Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Efficient Sequential Monte Carlo With Multiple Proposals and Control Variates

Wentao Li, Rong Chen & Zhiqiang Tan

To cite this article: Wentao Li, Rong Chen & Zhiqiang Tan (2016) Efficient Sequential Monte Carlo With Multiple Proposals and Control Variates, Journal of the American Statistical Association, 111:513, 298-313, DOI: 10.1080/01621459.2015.1006364

To link to this article: https://doi.org/10.1080/01621459.2015.1006364

Accepted author version posted online: 06 Feb 2015.
Published online: 05 May 2016.

Efficient Sequential Monte Carlo With Multiple Proposals and Control Variates

Wentao LI, Rong CHEN, and Zhiqiang TAN

Sequential Monte Carlo is a useful simulation-based method for online filtering of state-space models. For certain complex state-space models, a single proposal distribution is usually not satisfactory and using multiple proposal distributions is a general approach to address various aspects of the filtering problem. This article proposes an efficient method of using multiple proposals in combination with control variates. The likelihood approach of Tan (2004) is used in both resampling and estimation. The new algorithm is shown to be asymptotically more efficient than the direct use of multiple proposals and control variates. The guidance for selecting multiple proposals and control variates is also given. Numerical examples are used to demonstrate that the new algorithm can significantly improve over the bootstrap filter and auxiliary particle filter.

KEY WORDS: Auxiliary particle filter; Defensive proposal distribution; Filtering; Importance sampling; Regression.

1. INTRODUCTION

Since the introduction of particle filtering in Gordon, Salmond, and Smith (1993), simulation-based methods for online filtering of dynamic systems have been widely used in various fields, including target tracking (Chen and Liu 2000), signal processing (Wang, Chen, and Guo 2002), estimation of economics models (Shephard 2005), and counting contingency tables (Chen et al. 2005). In a dynamic system, there usually exist some unobservable state x_n and observable output y_n at time n. Online filtering aims to make inference on a *target function* $h(x_n)$ conditional on the entire observation history $y_1, \ldots, y_n$. At every n, the simulation-based methods generate a set of properly weighted Monte Carlo sample, called particles, to approximate the target distribution $\pi(x_n|y_1, \ldots, y_n)$. Various efforts have been devoted to improve the basic particle filtering method, including improving the sampling mechanism (Doucet, Briers, and Sénécal 2006), increasing the diversity of samples (Gilks and Berzuini 2001), and adaptively choosing the resampling schedule (Liu and Chen 1995). Sequential Monte Carlo (SMC) is a unified framework including most of these techniques, proposed by Doucet, Godsill, and Andrieu (2000) and Liu and Chen (1998). Reviews of SMC can be found in Doucet and Johansen (2009), Cappé, Godsill, and Moulines (2007) and Chen (2005). The SMC framework is marked by the central role of sequential importance sampling with resampling, where sequential importance sampling generates new weighted particles at each time point from the proposal distributions based on existed particles, and resampling mitigates the particle degeneracy, that is, the sample weights of a few particle dominate the others. A Markovian state-space model has the form of

$$\begin{aligned} x_n &= s_1(x_{n-1}, \varepsilon_n) \\ y_n &= s_2(x_n, e_n), \end{aligned} \tag{1}$$

where ε_n and e_n are random noises. The model is characterized by the state equation which defines a Markovian process with transition density $p(x_n|x_{n-1})$, and the observation equation where y_n only depends on the current state through density $p(y_n|x_n)$. This article considers the online filtering problem for (1).

The design of proposal distributions is essential to the SMC methods. The bootstrap filter (BF) of Gordon, Salmond, and Smith (1993) uses $p(x_n|x_{n-1})$, with its limitation that the sampling mechanism does not use the information in the observations y_n. Nevertheless, it is popular due to its simplicity and computational efficiency. The independent particle filter of Lin et al. (2005) uses $p(x_n \mid y_n)$, using only the observation, with some additional benefits. Using $p(x_n|x_{n-1}, y_n)$ as the proposal distribution is known to be optimal in the sense that the variance of particle weights conditional on x_{n-1} is 0 (Liu and Chen 1998; Pitt and Shephard 1999; Doucet, Godsill, and Andrieu 2000). Intuitively, it includes the full information of both state and observation equations. In practice, approximations of $p(x_n|x_{n-1}, y_n)$ are usually used. See Doucet, Godsill, and Andrieu (2000), Guo, Wang, and Chen (2005), Saha et al. (2009) and Pitt and Shephard (1999) for examples.

There are several limitations with the proposal distributions mentioned previously. First, the approximation to $p(x_n \mid x_{n-1}, y_n)$ may not give bounded variance for particle weights since the local approximation does not necessarily provide control on the tails of the proposal density. Second, the above approaches usually construct unimodal densities which are inefficient for multimodal target densities. Finally, although $p(x_n \mid x_{n-1}, y_n)$ includes both the information of the state and observation equations, it does not consider the target function $h(x_n)$ of the inference problem at hand. It may not be efficient in certain cases including estimating the probability of rare event.

Although there have been some discussions of the last issue under specific cases such as estimating the tail probability (Chan and Lai 2011; Cérou et al. 2012), there is no general guideline to deal with general target functions. For the first two issues mentioned previously, a general remedy is to consider a mixture

Wentao Li is Senior Research Associate, Department of Mathematics and Statistics, Fylde College, Lancaster University, Bailrigg, Lancaster LA1 4YF, United Kingdom (E-mail: *w.li@lancaster.ac.uk*). Rong Chen is Professor (E-mail: *rongchen@stat.rutgers.edu*), and Zhiqiang Tan is Associate Professor (E-mail: *ztan@stat.rutgers.edu*), Department of Statistics, Rutgers University, Piscataway, NJ 8854. Chen's research is sponsored in part by NSF grants DMS 0905763 and DMS0915139. Tan's research is sponsored in part by NSF grant DMS 0749418. The authors thank the two editors, an associate editor and two anonymous referees for their helpful comments.

Journal of the American Statistical Association
March 2016, Vol. 111, No. 513, Theory and Methods
DOI: 10.1080/01621459.2015.1006364

of multiple proposal distributions which may include proposals designed for controlling tails or concentrating on multiple modes. The usage of multiple proposals in importance sampling, which is essentially a single step of SMC, has been well discussed in the literature, including Hesterberg (1995), Owen and Zhou (2000), Oh and Berger (1993), Veach and Guibas (1995), Tan (2004), and Li, Tan, and Chen (2013). Combining multiple proposals with appropriate control variates can significantly increase efficiency and decrease contamination caused by mixing poor proposal distributions with good ones. Particularly, the likelihood estimator in Tan (2004) is shown to be optimal in a general class of estimators with multiple proposals. In SMC framework, the usage of multiple proposals and control variates only receives limited discussions for specific cases. Elinas, Sim, and Little (2006) and Fox et al. (2001) applied mixture proposal distribution in a Monte Carlo approach to a robot localization problem by combining information from camera/laser observations and a motion model. Singh et al. (2004) and Singh et al. (2007) considered control variates with target function $h(x_n)$ in a particle filter for target tracking sensor management.

In this article, we propose a novel algorithm under the SMC framework to tackle the three issues mentioned previously. Specifically, we use particles generated from a mixture proposal distribution with importance weights constructed by Tan's (2004) likelihood approach for both resampling and estimation in SMC. Under this new approach, a bounded variance is guaranteed by including proposals with heavier tails than the target distribution, and multimodality can be handled by including proposals to target each mode separately. The third issue can be dealt with by considering the target function in the construction of control variates and proposal distributions. A guideline for selecting component proposals and control variates are given for general target functions. The theoretical framework of the algorithm is constructed, and the asymptotic results show that the new algorithm is more efficient than a naive implementation of multiple proposals and control variates in SMC, and can be expected to increase efficiency significantly over standard SMC methods. Its effectiveness is illustrated through numerical studies on an AR(1) model observed with noise, serving as a benchmark model since all sequential distributions are analytically available, and on a stochastic volatility model with AR(1) dynamics which is widely used in economics and finance.

This article is organized as follows. Section 2 provides some overview of related methods of importance sampling with multiple proposals and some related SMC methods. In Section 3, the new algorithm is presented with some theoretical results. Guidance to several implementation issues are also provided. In Section 4, the new algorithm is demonstrated with several numerical examples.

2. OVERVIEW OF RELATED BACKGROUND

2.1 Basic SMC Method

For the state-space model (1), at time n, denote the random vector $(x_1, \ldots, x_n)$ by $x_{1:n}$ and observations $(y_1, \ldots, y_n)$ by $y_{1:n}$. To estimate the target function $h(x_{1:n})$, the optimal solution in the sense of Bayesian inference, where prior of $x_{1:n}$ is given by the state equation, is the posterior mean

$$\mu_n \equiv E\,[h(x_{1:n})|y_{1:n}] = \int h(x_{1:n})\pi_n^*(x_{0:n})dx_{0:n},$$

where x_0 is the initial state of system and $\pi_k^*(x_{0:n})$ is the posterior density of $x_{0:n}$ given $y_{1:k}$. By the posterior formula,

$$\pi_n^*(x_{0:n}) \propto \pi_n(x_{0:n}) \equiv \pi_0(x_0)\prod_{i=1}^{n} p(x_i|x_{i-1})p(y_i|x_i),$$

with normalizing constant c_n. Since n is usually large, standard importance sampling suffers from highly skew sample weights. For the introduction of importance sampling, see Robert and Casella (2004) and Liu (2008) and many other references. By combining the sequential importance sampling (SIS) and resampling algorithms, the high-dimensional problem can be mitigated:

Basic SMC Method:
At time n, assume weighted particles $\{(\widetilde{x}_{0:n-1}^{(j)}; \widetilde{w}_{n-1}^{(j)})\}_{j=1}^N$ are available. For $j = 1, \ldots, N$,

1. Propagation (Mutation): Generate $x_n^{(j)}$ from the proposal distribution $q(x_n|\widetilde{x}_{0:n-1}^{(j)})$ and let $x_{0:n}^{(j)} = (\widetilde{x}_{0:n-1}^{(j)}, x_n^{(j)})$.
2. Weighting (Correction): Assign $x_{0:n}^{(j)}$ with weight

$$w_n^{(j)} = \widetilde{w}_{n-1}^{(j)} \frac{p(x_n^{(j)}|\widetilde{x}_{n-1}^{(j)})p(y_n|x_n^{(j)})}{q(x_n^{(j)}|\widetilde{x}_{0:n-1}^{(j)})}. \tag{2}$$

3. Resampling (Selection): If the condition for resampling is satisfied, resample $\{x_{0:n}^{(j)}\}_{j=1}^N$ according to $\{w_n^{(j)}\}_{j=1}^N$ to obtain new weighted particles $\{(\widetilde{x}_{0:n}^{(j)}; \widetilde{\omega}_n^{(j)})\}_{j=1}^N$ where $\widetilde{\omega}_n^{(j)} = 1/N$; If the condition for resampling is not satisfied, let $(\widetilde{x}_{0:n}^{(j)}; \widetilde{\omega}_n^{(j)}) = (x_{0:n}^{(j)}, \omega_n^{(j)})$.

