further show that the unnormalized version of the measure $\hat{\eta}_n$ in (2) provides unbiased estimates.

Quantifying the variance reduction obtained by employing marginal particle filters instead of standard particle filter is a more challenging question, the answer to which is likely to be dependent on the specific state space model and proposals q_k. The variance reduction obtained by using MPF instead of PF should then be weighted against the additional computation cost required by MPF w.r.t. the $O(N)$ cost of PF. In their most naive implementation MPF require an $O(N^2)$ cost, which can however be reduced to $O(N \log N)$ using techniques from N-body learning (e.g. Gray and Moore, 2000; Lang et al., 2005) as shown in Klaas et al. (2005) or efficient implementations using GPUs (Charlier et al., 2021) as shown in Clarté et al. (2022, Section 4) for sums of the form of those in (5).

Marginalization techniques similar to those studied here are also commonplace in the sequential quasi Monte Carlo literature (Fearnhead, 2005; Gerber and Chopin, 2015), however, in those case the marginalization is motivated by other considerations not covered by our results. Finally, we expect that similar ideas could be applied to more sophisticated SMC algorithms employing marginalization techniques (e.g. Xu and Jasra, 2019; Crucinio and Johansen, 2023).

Data availability

No data was used for the research described in the article.

Acknowledgments

Support from the EPSRC, United Kingdom (grants # EP/R034710/1 and # EP/T004134/1) and the Lloyd's Register Foundation, United Kingdom Programme on Data-Centric Engineering at the Alan Turing Institute is acknowledged. The authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission. Data sharing is not applicable to this article.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spl.2023.109914.

References

Agapiou, S., Papaspiliopoulos, O., Sanz-Alonso, D., Stuart, A.M., 2017. Importance sampling: Intrinsic dimension and computational cost. Statist. Sci. 405–431.

Andrieu, C., Doucet, A., Holenstein, R., 2010. Particle Markov chain Monte Carlo methods. J. R. Stat. Soc. Ser. B Stat. Methodol. 72 (3), 269–342.

Berti, P., Pratelli, L., Rigo, P., 2006. Almost sure weak convergence of random probability measures. Stoch. Stoch. Rep. 78 (2), 91–97.

Boustati, A., Akyildiz, Ö.D., Damoulas, T., Johansen, A.M., 2020. Generalized Bayesian filtering via sequential Monte Carlo. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M.F., Lin, H. (Eds.), Advances in Neural Information Processing Systems, Vol. 33. Curran Associates, Inc., pp. 418–429.

Branchini, N., Elvira, V., 2021. Optimized auxiliary particle filters: adapting mixture proposals via convex optimization. In: 37th Conference on Uncertainty in Artificial Intelligence. pp. 1289–1299, Proceedings of Machine Learning Research.

Cappé, O., Moulines, E., Rydén, T., 2005. Inference in Hidden Markov Models. Springer.

Carpenter, J., Clifford, P., Fearnhead, P., 1999. An improved particle filter for non-linear problems. IEE Proc. Radar Sonar Navig. 146 (1), 2–7.

Charlier, B., Feydy, J., Glaunes, J.A., Collin, F., Durif, G., 2021. Kernel operations on the GPU, with autodiff, without memory overflows. J. Mach. Learn. Res. 22 (74), 1–6.

Chopin, N., 2004. Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. Ann. Statist. 32 (6), 2385–2411.

Chopin, N., Papaspiliopoulos, O., 2020. An Introduction to Sequential Monte Carlo. Springer, Cham.

Clarté, G., Diez, A., Feydy, J., 2022. Collective proposal distributions for nonlinear MCMC samplers: Mean-field theory and fast implementation. Electron. J. Stat. 22 (2), 6395–6460.

Crisan, D., Doucet, A., 2002. A survey of convergence results on particle filtering methods for practitioners. IEEE Trans. Signal Process. 50 (3), 736–746.

Crucinio, F.R., Johansen, A.M., 2023. A divide and conquer sequential Monte Carlo approach to high dimensional filtering. Statist. Sinica (in press).

Del Moral, P., 1996. Nonlinear filtering: Interacting particle resolution. Markov Process. Related Fields 2 (4), 555–580.

Del Moral, P., 2004. Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications. In: Probability and Its Applications, Springer Verlag, New York.

Del Moral, P., 2013. Mean Field Simulation for Monte Carlo Integration. Chapman and Hall/CRC, New York.

Del Moral, P., Doucet, A., Jasra, A., 2006. Sequential Monte Carlo samplers. J. R. Stat. Soc. Ser. B 63 (3), 411–436.

Didelot, X., Everitt, R.G., Johansen, A.M., Lawson, D.J., 2011. Likelihood-free estimation of model evidence. Bayesian Anal. 6 (1), 49–76.