After the weighting step, μ_n can be estimated by

$$\widehat{\mu}_{n,\text{basic}} = \frac{\sum_{j=1}^N h(x_{1:n}^{(j)})\omega_n^{(j)}}{\sum_{j=1}^N \omega_n^{(j)}}.$$

The SIS algorithm, containing the propagation and weighting steps, divides the sampling into sequential steps, which can be seen from the following equation

$$\frac{\pi_n(x_{0:n})}{q(x_{0:n})} = \frac{\pi_{n-1}(x_{0:n-1})}{q(x_{0:n-1})} \frac{\pi_n(x_{0:n})}{\pi_{n-1}(x_{0:n-1})q(x_n|x_{0:n-1})},$$

where $q(x_{0:n}) = \pi_0(x_0)\prod_{t=1}^n q(x_t|x_{0:t-1})$ and $\pi_n(x_{0:n})/\pi_{n-1}(x_{0:n-1}) = p(x_n|x_{n-1})p(y_n|x_n)$. This sequential implementation is appropriate for online analysis. The resampling step performs resampling to drop the samples with small weights and duplicate the samples with large weights, hence mitigating the particle degeneracy. The schedule of performing resampling can be either fixed or adaptive according to certain quality indicator of the particles (Liu 2008). After resampling, $\{\widetilde{x}_{0:n}^{(j)}\}_{j=1}^N$ are approximately distributed according to $\pi_n^*(x_{0:n})$ and equal weights are assigned to each sample. The algorithm may be initialized by generating iid samples $\{x_0^{(j)}\}_{j=1}^n$ from $\pi_0(x_0)$ and setting $w_0^{(j)} = 1/N$ for all j.

For the choice of the proposal density $q(x_n|x_{0:n-1})$, $p(x_n|x_{n-1})$ is often used but $p(x_n|y_n, x_{n-1}) \propto p(x_n|x_{n-1})p(y_n|x_n)$ is considered optimal. When using the prior $p(x_n|x_{n-1})$, observations are only used in the calculation of the importance weights

$\{w_n^{(j)}\}_{j=1}^N$, but not in the propagation step. When observations have outliers or are very informative (with small variance), the number of particles after propagation in the high posterior likelihood area will be small, leading to particle degeneracy and therefore poor algorithm performance (Cappé, Godsill, and Moulines 2007). When $p(x_n|y_n, x_{n-1})$ is analytically unavailable, often due to nonlinearity in the system, its approximation by linearizing $\log(p(x_n|x_{n-1}))$ or $\log(p(y_n|x_n))$ can be used instead.

As a preliminary for the theoretical analysis of the new algorithm, we restate the central limit theorem (CLT) of $\widehat{\mu}_{n,\text{basic}}$ in Douc and Moulines (2008) here. Suppose multinomial resampling is performed at every step. Let Θ_n be the domain of $x_{0:n}$ and $L^1(\pi_n) = \{h : \int |h(x_{0:n})|\pi_n(x_{0:n})dx_{0:n} < \infty\}$. A set C of real function is said to be proper if: (1) C is a linear space; (2) if $g \in C$ and f is measurable, $|f| < |g|$ implies that $f \in C$; (3) For all c, the constant function $f \equiv c$ belongs to C. Assume that:

(C1) For some proper set A_0, the following convergence hold:

$$\frac{1}{n}\sum_{j=1}^{N} h(x_0^{(j)}) \xrightarrow{P} \int h(x_0)\pi_0(x_0)dx_0, \ \forall h \in L^1(\pi_0),$$

$$\sqrt{N}\left[\frac{1}{N}\sum_{j=1}^{N} h(x_0^{(j)}) - \int h(x_0)\pi_0(x_0)dx_0\right] \xrightarrow{\mathcal{L}} N(0, \sigma_0(h)), \quad \forall h \in A_0.$$

With a proper set A_0, define

$$A_n = \left\{h : \Theta_n \to \mathbb{R} | E_{\pi_n^*}[h^2] < \infty, E_q\left[\frac{\pi_n^*}{\pi_{n-1}^* q} h | x_{0:n-1}\right] \in A_{n-1}, E_q\left[\left(\frac{\pi_n^*}{\pi_{n-1}^* q} h\right)^2 | x_{0:n-1}\right] \in W_{n-1}\right\}$$

and $W_n = \left\{h : \Theta_n \to \mathbb{R} | E_{\pi_n^*}[|h|] < \infty, E_q\left[\left(\frac{\pi_n^*}{\pi_{n-1}^* q}\right)^2 |h| | x_{0:n-1}\right] \in W_{n-1}\right\}$,

where $W_0 = L^1(\pi_0)$.

(C2) The unit function $I_n : \Theta_n \to 1$ belongs to A_n.

Theorem 1 (Douc and Moulines 2008). Let $V_{2,0}(h) = \text{var}_{\pi_0}(h)$ and by induction, for $n > 0$, define

$$V_{1,n}(h) = V_{2,n-1}\left(E_q\left[h_n(x_{0:n})\right]\right) + E_{\pi_{n-1}^*}\left(\text{var}_q\left[h_n(x_{0:n})\right]\right),$$

where $h_n(x_{0:n}) = \frac{\pi_n^*(x_{0:n})(h(x_{1:n}) - \mu_n)}{\pi_{n-1}^*(x_{0:n-1})q(x_n|x_{0:n-1})}$,

$$V_{2,n}(h) = V_{1,n}(h(x_{1:n})) + \text{var}_{\pi_n^*}[h(x_{1:n})].$$

Suppose conditions (C1) and (C2) are satisfied. Then for any n, A_n and W_n are proper, and the following convergence results hold:

$$\frac{\sum_{j=1}^N h(x_n^{(j)})w_n^{(j)}}{\sum_{j=1}^N w_n^{(j)}} \xrightarrow{P} \mu_n, \forall h \in L^1(\pi_n),$$

$$\sqrt{N}\left[\frac{\sum_{j=1}^N h(x_n^{(j)})w_n^{(j)}}{\sum_{j=1}^N w_n^{(j)}} - \mu_n\right] \xrightarrow{\mathcal{L}} N(0, V_{1,n}(h)), \forall h \in A_n.$$

Specifically,

$$V_{1,n}(h) = \int \frac{\pi_n^*(x_{0:1})^2(\mu_1(x_{0:1}) - \mu_n)^2}{\pi_0(x_0)q(x_1|x_0)}dx_{0:1} + \sum_{t=2}^{n} \frac{\pi_n^*(x_{0:t})^2(\mu_t(x_{0:t}) - \mu_n)^2}{\pi_{t-1}^*(x_{0:t-1})q(x_t|x_{0:t-1})}dx_{0:t}, \quad (3)$$

where $\mu_t(x_{0:t}) = \int h(x_{1:n})\pi_n^*(x_{t+1:n}|x_{0:t})dx_{t+1:n}$ and $\pi_n^*(x_{0:t})$ is the marginal density of $\pi_n^*(x_{0:n})$.

Condition (C1) initializes the asymptotic results. The definition of proper sets $\{A_n\}$ assures that, for $h \in A_n$, under each of the sequential distributions $\{\pi_n^*\}$ and $\{\pi_{n-1}^* q\}$, the first two moments of $h(x_{1:n})$ and $\pi_n^* h(x_{1:n})/(\pi_{n-1}^* q)$ exist, which is required for the law of large number and asymptotic normality of a triangular array conditional on $\{x_{0:n-1}^{(j)}\}_{j=1}^N$. Condition (C2) are required for A_n to be proper. One useful interpretation of (3) is that each term can be interpreted as an importance sampling variance where the target integral is $\int \mu_t(x_{0:t})\pi_n^*(x_{0:t})dx_{0:t}$ and the importance distribution is $\pi_{t-1}^*(x_{0:t-1})q(x_t|x_{0:t-1})$ (Johansen and Doucet 2008). This interpretation allows us to consider the techniques of improving importance sampling in the design of SMC.

2.2 Importance Sampling With Multiple Proposals and Control Variates

Consider estimating integral $\mu = \int h(x)\pi^*(x)dx$, where $h(x)$ is a real function and the probability density $\pi^*(x)$ can only be evaluated up to a normalizing constant by $\pi(x) \propto \pi^*(x)$, and the importance proposal is $q(x)$. It is well known that a good $q(x)$ needs to satisfy two requirements: one is that the tail of $q(x)$ should be heavier than the tail of $\pi(x)$, so that the estimation variance is bounded, and the other is the shapes of $q(x)$ and $\pi(x)$ should be close. Using a mixture of multiple proposals has been shown to be helpful for satisfying these requirements (Hesterberg 1995). In addition, it is more efficient and safer to combine it with appropriates control variates (Owen and Zhou 2000; Tan 2004). Given a mixture of p proposal densities and a proportions vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)$ to form the mixture proposal density $q_{\boldsymbol{\alpha}}(x)$, stratified observations $\{x_{ki}\}_{i=1}^{n_k}$ are taken from the kth proposal $q_k(x)$ where $n_k = n\alpha_k$. With control variates $\boldsymbol{g}(x)/q_{\boldsymbol{\alpha}}(x)$, where $\boldsymbol{g}(x) = (q_2(x) - q_1(x), \ldots, q_p(x) - q_1(x))^T$, a natural extension of Owen and Zhou (2000)'s regression approach to estimate μ is

$$\widehat{\mu}_{\text{Reg}} = \frac{\frac{1}{n}\sum_{k=1}^p \sum_{i=1}^{n_k}\left[h(x_{ki})\pi(x_{ki}) - \widehat{\boldsymbol{\beta}}_1^T \boldsymbol{g}(x_{ki})\right]/q_{\boldsymbol{\alpha}}(x_{ki})}{\frac{1}{n}\sum_{k=1}^p \sum_{i=1}^{n_k}\left[\pi(x_{ki}) - \widehat{\boldsymbol{\beta}}_2^T \boldsymbol{g}(x_{ki})\right]/q_{\boldsymbol{\alpha}}(x_{ki})}, \quad (4)$$

where $\widehat{\boldsymbol{\beta}}_1 = \widetilde{\text{var}}\left[\frac{\boldsymbol{g}(X)}{q_{\boldsymbol{\alpha}}(X)}\right]^{-1} \widetilde{\text{cov}}^T\left[\frac{h(X)\pi(X)}{q_{\boldsymbol{\alpha}}(X)}, \frac{\boldsymbol{g}(X)}{q_{\boldsymbol{\alpha}}(X)}\right]$,

$$\widehat{\boldsymbol{\beta}}_2 = \widetilde{\text{var}}\left[\frac{\boldsymbol{g}(X)}{q_{\boldsymbol{\alpha}}(X)}\right]^{-1} \widetilde{\text{cov}}^T\left[\frac{\pi(X)}{q_{\boldsymbol{\alpha}}(X)}, \frac{\boldsymbol{g}(X)}{q_{\boldsymbol{\alpha}}(X)}\right],$$

and $\widetilde{\text{var}}$ and $\widetilde{\text{cov}}$ denote the pooled-sample variance and covariance.

Another estimator combining multiple proposals and control variates is proposed by Tan (2004) based on the perspective of Kong et al. (2003). In Kong et al. (2003), Monte Carlo integration was treated as a statistical inference problem where the Monte Carlo sample serves as observations and the under-

lying measure in target integral, usually Lebesgue or counting measure, is treated as an unknown nonnegative measure ν and modeled semiparametrically. Then by nonparametric maximum likelihood, ν is estimated by a discrete measure with the Monte Carlo sample as its support, and the target integral is estimated by an integration over this discrete measure. With multiple proposals, Tan (2004) proposed to restrict the measure ν to the set $\{\nu : \int q_k(x)dx = \int q_1(x)dv,\ k = 1, \ldots, p\}$. The nonparametric MLE of ν under such a restriction is

$$\widehat{\nu} \propto \frac{\widehat{P}(\{x\})}{q_{\boldsymbol{\alpha}}(x) + \widehat{\boldsymbol{\zeta}}^T \boldsymbol{g}(x)},$$
$$\text{where } \widehat{\boldsymbol{\zeta}} = \operatorname*{argmax}_{\boldsymbol{\zeta}} \sum_{k=1}^{p} \sum_{i=1}^{n_k} \log\left[q_{\boldsymbol{\alpha}}(x_{ki}) + \boldsymbol{\zeta}^T \boldsymbol{g}(x_{ki})\right].$$

Replacing the underlying measure of μ by $\widehat{\nu}$ gives an extension of Tan (2004)'s MLE approach to estimate μ as following:

$$\widehat{\mu}_{\text{MLE}} = \frac{\frac{1}{n}\sum_{k=1}^{p}\sum_{i=1}^{n_k} h(x_{ki})\pi(x_{ki}) / \left[q_{\boldsymbol{\alpha}}(x_{ki}) + \widehat{\boldsymbol{\zeta}}^T \boldsymbol{g}(x_{ki})\right]}{\frac{1}{n}\sum_{k=1}^{p}\sum_{i=1}^{n_k} \pi(x_{ki}) / \left[q_{\boldsymbol{\alpha}}(x_{ki}) + \widehat{\boldsymbol{\zeta}}^T \boldsymbol{g}(x_{ki})\right]}.$$

An interesting interpretation is that, by the construction of $\boldsymbol{g}$, the function $q_{\boldsymbol{\alpha}} + \widehat{\boldsymbol{\zeta}}^T \boldsymbol{g}$ is a weighted average of $q_1, \ldots, q_p$, with the weight of q_k equal to $\alpha_k + \hat{\zeta}_{k-1}$ for $k \geq 2$ and that of q_1 equal to $\alpha_1 - \hat{\zeta}_1 - \cdots - \hat{\zeta}_{p-1}$. It is shown that the MLE approach is optimal in a broad class of estimators with multiple proposals, and it is asymptotically equivalent to the regression approach in the first order, that is, they have the same asymptotic variance. For more discussion, see Li, Tan, and Chen (2013).

2.3 Generalized SMC With Mixture Proposal, Control Variates, and Auxiliary Function

Since SMC is a sequential implementation of importance sampling, it is natural to include multiple proposals and control variance in it. For example, a strategy is to use one proposal using the state equation and the other using the most recent observation (Thrun et al. 2001; Elinas, Sim, and Little 2006). For the basic SMC method, one implementation is to introduce control variates $\boldsymbol{S}(x_{1:n})$ satisfying $\int \boldsymbol{S}(x_{1:n})\pi_n^*(x_{0:n})dx_{0:n} = 0$ and replace the target function $h(x_{1:n})$ in $\widehat{\mu}_{n,\text{basic}}$ with $h(x_{1:n}) + \widehat{\boldsymbol{\gamma}}_n^T \boldsymbol{S}(x_{1:n})$ where $\widehat{\boldsymbol{\gamma}}_n$ is the estimated optimal coefficients (Singh et al. 2004). In this setting, the estimator is still asymptotically unbiased.

It is well known that in the selection (resampling) step, the resampling probability does not need to be proportional to the importance weight $w_n^{(j)}$ (Liu and Chen 1998; Carpenter, Clifford, and Fearnhead 1999; Pitt and Shephard 1999; Zhang et al. 2004; Lin, Chen, and Mykland 2010). Other resampling probability can be used to achieve various objectives. The popular auxiliary particle filter (APF) of Pitt and Shephard (1999) and Carpenter, Clifford, and Fearnhead (1999) use future observation y_{n+1} in the resampling. Specifically, since $\pi_{n+1}(x_{0:n})/\pi_n(x_{0:n})$ does not depend on x_{n+1}, by resampling according to $\{w_n^{(j)}\pi_{n+1}(x_{0:n}^{(j)})/\pi_n(x_{0:n}^{(j)})\}$ instead of $\{w_n^{(j)}\}$ in the resampling step, the selected $\{\widetilde{x}_{0:n}^{(j)}\}$ are more likely to reside in the high likelihood area of $\pi_{n+1}^*(x_{0:n})$. The analytical form of $\pi_{n+1}(x_{0:n})/\pi_n(x_{0:n})$ is usually unavailable in practice and some approximation $\eta(x_{0:n})$, named auxiliary function, is used instead. Douc, Moulines, and Olsson (2009) discussed several implementations of APF and derived the asymptotic properties of APF, based on which the optimal form of $\eta(x_{0:n})$ was given.

Combining these techniques, with $\eta(x_{0:n})$ being the auxiliary function, $\boldsymbol{S}(x_{0:n})$ being the control variates, and $q_{\boldsymbol{\alpha}}(x_n|x_{1:n-1})$ being the mixture proposal density where $q_{\boldsymbol{\alpha}}(x_n|x_{1:n-1}) = \sum_{k=1}^{p} \alpha_k q_k(x_n|x_{1:n-1})$ given p proposal densities $q_1(x_n|x_{1:n-1}), \ldots, q_p(x_n|x_{1:n-1})$ and mixture proportion vector $\boldsymbol{\alpha}$ at every time n, we have a more general SMC algorithm as follows:

Generalized SMC Method:

At time n, assume weighted samples $\{(\widetilde{x}_{0:n-1}^{(j)}; \widetilde{w}_{n-1}^{(j)})\}_{j=1}^N$ are available. For $j = 1, \ldots, N$,

1. Propagation (Mutation): Generate $x_n^{(j)}$ from the proposal distribution $q_{\boldsymbol{\alpha}}(x_n|\widetilde{x}_{0:n-1}^{(j)})$ and let $x_{0:n}^{(j)} = (\widetilde{x}_{0:n-1}^{(j)}, x_n^{(j)})$.
2. Weighting (Correction): Assign $x_{0:n}^{(j)}$ with weight

$$w_n^{(j)} = \widetilde{w}_{n-1}^{(j)} \frac{p\left(x_n^{(j)}|\widetilde{x}_{n-1}^{(j)}\right) p(y_n|x_n^{(j)})}{q_{\boldsymbol{\alpha}}\left(x_n^{(j)}|\widetilde{x}_{0:n-1}^{(j)}\right)}. \tag{5}$$

3. Resampling (Selection): If the condition for resampling is satisfied, resample $\{x_{0:n}^{(j)}\}_{j=1}^N$ according to $\{\eta(x_{0:n}^{(j)})w_n^{(j)}\}_{j=1}^N$ to obtain new weighted particles $\{(\widetilde{x}_{0:n}^{(j)}; \widetilde{\omega}_n^{(j)})\}_{j=1}^N$ where $\widetilde{\omega}_n^{(j)} = 1/\eta(\widetilde{x}_{0:n}^{(j)})$; if the condition for resampling is not satisfied, let $(\widetilde{x}_{0:n}^{(j)}; \widetilde{\omega}_n^{(j)}) = (x_{0:n}^{(j)}; \omega_n^{(j)})$.

Estimation is often made after the weighting step, but before the resampling step, where μ_n can be estimated by

$$\widehat{\mu}_{n,\text{generalized}} = \frac{\sum_{j=1}^{N}\left[h(x_n^{(j)}) + \widehat{\boldsymbol{\gamma}}_n^T \boldsymbol{S}(x_{0:n})\right]\omega_n^{(j)}}{\sum_{j=1}^{N}\omega_n^{(j)}}.$$

At the propagation step, particles can be generated in stratification to increase both computation and estimation efficiency.

The above generalized SMC method is more flexible and expected to perform better than the basic SMC, but there are several key implementation issues regarding the additional elements. First, it is necessary for the proposal densities to have heavier tails than $p(x_n|x_{n-1})p(y_n|x_n)$ so that sample weight $w_n^{(j)}$ has a bounded variance. But it is difficult to obtain an accurate approximation of $p(x_n|y_n, x_{n-1})$ with heavier tails, except in some special cases such as a log-concave $p(x_n|x_{n-1})p(y_n|x_n)$ approximated by a first-order Taylor expansion. Second, there is no general guidance on how to select appropriate control variates. The requirement $\int \boldsymbol{S}(x_{1:n})\pi_n^*(x_{0:n})dx_{0:n} = 0$ in Singh et al. (2004) is difficult to satisfy due to high dimension of $\boldsymbol{S}$. Third, the optimal auxiliary function given in Douc, Moulines, and Olsson (2009) is often not available; hence, a carefully constructed approximation is essential in the implementation.

In the next section, we propose a new algorithm that provides a unified strategy to design the proposals, control variates, and the auxiliary function, based on a novel theoretical understanding of the control variates.