Douc, R., Moulines, R., 2008. Limit theorems for weighted samples with applications to sequential Monte Carlo methods. Ann. Statist. 36 (5), 2344–2376.

Elvira, V., Martino, L., Bugallo, M.F., Djuric, P.M., 2019. Elucidating the auxiliary particle filter via multiple importance sampling [lecture notes]. IEEE Signal Process. Mag. 36 (6), 145–152.

Everitt, R.G., Prangle, D., Maybank, P., Bell, M., 2017. Marginal sequential Monte Carlo for doubly intractable models. arXiv preprint arXiv:1710.04382.

Fearnhead, P., 1998. Sequential Monte Carlo Methods in Filter Theory (Ph.D. thesis). University of Oxford.

Fearnhead, P., 2005. Using random quasi-Monte-Carlo within particle filters, with application to financial time series. J. Comput. Graph. Statist. 14 (4), 751–769.

Fearnhead, P., Papaspiliopoulos, O., Roberts, G.O., 2008. Particle filters for partially observed diffusions. J. R. Stat. Soc. Ser. B Stat. Methodol. 70 (4), 755–777.

Gerber, M., Chopin, N., 2015. Sequential quasi-Monte Carlo. J. R. Stat. Soc. Ser. B 3 (77), 509–579.

Gerber, M., Chopin, N., Whiteley, N., 2019. Negative association, ordering and convergence of resampling methods. Ann. Statist. 47 (4), 2236–2260.

Gordon, N.J., Salmond, S.J., Smith, A.F.M., 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proc. F 140 (2), 107–113.

Gray, A., Moore, A., 2000. 'N-body' problems in statistical learning. In: Advances in Neural Information Irocessing Systems. pp. 521–527.

Johansen, A.M., Doucet, A., 2007. Auxiliary Variable Sequential Monte Carlo Methods. University of Bristol, Department of Mathematics – Statistics Group, University Walk, Bristol, BS8 1TW, UK.

Johansen, A.M., Doucet, A., 2008. A note on auxiliary particle filters. Statist. Probab. Lett. 78 (12), 1498–1504.

Klaas, M., de Freitas, N., Doucet, A., 2005. Towards practical N^2 Monte Carlo: The marginal particle filter. In: Proceedings of Uncertainty in Artificial Intelligence. pp. 308–315.

Lai, J., Domke, J., Sheldon, D., 2022. Variational marginal particle filters. In: International Conference on Artificial Intelligence and Statistics. PMLR, pp. 875–895.

Lang, D., Klaas, M., de Freitas, N., 2005. Empirical Testing of Fast Kernel Density Estimation Algorithms. Technical Report TR2005-03, Department of Computer Science, University of British Columbia.

Li, W., Chen, R., Tan, Z., 2016. Efficient sequential Monte Carlo with multiple proposals and control variates. J. Amer. Statist. Assoc. 111 (513), 298–313.

Lin, M.T., Zhang, J.L., Cheng, Q., Chen, R., 2005. Independent particle filters. J. Amer. Statist. Assoc. 100 (472), 1412–1421.

Míguez, J., Crisan, D., Djurić, P.M., 2013. On the convergence of two sequential Monte Carlo methods for maximum a posteriori sequence estimation and stochastic global optimization. Stat. Comput. 23 (1), 91–107.

Olsson, J., Rydén, T., 2004. The Bootstrap Particle Filtering Bias. Lund University, Technical Report 929081, Manne Siegbahn Institute.

Pitt, M.K., Shephard, N., 1999. Filtering via simulation: Auxiliary particle filters. J. Amer. Statist. Assoc. 94 (446), 590–599.

Poyiadjis, G., Doucet, A., Singh, S.S., 2011. Particle approximations of the score and observed information matrix in state space models with application to parameter estimation. Biometrika 98 (1), 65–80.

Schmon, S.M., Deligiannidis, G., Doucet, A., Pitt, M.K., 2021. Large sample asymptotics of the pseudo-marginal method. Biometrika 108 (1), 37–51.

Sisson, S.A., Fan, Y., Tanaka, M.M., 2007. Sequential Monte Carlo without likelihoods. Proc. Natl. Acad. Sci. USA 104 (4), 1760–1765.

Walker, S.G., 2014. On a lower bound for the Jensen inequality. SIAM J. Math. Anal. 46 (5), 3151–3157.

Whiteley, N., 2013. Stability properties of some particle filters. Ann. Appl. Probab. 23 (6), 2500–2537.

Xu, Y., Jasra, A., 2019. Particle filters for inference of high-dimensional multivariate stochastic volatility models with cross-leverage effects. Found. Data Sci. 1 (1), 61–85.