3. LIKELIHOOD-BASED MIXTURE SMC

3.1 The New Algorithm

Given p proposal densities $q_1(x_n|x_{1:n-1}), \ldots, q_p(x_n|x_{1:n-1})$ and an auxiliary function $\eta(x_{0:n})$ for every n. For $t_0 < n$, let $q_k(x_{t_0+1:n}|x_{1:t_0}) = \prod_{t=t_0+1}^{n} q_k(x_t|x_{1:t-1})$ for $k = 1, \ldots, p$, $\boldsymbol{g}(x_{t_0+1:n}|x_{1:t_0}) = (q_1 - q_2, \ldots, q_1 - q_p)(x_{t_0+1:n}|x_{1:t_0})$ and $q_{\boldsymbol{\alpha}}(x_{t_0+1:n}|x_{1:t_0}) = \sum_{k=1}^{p} \alpha_k q_k(x_{t_0+1:n}|x_{1:t_0})$ with mixture proportion vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)$. The following algorithm is proposed to use multiple proposals and control variates in the SMC framework:

Likelihood-Based Mixture SMC (LM-SMC):

At time n, assume particles $\{\widetilde{x}_{0:n-1}^{(j)}\}_{j=1}^{N}$ and indicator sets $I_1, \ldots, I_p$ are available where $\cup_{k=1}^{p} I_k = \{1, \ldots, N\}$. Let n_0 be the time of last resampling. For $j \in I_k$,

1. Propagation (Mutation): Generate $x_n^{(j)}$ from the proposal $q_k(x_n|\widetilde{x}_{0:n-1}^{(j)})$ and let $x_{0:n}^{(j)} = (\widetilde{x}_{0:n-1}^{(j)}, x_n^{(j)})$.
2. Weighting (Correction): Assign $x_{0:n}^{(j)}$ with weight

$$v_n^{(j)} = \frac{\pi_n(x_{0:n}^{(j)})}{\pi_{n_0}\left(\widetilde{x}_{0:n_0}^{(j)}\right)\eta\left(\widetilde{x}_{0:n_0}^{(j)}\right)\left[q_{\boldsymbol{\alpha}}\left(x_{n_0+1:n}^{(j)}|\widetilde{x}_{0:n_0}^{(j)}\right) + \widehat{\boldsymbol{\zeta}}_n^T \boldsymbol{g}\left(x_{n_0+1:n}^{(j)}|\widetilde{x}_{0:n_0}^{(j)}\right)\right]}, \quad (6)$$

where

$$\widehat{\boldsymbol{\zeta}}_n = \underset{\boldsymbol{\zeta}}{\operatorname{argmax}} \sum_{j=1}^{N} \log\left[q_{\boldsymbol{\alpha}}\left(x_{n_0+1:n}^{(j)}|\widetilde{x}_{0:n_0}^{(j)}\right) + \boldsymbol{\zeta}^T \boldsymbol{g}\left(x_{n_0+1:n}^{(j)}|\widetilde{x}_{0:n_0}^{(j)}\right)\right].$$

3. Resampling (Selection): If the condition for resampling is satisfied, resample $\{x_{0:n}^{(j)}\}_{j=1}^{N}$ according to $\{\eta(x_{0:n}^{(j)})v_n^{(j)}\}_{j=1}^{N}$ to obtain new particles $\{\widetilde{x}_{0:n}^{(j)}\}_{j=1}^{N}$, each with weight $1/N$, and divide $\{1, \ldots, N\}$ with equal probabilities into new indicator sets $I_1, \ldots, I_p$ satisfying $\#I_k = \alpha_k N$; if resampling is not needed, let $\widetilde{x}_{0:n}^{(j)} = x_{0:n}^{(j)}$.

After the weighting step, estimate μ_n by

$$\widehat{\mu}_{n,\mathrm{MLE}} = \frac{\sum_{j=1}^{N} h(x_n^{(j)})v_n^{(j)}}{\sum_{j=1}^{N} v_n^{(j)}}.$$

Remark 1. One of the novelties of this algorithm is that the control variates are included in both resampling and estimation. An alternative implementation is, in the generalized SMC method, to estimate μ_n by Owen and Zhou's estimator $\widehat{\mu}_{\mathrm{Reg}}$ in (4) as follows:

$$\widehat{\mu}_{n,\mathrm{Reg}} = \frac{\sum_{j=1}^{N}\left[h(x_n^{(j)})w_n^{(j)} - \widehat{\boldsymbol{\tau}}_{1n}^T \boldsymbol{g}(x_{n_0+1:n}^{(j)}|\widetilde{x}_{0:n_0}^{(j)})/q_{\boldsymbol{\alpha}}(x_{n_0+1:n}^{(j)}|\widetilde{x}_{0:n_0}^{(j)})\right]}{\sum_{j=1}^{N}\left[w_n^{(j)} - \widehat{\boldsymbol{\tau}}_{2n}^T \boldsymbol{g}(x_{n_0+1:n}^{(j)}|\widetilde{x}_{0:n_0}^{(j)})/q_{\boldsymbol{\alpha}}(x_{n_0+1:n}^{(j)}|\widetilde{x}_{0:n_0}^{(j)})\right]} \quad (7)$$

where

$$\widehat{\boldsymbol{\tau}}_{1n} = \widehat{\mathrm{var}}\left[\frac{\boldsymbol{g}}{q_{\boldsymbol{\alpha}}}\right]^{-1} \widehat{\mathrm{cov}}\left[\frac{h\pi_n}{\pi_{n_0} c q_{\boldsymbol{\alpha}}}, \frac{\boldsymbol{g}}{q_{\boldsymbol{\alpha}}}\right], \quad \widehat{\boldsymbol{\tau}}_{2n} = \widehat{\mathrm{var}}\left[\frac{\boldsymbol{g}}{q_{\boldsymbol{\alpha}}}\right]^{-1} \times \widehat{\mathrm{cov}}\left[\frac{\pi_n}{\pi_{n_0} c q_{\boldsymbol{\alpha}}}, \frac{\boldsymbol{g}}{q_{\boldsymbol{\alpha}}}\right],$$

We call this regression-based mixture SMC (RM-SMC). It only uses the control variates in the estimation without changing the distribution of particles. Although the regression approach assigns proper importance weights for each particle, they are not necessarily positive and hence cannot be directly used in resampling. The likelihood approach always gives positive importance weights, after incorporating the effect of control variates. The asymptotic results in the next section show that the new algorithm outperforms both the generalized SMC method without control variates and RM-SMC.

Remark 2. In the algorithm, particles are mutated within each group and are only mixed across groups at the time of resampling. Hence, the proposal distribution of $x_{n_0+1:n}$ is the mixture of p distributions and the number of control variates in $\boldsymbol{g}$ is $p - 1$. Similarly as mentioned in Section 2.2, the function $q_{\boldsymbol{\alpha}} + \widehat{\boldsymbol{\zeta}}_n^T \boldsymbol{g}$ is a weighted average of $q_1, \ldots, q_p$, with the weights summing to 1 but different from $\boldsymbol{\alpha}$.

Remark 3. For generalized SMC method, the control variates $\boldsymbol{S}(x_{1:n})$ need to satisfy $\int \boldsymbol{S}(x_{1:n})\pi_n^*(x_{0:n})dx_{0:n} = 0$ to make the estimator asymptotically unbiased, which makes the construction of $\boldsymbol{S}(x_{1:n})$ difficult. The new algorithm avoids the difficulty and is easy to implement.

3.2 Theoretical Results

Here, we present a CLT for $\widehat{\mu}_{n,\mathrm{MLE}}$, similar to that in Douc and Moulines (2008). For simplicity, only the scheme of multinomial resampling at every step is discussed. Let $\widetilde{v}_n^{(j)} = \eta(x_{0:n}^{(j)})v_n^{(j)}$ be the weights used in resampling, $\widetilde{\pi}_n(x_{0:n}) = \eta(x_{0:n})\pi_n(x_{0:n})$ be the unnormalized new target density after resampling with the auxiliary function, $\widetilde{\pi}_n^*(x_{0:n}) = e_n^{-1}\widetilde{\pi}_n(x_{0:n})$ where e_n is the normalizing constant, and $\widetilde{\mu}_n = \int h(x_{1:n})\widetilde{\pi}_n^*(x_{0:n})dx_{0:n}$. Let $A_0' = A_0$, $W_0' = W_0$ and for $n > 0$, recursively define

$$\begin{aligned}
A_n' = \Big\{h : \Theta_n \to \mathbb{R} \,\Big|\, & E_{\pi_n^*}[h^2] < \infty, E_{\widetilde{\pi}_n^*}[h^2] < \infty, \\
& E_{q_{\boldsymbol{\alpha}}}\left[\frac{\pi_n^*}{\widetilde{\pi}_{n-1}^* q_{\boldsymbol{\alpha}}}|h||x_{0:n-1}\right] \in A_{n-1}', \\
& E_{q_{\boldsymbol{\alpha}}}\left[\frac{\widetilde{\pi}_n^*}{\widetilde{\pi}_{n-1}^* q_{\boldsymbol{\alpha}}}|h||x_{0:n-1}\right] \in A_{n-1}', \\
& E_{q_{\boldsymbol{\alpha}}}\left[\left(\frac{\pi_n^*}{\widetilde{\pi}_{n-1}^* q_{\boldsymbol{\alpha}}}h\right)^2 |x_{0:n-1}\right] \in W_{n-1}', \\
& E_{q_{\boldsymbol{\alpha}}}\left[\left(\frac{\widetilde{\pi}_n^*}{\widetilde{\pi}_{n-1}^* q_{\boldsymbol{\alpha}}}h\right)^2 |x_{0:n-1}\right] \in W_{n-1}'\Big\} \\
W_n' = \Big\{h : \Theta_n \to \mathbb{R} \,\Big|\, & E_{\pi_n^*}[|h|] < \infty, E_{\widetilde{\pi}_n^*}[|h|] < \infty, \\
& E_{q_{\boldsymbol{\alpha}}}\left[\left(\frac{\pi_n^*}{\widetilde{\pi}_{n-1}^* q_{\boldsymbol{\alpha}}}\right)^2 |h||x_{0:n-1}\right] \in W_{n-1}', \\
& E_{q_{\boldsymbol{\alpha}}}\left[\left(\frac{\widetilde{\pi}_n^*}{\widetilde{\pi}_{n-1}^* q_{\boldsymbol{\alpha}}}\right)^2 |h||x_{0:n-1}\right] \in W_{n-1}'\Big\}.
\end{aligned}$$

Let $\sigma_{2,0}^2(h) = \mathrm{var}_{\pi_0}[h]$ and for $n > 0$, recursively define

$$\begin{aligned}
\sigma_{1,n}^2(h) &= \sigma_{2,n-1}^2\left(E_{q_{\boldsymbol{\alpha}}}[h_n(x_{0:n})]\right) + E_{\widetilde{\pi}_{n-1}^*}\left(\mathrm{var}_{q_{\boldsymbol{\alpha}}}[h_n(x_{0:n}) - \boldsymbol{\beta}_n^T \boldsymbol{g}_n(x_n|x_{0:n-1})]\right), \\
\text{where } h_n(x_{0:n}) &= \frac{\pi_n^*(x_{0:n})(h - \mu_n)}{\widetilde{\pi}_{n-1}^*(x_{0:n-1})q_{\boldsymbol{\alpha}}(x_n|x_{0:n-1})}, \boldsymbol{g}_n(x_n|x_{0:n-1}) \\
&= \frac{\boldsymbol{g}(x_n|x_{0:n-1})}{q_{\boldsymbol{\alpha}}(x_n|x_{0:n-1})} \text{ and } \boldsymbol{\beta}_n = \mathrm{var}[\boldsymbol{g}_n]^{-1}\mathrm{cov}[h_n, \boldsymbol{g}_n], \\
\sigma_{2,n}^2(h) &= \sigma_{2,n-1}^2\left(E_{q_{\boldsymbol{\alpha}}}\left[\widetilde{h}_n(x_{0:n})\right]\right) + E_{\widetilde{\pi}_{n-1}^*}\left(\mathrm{var}_{q_{\boldsymbol{\alpha}}}\left[\widetilde{h}_n(x_{0:n}) - \widetilde{\boldsymbol{\beta}}_n^T \boldsymbol{g}_n(x_n|x_{0:n-1})\right]\right) + \mathrm{var}_{\widetilde{\pi}_n^*}[h], \\
\text{where } \widetilde{h}_n(x_{0:n}) &= \frac{\widetilde{\pi}_n^*(x_{0:n})(h - \widetilde{\mu}_n)}{\widetilde{\pi}_{n-1}^*(x_{0:n-1})q_{\boldsymbol{\alpha}}(x_n|x_{0:n-1})} \text{ and} \\
\widetilde{\boldsymbol{\beta}}_n &= \mathrm{var}[\boldsymbol{g}_n]^{-1}\mathrm{cov}[\widetilde{h}_n, \boldsymbol{g}_n].
\end{aligned}$$

Assume that:

(C2′) The unit function $I_n : \Theta_n \to 1$ belongs to A'_n.
(C3) $\text{var}_{\widetilde{\pi}^*_n q_{\boldsymbol{\alpha}}}[\boldsymbol{g}/q_{\boldsymbol{\alpha}}] < \infty$ and is positive definite.
(C4) $E_{\widetilde{\pi}^*_n q_{\boldsymbol{\alpha}}}[g_{i_1} g_{i_2} g_{i_3}/q^3_{\boldsymbol{\alpha}}] < \infty$ for any three coordinates $g_{i_1}, g_{i_2}, g_{i_3}$ of $\boldsymbol{g}$.
(C5) $\int \boldsymbol{g}^2(x_n|x_{0:n-1})/q_{\boldsymbol{\alpha}}(x_n|x_{0:n-1})dx_n \in W'_{n-1}$.
(C6) $\int \boldsymbol{g}(x_n|x_{0:n-1})dx_n \equiv 0$.

Theorem 2. Suppose conditions (C1), (C2′), and (C3)–(C6) are satisfied. Then for any n, A'_n and W'_n are proper, $\sigma^2_{1,n}(h)$ is finite, and the following convergence holds:

$$\widehat{\mu}_{n,\text{MLE}} \xrightarrow{P} \mu_n, \forall h \in L^1(\pi^*_n), \tag{8}$$

$$\sqrt{N}\left(\widehat{\mu}_{n,\text{MLE}} - \mu_n\right) \xrightarrow{\mathcal{L}} N\left(0, \sigma^2_{1,n}(h)\right), \forall h \in A'_n. \tag{9}$$

Specifically,

$$\sigma^2_{1,n}(h) = \sum_{t=1}^{n} \int \times \frac{\left[\pi^*_n(x_{0:t})(\mu_t(x_{0:t}) - \mu_n) - \boldsymbol{\beta}^T_{tn}\widetilde{\pi}^*_{t-1}(x_{0:t-1})\boldsymbol{g}(x_t|x_{0:t-1})\right]^2}{\widetilde{\pi}^*_{t-1}(x_{0:t-1})q_{\boldsymbol{\alpha}}(x_t|x_{0:t-1})} dx_{0:t}, \tag{10}$$

where

$$\boldsymbol{\beta}_{tn} = \text{var}_{\widetilde{\pi}^*_{t-1}q_{\boldsymbol{\alpha}}}\left[\frac{\boldsymbol{g}(x_t|x_{0:t-1})}{q_{\boldsymbol{\alpha}}(x_t|x_{0:t-1})}\right]^{-1} \text{cov}_{\widetilde{\pi}^*_{t-1}q_{\boldsymbol{\alpha}}} \times \left[\frac{\pi^*_n(x_{0:t})(\mu_t(x_{0:t}) - \mu_n)}{\widetilde{\pi}^*_{t-1}(x_{0:t-1})q_{\boldsymbol{\alpha}}(x_t|x_{0:t-1})}, \frac{\boldsymbol{g}(x_t|x_{0:t-1})}{q_{\boldsymbol{\alpha}}(x_t|x_{0:t-1})}\right],$$

and $\widetilde{\pi}^*_0(x_0) = \pi^*_0(x_0)$.

The proof of the theorem is given in the Appendix.

The sets A'_n and W'_n are defined similarly as A_n and W_n so that for $h \in A'_n$, under each of the sequential distributions $\{\pi^*_n\}$, $\{\widetilde{\pi}^*_n\}$, $\{\pi^*_{n-1}q\}$, and $\{\widetilde{\pi}^*_{n-1}q\}$, the asymptotic results of certain triangular arrays conditional on $\{x^{(j)}_{0:n}\}^N_{j=1}$ or $\{\widetilde{x}^{(j)}_{0:n-1}\}^N_{j=1}$ hold. One sufficient condition for $h \in A'_n$ is

$$\frac{p(x_n|x_{n-1})p(y_n|x_n)}{\eta(x_{0:n-1})q_{\boldsymbol{\alpha}}(x_n|x_{0:n-1})} \text{ and } \frac{\eta(x_{0:n})p(x_n|x_{n-1})p(y_n|x_n)}{\eta(x_{0:n-1})q_{\boldsymbol{\alpha}}(x_n|x_{0:n-1})} \text{ are bounded from above,} \tag{11}$$

which can be satisfied by including heavy tail component proposals, and the first two moments of h under densities π^*_n, $\widetilde{\pi}^*_n$, $\pi^*_{n-1}q$, and $\widetilde{\pi}^*_{n-1}q$ exist for every n. Condition (C3) ensures that the optimization for MLE weights $\widehat{\boldsymbol{\zeta}}$ gives stable results. Condition (C4) is required for the remainder of certain Taylor expansion to be small. Conditions (C3)–(C5) can be satisfied when all mixture proportions of $q_{\boldsymbol{\alpha}}$ are nonzero, or the tails of component proposals with nonzero proportions cover those of all other component proposals. Condition (C6) is necessary for $\sigma^2_{1,n}(h)$ to have the clean expression in Theorem 2, which is automatically satisfied in LM-SMC and is stated here to emphasize the requirement for control variates.

The analytical form of $\sigma^2_{2,n}(h)$ clearly demonstrates the impact of the control variates in two aspects. First, every term in the asymptotic variance contains the control variates, which is resulted from including control variates in the resampling step. Second, the coefficient vector for control variates is optimal for every n despite the target density for each term changes as n increases. This is because in the likelihood approach, the estimated coefficient vector $\widehat{\boldsymbol{\zeta}}_n$ does not depend on the target density, and hence the same sample and coefficients can be adapted for different target density automatically. Therefore, LM-SMC outperforms the generalized SMC method without using control variates. Since RM-SMC only includes the control variates in the estimator, it only affects the last term of the asymptotic variance (shown in the Appendix). Therefore, LM-SMC also outperforms RM-SMC. Section 4.1.3 gives a specific example illustrating these comparisons on variance expansions.

The following corollary shows that the asymptotic variance of LM-SMC is always smaller than that of the APF using $N\alpha_k$ particles from the proposal q_k, for any $k = 1, \ldots, p$.

Corollary 1. For any $k = 1, \ldots, p$ such that $\alpha_k > 0$,

$$\sigma^2_{1,n}(h) \le \frac{1}{\alpha_k}\sum_{t=1}^{n} \int \frac{\left[\pi^*_n(x_{0:t})(\mu_t(x_{0:t}) - \mu_n)\right]^2}{\widetilde{\pi}^*_{t-1}(x_{0:t-1})q_k(x_t|x_{0:t-1})} dx_{0:t}.$$

3.3 Guideline on Choosing Component Proposals and Auxiliary Function

The choices of component proposals $q_1, \ldots, q_p$ are critical to the performance of LM-SMC. A common criterion of selecting proposal is the conditional variance of importance weights given the previous weighted particles, to limit particle degeneracy. Alternatively, at each time step n, we select the component proposals so that $\mu_{n,\text{MLE}}$ has the smallest asymptotic variance $\sigma_{1,n}(h)$ at this single step than all other choices of proposals. This strategy is similar to the one adapted in Douc, Moulines, and Olsson (2009) for selecting auxiliary function. Note that in (10), at time step n, the choice of $q_1(x_n|x_{0:n-1}), \ldots, q_p(x_n|x_{0:n-1})$ and $\eta(x_{0:n-1})$ only affects the last term of $\sigma_{1,n}(h)$ which is

$$\int \frac{\left[\pi^*_n(x_{0:n})(h(x_{1:n}) - \mu_n) - \boldsymbol{\beta}^T_{nn}\boldsymbol{g}(x_n|x_{0:n-1})\widetilde{\pi}^*_{n-1}(x_{0:n-1})\right]^2}{\widetilde{\pi}^*_{n-1}(x_{0:n-1})q_{\boldsymbol{\alpha}}(x_n|x_{0:n-1})} dx_{0:t}, \tag{12}$$

and therefore is to be minimized. By the theory of the likelihood approach (Tan 2004), (12) is equal to 0 when $\pi^*_n(h - \mu_n)$ is a linear combination of $\widetilde{\pi}^*_{n-1}q_1, \ldots, \widetilde{\pi}^*_{n-1}q_p$. This property suggests to choose $q_1, \ldots, q_p$ and $\eta(x_{0:n-1})$ such that $\pi^*_n(h - \mu_n)$ is close to some linear combination of $\widetilde{\pi}^*_{n-1}q_1, \ldots, \widetilde{\pi}^*_{n-1}q_p$. Since $\pi^*_n(x_{0:n})(h(x_{1:n}) - \mu_n) \propto \pi^*_{n-1}(x_{0:n-1})p(x_n|x_{n-1})p(y_n|x_n)(h(x_{1:n}) - \mu_n)$, the construction of component proposals can be done by approximating and decomposing $p(x_n|x_{n-1})p(y_n|x_n)(h - \mu_n)$. Specifically, if $p(x_n|x_{n-1})p(y_n|x_n)$ can be approximated by $\eta(x_{0:n-1})q(x_n|x_{0:n-1})$ where $\eta(x_{0:n-1})$ is a positive function and $q(x_n|x_{0:n-1})$ is a probability density having a decomposition

$$q(x_n|x_{0:n-1})(h(x_{1:n}) - \mu_n) = \sum_{k=1}^{p} c_k r_k(x_{1:n-1})q_k(x_n|x_{0:n-1}) \tag{13}$$

with real functions $r_k(x_{0:n-1})$, densities $q_k(x_n|x_{0:n-1})$, and constants c_k, then $q_1, \ldots, q_p$ and $\eta(x_{0:n-1})$ can be used as the component proposals and the auxiliary function.

Meanwhile, to ensure the sample weights have bounded variance, one should include a heavy tail component proposal and add a positive constant into $\eta(x_{0:n-1})$ to have $\eta(x_{0:n-1})$ bounded away from 0, so that the sufficient condition (11) is satisfied. A simple choice of the heavy tailed proposal is the prior density $p(x_n|x_{n-1})$ from the state equation.

The above strategy can be illustrated using the following special case of model (1),

$$\begin{aligned} x_n &= s_1(x_{n-1}) + \varepsilon_n \\ y_n &= s_2(x_n) + e_n \end{aligned} \tag{14}$$

with normal noises ε_n and e_n and a nonlinear function $s_2(x)$. Suppose the target function $h(x_{1:n})$ is x_n.

Let $\log(p_2(y_n|x_n))$ be the second-order Taylor expansion of $\log(p(y_n|x_n))$ around the mode of $p(y_n|x_n)$, $q_2(x_n|y_n, x_{n-1})$ be the normalized $p(x_n|x_{n-1})p_2(y_n|x_n)$, and $r(x_{n-1})$ be the corresponding normalizing constant. By approximating the center of normalized $p(x_n|x_{n-1})p(y_n|x_n)$ by $q_2(x_n|y_n, x_{n-1})$, controlling its tail by $p(x_n|x_{n-1})$, and adding a positive constant c to $r(x_{n-1})$, $p(x_n|x_{n-1})p(y_n|x_n)$ can be approximated by

$$\begin{aligned} \eta(x_{0:n-1})q(x_n|x_{0:n-1}) \equiv\ & [r(x_{n-1}) + c]\,[\gamma_1 q_2(x_n|y_n, x_{n-1}) \\ & + \gamma_2 p(x_n|x_{n-1})], \end{aligned} \tag{15}$$

where the values of γ_1 and γ_2 can be arbitrary. The value c should not be too large or too small compared to $r(x_{n-1})$. The expectation $E_{\pi^*_{n-2}q_{\boldsymbol{\alpha}}}[r(x_{n-1})]$ is a reasonable choice and can be estimated by sample average.

Then, the component proposal can be constructed as follows. Note that $q_2(x_n|y_n, x_{n-1})$ is a normal density with mean $\theta(x_{n-1})$. Decompose $q(x_n|x_{0:n-1})(x_n - \mu_n)$ by

$$\begin{aligned} & [\gamma_1 q_2(x_n|y_n, x_{n-1}) + \gamma_2 p(x_n|x_{n-1})]\,(x_n - \mu_n) \\ &\quad = \gamma_1\,[x_n - \theta(x_{n-1})]\,q_2(x_n|y_n, x_{n-1}) + \gamma_1\,[\theta(x_{n-1}) \\ &\qquad - \mu_n]\,q_2(x_n|y_n, x_{n-1}) \\ &\qquad + \gamma_2\,[x_n - s_1(x_{n-1})]\,p(x_n|x_{n-1}) + \gamma_2\,[s_1(x_{n-1}) \\ &\qquad - \mu_n]\,p(x_n|x_{n-1}) \end{aligned} \tag{16}$$

and use the following component proposal distributions: the normalized $[x_n - \theta(x_{n-1})]^+ q_2(x_n|y_n, x_{n-1})$, normalized $[x_n - \theta(x_{n-1})]^- q_2(x_n|y_n, x_{n-1})$, $q_2(x_n|y_n, x_{n-1})$, normalized $[x_n - s_1(x_{n-1})]^+ p(x_n|x_{n-1})$, normalized $[x_n - s_1(x_{n-1})]^- p(x_n|x_{n-1})$ and $p(x_n|x_{n-1})$. They are either Weibull distribution or Normal distribution, which can be sampled from easily. When the mixture proportion for $p(x_n|x_{n-1})$ is nonzero ($\alpha_6 > 0$), then (11) is satisfied by the fact that $\eta(x_{0:n-1})$ is bounded and $\eta(x_{0:n-1})q_{\boldsymbol{\alpha}}(x_n|x_{0:n-1}) > \alpha_6 c p(x_n|x_{n-1})$.

3.4 The Choice of Mixture Proportions

Based on the theoretical results, the following guidelines are useful for choosing the mixture proportions for the candidates component proposals.

First, the proportions of some proposals can be set to zero, but these proposals remain as part of the control variates. That is, some of the proposal distributions are not used in generating samples, but only used as control variates. Typically, these are part of (13) but are computationally expensive to sample from, though easy to evaluate. This is valid because when $\pi^*_n(h - \mu_n)$ is a linear combination of $\widetilde{\pi}^*_{n-1}q_1, \ldots, \widetilde{\pi}^*_{n-1}q_p$, (12) still equals 0 even if some α_i are zero. This strategy allows more flexibility in decomposition (13) to include terms which are easy to be normalized but difficult to sample from. Then, the values of nonzero proportions can follow some heuristic rules derived from experience or interpretation of proposals. Equal proportions are often a good starting point.

Second, the component proposals with nonzero proportions should satisfy conditions (C3)–(C5). These conditions hold when $q_{\boldsymbol{\alpha}}$ has a heavier tail than those proposals with zero proportions. That is, the component for protection should have nonzero proportion.

Third, the proportion of a very good component proposal (if there is one) can be close to 1 so that the performance of LM-SMC is at least not worse than the performance of that component proposal. This is ensured by Corollary 1.

The first strategy raised the question of whether sampling from a subset of $q_1, \ldots, q_p$ will decrease the estimation efficiency and offset the saving of computational resource. As we illustrate through the following example, when the sample coverage is not of a major concern, the improvement by the likelihood approach may be mainly caused by the use of control variates. Therefore in practice, one can construct several easy-to-sample proposals to cover the high likelihood area of the target density with some additional more sophisticated proposal densities as covariates to achieve more accurate approximation. The following example is considered for importance sampling, though similar features can be seen in SMC as well.

Example. Suppose a random vector $\boldsymbol{X} = (X_1, \ldots, X_{10})$ follows

$$\pi(\boldsymbol{x}) = 0.8 \prod_{p=1}^{10} \phi(x_p) + 0.2 \prod_{p=1}^{10} \psi_4(x_p),$$

where $\phi(x)$ is the standard normal density and $\psi_d(x)$ is the student t density with degrees of freedom d. This distribution is used in Tan (2004). The target of interest is the expectation $\mu = E\,[f(\boldsymbol{X})]$ under $\pi(\boldsymbol{x})$, where $f(\boldsymbol{x}) = \sum_{p=1}^{10} x_p/10$. Here, we compare four estimators to illustrate the effects of setting some mixture proportions to 0.

The first two estimators are based on the proposal choices in Tan (2004). Let $q_1(\boldsymbol{x}) = \prod_{p=1}^{10} \phi(x_p)$, $q_2(\boldsymbol{x}) = \prod_{p=1}^{10} \psi_2(x_p)$, $g_1(\mathbf{x}) = q_1(\boldsymbol{x}) - q_2(\boldsymbol{x})$, and $q_{.5,2}(\boldsymbol{x}) = 0.5q_1(\boldsymbol{x}) + 0.5q_2(\boldsymbol{x})$ where $q_{\boldsymbol{\alpha},k}(\boldsymbol{x})$ denotes the mixture of k proposals with proportions $\boldsymbol{\alpha}$. Among the component proposals, q_1 approximates the center of π and q_2 controls the tail of π. Suppose $\{\boldsymbol{x}_i\}_{i=1}^n$ are generated from $q_{.5,2}(\boldsymbol{x})$ in stratification. Then IS and Tan's likelihood approach give the estimator

$$\begin{aligned} \widehat{\mu}_{P_2} &= \frac{\sum_{i=1}^n f(\boldsymbol{x}_i)\pi(\boldsymbol{x}_i)/q_{.5,2}(\boldsymbol{x}_i)}{\sum_{i=1}^n \pi(\boldsymbol{x}_i)/q_{.5,2}(\boldsymbol{x}_i)}, \\ \widehat{\mu}_{P_2C_1} &= \frac{\sum_{i=1}^n f(\boldsymbol{x}_i)\pi(\boldsymbol{x}_i)/\left[q_{.5,2}(\boldsymbol{x}_i) + \widehat{\zeta}_{P_2C_1} g_1(\boldsymbol{x}_i)\right]}{\sum_{i=1}^n \pi(\boldsymbol{x}_i)/\left[q_{.5,2}(\boldsymbol{x}_i) + \widehat{\zeta}_{P_2C_1} g_1(\boldsymbol{x}_i)\right]}, \end{aligned}$$

where $\widehat{\zeta}_{P_2C_1} = \underset{\zeta}{\operatorname{argmax}} \sum_{i=1}^n \log\left(q_{.5,2}(\boldsymbol{x}_i) + \zeta g_1(\boldsymbol{x}_i)\right)$ and $\widehat{\mu}_{P_kC_j}$ denotes the estimator with k sampling proposals and j control variates.

The third and fourth estimators are based on more sophisticated component proposals following the discussions in Li, Tan, and Chen (2013) which suggested to decompose an approximation of $(f(\boldsymbol{x}) - \mu)\pi(\boldsymbol{x})$, similar to the discussions in Section 3.3. By approximating $\pi(\boldsymbol{x})$ with the mixture of $q_1(\boldsymbol{x})$ and $q_2(\boldsymbol{x})$

as in the first two estimators,

$$\left(\frac{1}{10}\sum_{p=1}^{10}x_p-\mu\right)\pi(\boldsymbol{x})\approx\left(\frac{1}{10}\sum_{p=1}^{10}x_p-\mu\right)[\gamma_1 q_1(\boldsymbol{x})+\gamma_2 q_2(\boldsymbol{x})]$$
$$=\sum_{p=1}^{10}\tau_{1p}x_p^+\phi(x_p)\prod_{i\neq p}\phi(x_i)-\sum_{p=1}^{10}\tau_{2p}x_p^-\phi(x_p)\prod_{i\neq p}\phi(x_i)$$
$$+\sum_{p=1}^{10}\tau_{3p}x_p^+\psi_2(x_p)\prod_{i\neq p}\psi_2(x_i)-\sum_{p=1}^{10}\tau_{4p}x_p^-\psi_2(x_p)\prod_{i\neq p}\psi_2(x_i)$$
$$+\tau_5 q_1(\boldsymbol{x})+\tau_6 q_2(\boldsymbol{x}),$$

where γ_1, γ_2, $\tau_1,\ldots,\tau_6$ are constants. Therefore, we choose the following component proposals: $q_{1j+}(\boldsymbol{x})\propto x_j^+\phi(x_j)\prod_{i\neq j}\phi(x_i)$, $q_{1j-}(\boldsymbol{x})\propto x_j^-\phi(x_j)\prod_{i\neq j}\phi(x_i)$, $q_{2j+}(\boldsymbol{x})\propto x_j^+\psi_2(x_j)\prod_{i\neq j}\psi_2(x_i)$, $q_{2j-}(\boldsymbol{x})\propto x_j^-\psi_2(x_j)\prod_{i\neq j}\psi_2(x_i)$, for $j=1,\ldots,10$ and $q_1(\boldsymbol{x})$ and $q_2(\boldsymbol{x})$. There are a total of 42 component proposals and all can be sampled relatively easily. Let $q_{\boldsymbol{\alpha},42}(\boldsymbol{x})$ be the mixture of these component proposals with equal proportions and $\boldsymbol{g}(\boldsymbol{x})$ be the corresponding vector of control variates. Suppose $\{\boldsymbol{x}_i\}$ are generated from $q_{\boldsymbol{\alpha},42}(\boldsymbol{x})$ in stratification. The likelihood approach gives the third estimator

$$\widehat{\mu}_{P_{42}C_{41}}=\frac{\sum_{i=1}^n f(\boldsymbol{x}_i)\pi(\boldsymbol{x}_i)/\left[q_{\boldsymbol{\alpha},42}(\boldsymbol{x}_i)+\widehat{\boldsymbol{\zeta}}_{P_{42}C_{41}}^T\boldsymbol{g}(\boldsymbol{x}_i)\right]}{\sum_{i=1}^n \pi(\boldsymbol{x}_i)/\left[q_{\boldsymbol{\alpha},42}(\boldsymbol{x}_i)+\widehat{\boldsymbol{\zeta}}_{P_{42}C_{41}}^T\boldsymbol{g}(\boldsymbol{x}_i)\right]},$$

where $\widehat{\boldsymbol{\zeta}}_{P_{42}C_{41}}=\operatorname{argmax}_{\boldsymbol{\zeta}}\sum_{i=1}^n\log\left(q_{\boldsymbol{\alpha},42}(\boldsymbol{x}_i)+\boldsymbol{\zeta}^T\boldsymbol{g}(\boldsymbol{x}_i)\right)$.

The fourth estimator is constructed by setting the mixture proportions of all component proposals except $q_1(\boldsymbol{x})$ and $q_2(\boldsymbol{x})$ to 0 in $\widehat{\mu}_{P_{42}C_{41}}$, which means the sample is only generated from $q_{.5,2}(\boldsymbol{x})$ but the control variates $\boldsymbol{g}(\boldsymbol{x})$ uses all components. This gives the fourth estimator

$$\widehat{\mu}_{P_2C_{41}}=\frac{\sum_{i=1}^n f(\boldsymbol{x}_i)\pi(\boldsymbol{x}_i)/\left[q_{.5,2}(\boldsymbol{x}_i)+\widehat{\boldsymbol{\zeta}}_{P_2C_{41}}^T\boldsymbol{g}(\boldsymbol{x}_i)\right]}{\sum_{i=1}^n \pi(\boldsymbol{x}_i)/\left[q_{.5,2}(\boldsymbol{x}_i)+\widehat{\boldsymbol{\zeta}}_{P_2C_{41}}^T\boldsymbol{g}(\boldsymbol{x}_i)\right]},$$

where $\widehat{\boldsymbol{\zeta}}_{P_2C_{41}}=\operatorname{argmax}_{\boldsymbol{\zeta}}\sum_{i=1}^n\log\left(q_{.5,2}(\boldsymbol{x}_i)+\boldsymbol{\zeta}^T\boldsymbol{g}(\boldsymbol{x}_i)\right)$. Note that $\widehat{\mu}_{P_2C_{41}}$ has a bounded variance since the heavy tailed proposal $q_2(\boldsymbol{x})$ is included.

The MSE of these four estimators are reported in Table 1. It can be seen that $\widehat{\mu}_{P_2}$ and $\widehat{\mu}_{P_2C_1}$ have similar MSEs, indicating no improvement by a control variate $g_1(\boldsymbol{x})$ without considering the target function $f(\boldsymbol{x})$. When control variates constructed using the information of $f(\boldsymbol{x})$ are included, the resulted estimator $\widehat{\mu}_{P_2C_{41}}$ improves the MSE of $\widehat{\mu}_{P_2}$ by more than one order of magnitude. If the sampling proposals also use the information of $f(\boldsymbol{x})$, the resulting estimator $\widehat{\mu}_{P_{42}C_{41}}$ improves the MSE of $\widehat{\mu}_{P_2C_{41}}$ by about 20%. It shows that the main contribution of improvement comes from the control variates instead of proposal distributions.

4. NUMERICAL STUDIES

Here, we present several examples to illustrate the performance of the new algorithm. In these examples, the target function is $h(x_n)=x_n$ and the posterior mean of the current state is of interest. We compare five methods: The BF, APF, the generalized SMC method without using control variates (GSMC), RM-SMC of (7), and LM-SMC. The generalized SMC used here does not include control variates due to the difficulties discussed in Remark 3 of Section 3.1. In all examples, systematic resampling (Carpenter, Clifford, and Fearnhead 1999; Douc and Cappé 2005) is used at every step, each step uses 2000 particles, and the simulation is replicated 200 times independently. The trust region optimization algorithm (Nocedal and Wright 1999) is used for calculating MLE weights in all examples. We report the average of MSE over 100 steps, that is,

$$\overline{\text{MSE}}=\frac{1}{100}\sum_{t=1}^{100}\text{MSE}_t,\ \text{where MSE}_t=\frac{1}{200}\sum_{i=1}^{200}(\widehat{\mu}_{ti}-\mu_t)^2,$$

where $\widehat{\mu}_{ti}$ is the estimator at the tth step of tth replication and μ_t is the theoretical posterior mean of x_t, and the comparison between LM-SMC and the tth method with consideration of computing time, that is, the ratio

$$R_i=\frac{\overline{\text{MSE}}_{\text{LM-SMC}}\times T_{\text{LM-SMC}}}{\overline{\text{MSE}}_i\times T_i},$$

where T is the computing time, for $i=1,\ldots,4$ are reported.

Table 1. Comparison of the four estimators in example of Section 3.4. Simulation is replicated for 1000 times independently and each replicate uses 4000 draws. The mean square errors (MSEs) are reported

	$\widehat{\mu}_{P_2}$	$\widehat{\mu}_{P_2C_1}$	$\widehat{\mu}_{P_{42}C_{41}}$	$\widehat{\mu}_{P_2C_{41}}$
MSE	$1.4E-1$	$1.4E-1$	$4.1E-3$	$5.0E-3$

4.1 AR(1) Observed With Noise

Consider the following process

$$x_n=\phi x_{n-1}+\varepsilon_n,\ \varepsilon_n\sim N(0,\sigma^2)$$
$$y_n=x_n+e_n,\ e_n\sim N(0,1).$$

It is often used as a benchmark model for comparing SMC methods since all sequential distributions are analytically available through Kalman filter.

For APF, since the density $p(x_n|x_{n-1},y_n)$ is normal and $\eta(x_{n-1})=\pi_n(x_{0:n-1})/\pi_{n-1}(x_{0:n-1})$ can be evaluated, in this example we use them as the proposal density and the auxiliary function. This is referred to as "perfect adaption" of the APF in Pitt and Shephard (1999).

4.1.1 Construction of Component Proposals. For the mixture methods, the component proposals can be obtained by the following decomposition. Denote the mean of x_n with density $p(x_n|x_{n-1},y_n)$ conditional on x_{n-1} by $\theta_n(x_{n-1})$. We have

$$p(x_n|x_{n-1})p(y_n|x_n)(x_n-\mu_n)$$
$$=\eta(x_{n-1})\left[(x_n-\theta_n(x_{n-1}))^+p(x_n|x_{n-1},y_n)\right.$$
$$+(x_n-\theta_n(x_{n-1}))^-p(x_n|x_{n-1},y_n)$$
$$\left.+(\theta_n(x_{n-1})-\mu_n)p(x_n|x_{n-1},y_n)\right]. \quad (17)$$

Then, we can use the same $\eta(x_{n-1})$ in APF as the auxiliary function, and the following component proposals: $p(x_n|x_{n-1},y_n)$, normalized $(x_n-\theta_n(x_{n-1}))^+p(x_n|x_{n-1},y_n)$ and normalized $(x_n-\theta_n(x_{n-1}))^-p(x_n|x_{n-1},y_n)$. Since $p(x_n|x_{n-1},y_n)$ is known to be the optimal choice in the sense of minimizing the conditional variance of the importance weights, following the

Table 2. Comparison of five methods in Example 1. $\overline{\text{MSE}}$ is reported and the ratio of $\overline{\text{MSE}}$ multiplied with computing time between LM-SMC and the corresponding method is reported in the parenthesis

SNR = 10	$\phi = 0.7, \sigma = 2.3$	$\phi = 0.8, \sigma = 1.9$	$\phi = 0.9, \sigma = 1.4$	Time s
BF	1.7E–3(0.0004)	8.5E–4(0.0022)	8.0E–4(0.015)	44
APF	4.3E–4(0.0015)	4.0E–4(0.0044)	3.7E–4(0.033)	46
GSMC	3.9E–4(0.0013)	3.9E–4(0.0037)	3.5E–4(0.027)	57
RM-SMC	5.5E–6(0.069)	1.2E–5(0.088)	3.5E–5(0.2)	78
LM-SMC	3.1E–7(1.0)	8.5E–7(1.0)	5.7E–6(1.0)	97
SNR = 1	$\phi = 0.7, \sigma = 0.7$	$\phi = 0.8, \sigma = 0.6$	$\phi = 0.9, \sigma = 0.44$	Time s
BF	3.3E–4(0.042)	3.3E–4(0.11)	2.9E–4(0.30)	44
APF	2.1E–4(0.065)	2.0E–4(0.18)	1.9E–4(0.49)	46
GSMC	2.0E–4(0.054)	1.9E–4(0.15)	1.8E–4(0.42)	57
RM-SMC	4.4E–5(0.18)	6.4E–5(0.32)	1.0E–4(0.52)	78
LM-SMC	6.7E–6(1.0)	1.7E–5(1.0)	4.6E–5(1.0)	97
SNR = 0.5	$\phi = 0.7, \sigma = 0.5$	$\phi = 0.8, \sigma = 0.42$	$\phi = 0.9, \sigma = 0.31$	Time s
BF	2.2E–4(0.08)	2.0E–4(0.19)	1.5E–4(0.44)	44
APF	1.4E–4(0.11)	1.4E–4(0.27)	1.2E–4(0.51)	46
GSMC	1.4E–4(0.09)	1.3E–4(0.22)	1.2E–4(0.44)	57
RM-SMC	4.3E–5(0.21)	6.2E–5(0.35)	7.9E–5(0.47)	78
LM-SMC	7.5E–6(1.0)	1.8E–5(1.0)	3.0E–5(1.0)	97

guidelines in Section 3.4, the mixture proportions of the above proposals are set to be 0.98, 0.01, 0.01.

4.1.2 Results. The five methods are compared under different settings of signal to noise ratio (SNR), defined as the ratio between the variances of x_n and e_n. The values of parameter ϕ are set to be 0.7, 0.8, 0.9, and σ are determined so that the SNR of the process is controlled at 10, 1, and 0.5. For each setting, a series with length 100 is generated as observations. Results are listed in Table 2, and the trajectories of logarithm of MSE for all five methods are given in Figure 1.

In Figure 1, it is seen that the MSE of LM-SMC is always smaller than the others, which is consistent with the theoretical comparisons in Section 3.2. In Table 2, it is seen that in all parameter settings, with both variance and computing time considered, LM-SMC outperforms the other methods significantly, ranging from 60% to several orders of magnitude. It is not surprising since each term of the linear decomposition (17) is used directly as the control variate, and the target function is very close to a linear combination of the control variates.

From Table 2, it is seen that the improvement of LM-SMC is more significant for higher SNR or smaller ϕ. This phenomena gives some insights of when to use LM-SMC. First, higher SNR indicates higher signal strength. Since the control variates in LM-SMC involve the signal level x_n, the higher SNR is, the more information about x_n are contained in the control variates, which gives LM-SMC more advantages to estimate x_n. In fact, LM-SMC always achieves smaller MSEs with a higher SNR than a lower one. This is not necessarily true for BF, APF, and GSMC. Therefore, the improvement LM-SMC can achieve is related to the signal strength. Second, smaller ϕ means larger uncertainty in x_n. BF and APF heavily rely on the quality of x_n and therefore deteriorates for smaller ϕ. In contrast, LM-SMC becomes increasingly more effective. Therefore, our method gives improvement in situations where the existing methods perform poorly.

4.1.3 Comparing Expansion Terms of Variances. Compared with BF, the most basic method, the improvement of LM-SMC includes the effects of the auxiliary function, better proposal densities, and control variates. To illustrate the sole effect of control variates by separating it from the others, here we provide some comparisons among the terms of theoretical variance expansions (3) and (10) of all five methods. Due to the linear Gaussian nature of this example, each term of the variance expansions can be estimated accurately. Consider the parameter setting $\phi = 0.9$ and $\sigma = 1.4$ in Table 2 and the corresponding observation series that has length 100, and the same design for each method as previously. At each time step n, all n terms of the variance expansion of each method are estimated using 10^6 Monte Carlo samples. The results on the last three terms are reported in Table 3. The other terms follow a similar pattern. Note that different from previous simulations using systematic resampling, the theoretical variances here are for algorithms using multinomial resampling.

In Table 3, the comparisons between APF and BF, GSMC and APF, RM-SMC and GSMC, and LM-SMC and RM-SMC reveal the improvements due to the auxiliary function, the multiple proposals, the control variates of the last expansion term, and the control variates of earlier terms, respectively. First, it can be seen that most of the improvement comes from control variates, by noting that in terms of total variances, LM-SMC improves upon GSMC by over 89% for all n, while GSMC improves upon BF by less than 25% for $n = 3$ and 7 and 75% for $n = 9$ due to the failure of BF. Second, the variance reduction contributed by the second-to-last term is also significant, which makes the variances of LM-SMC is almost 50% of those of RM-SMC. Therefore, LM-SMC has significant improvement on efficiency by achieving variance reduction beyond the last term of the variance expansion.

Similar patterns are seen in all other time step n (not shown here). We also observe that, for each method, the relative

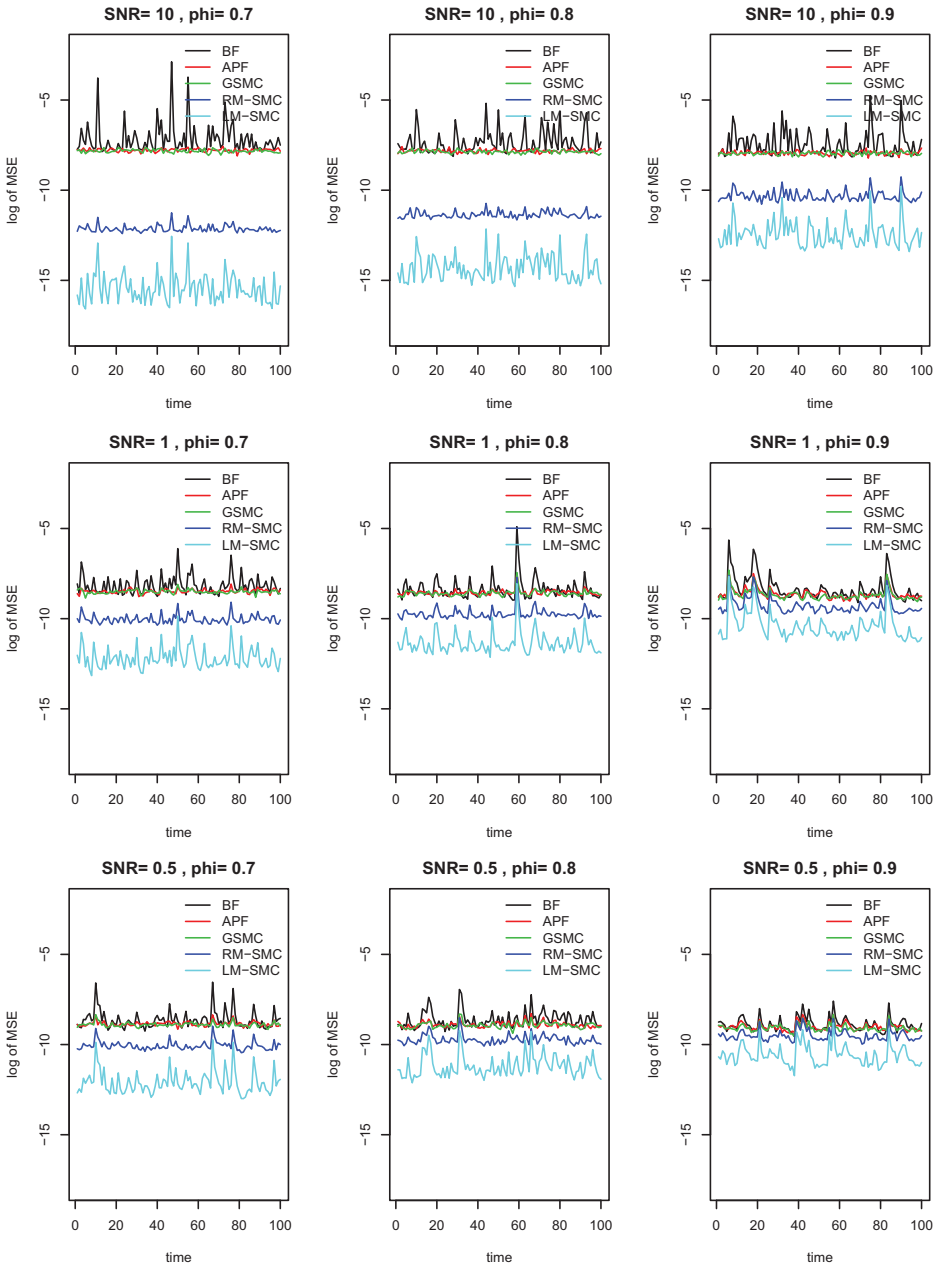

Figure 1. Trajectories of logarithm of MSE for the five estimators in all cases of Example 1.

contribution of the kth term toward the total variance decreases quickly as k decreases. Hence, the earlier terms are almost negligible, and the last three steps are the dominating terms.

4.2 Stochastic Volatility With AR(1) Dynamics

Consider the stochastic volatility model in Sandmann and Koopman (1998)

$$\begin{cases} x_n &= \phi x_{n-1} + \eta_n,\ \eta_n \sim N(0, \sigma_\eta^2) \\ y_n &= \overline{\sigma} e^{\frac{x_n}{2}} \xi_n,\ \xi_n \sim N(0, 1), \end{cases}$$

where η_n and ξ_n are independent, y_n is the demeaned return of a portfolio obtained by subtracting the average of all returns from the actual return, and $\overline{\sigma}$ is the average volatility level. Due to the nonlinear structure of the observation equation, the analytical form of $p(x_n|x_{n-1}, y_n)$ is unavailable.

For the APF method, Kim, Shephard, and Chib (1998) and Pitt and Shephard (1999) suggested to use the normal density $q_1(x_n|x_{n-1}, y_n) \propto p(x_n|x_{n-1})p_1(y_n|x_n)$ as the proposal and the corresponding normalizing constant as the auxiliary function, where $\log(p_1(y_n|x_n))$ is the first-order Taylor expansion of $\log(p(y_n|x_n))$ around $\phi \sum_{j=1}^{N} x_{n-1}^{(j)}/N$. Since $p(y_n|x_n)$ is log-concave, $p_1(y_n|x_n)$ has heavier tails than $p(y_n|x_n)$, and hence $q_1(x_n|x_{n-1}, y_n)$ and the auxiliary function satisfy the tail requirement of the proposal distribution. In this example, we select $q_1(x_n|x_{n-1}, y_n)$ and $p(x_n|x_{n-1})p_1(y_n|x_n)/q_1(x_n|x_{n-1}, y_n)$ as the proposal and the auxiliary function for the APF method.

4.2.1 Construction of Component Proposals. For the mixture methods, let $q_2(x_n|x_{n-1}, y_n)$ being the normalized $p(x_n|x_{n-1})p_2(y_n|x_n)$ where $\log(p_2(y_n|x_n))$ is the second-order Taylor expansion of $p(y_n|x_n)$ around the

Table 3. Comparison of variance expansion terms of five methods in Example 1. Parameter settings are $\phi = 0.9$ and $\sigma = 1.4$. The asymptotic variances at time n, and the last three terms, denoted by kth, of each variance expansion are reported

$n = 3$	first	second	third	Total variance
BF	3.4e-3	7.1e-2	9.1e-1	9.8e-1
APF	4.0e-3	6.1e-2	7.1e-1	7.8e-1
GSMC	3.8e-3	5.9e-2	6.8e-1	7.4e-1
RM-SMC	3.8e-3	5.9e-2	5.7e-2	1.2e-1
LM-SMC	4.3e-4	6.6e-3	5.7e-2	6.4e-2
$n = 7$	fifth	sixth	seventh	Total variance
BF	3.5e-3	2.0e-1	6.8e-1	8.8e-1
APF	6.5e-3	5.3e-2	7.2e-1	7.7e-1
GSMC	6.3e-3	5.1e-2	6.8e-1	7.4e-1
RM-SMC	6.3e-3	5.1e-2	5.7e-2	1.1e-1
LM-SMC	1.4e-3	5.8e-3	5.7e-2	6.4e-2
$n = 9$	seventh	eighth	ninth	Total variance
BF	3.4e-3	2.0e-1	3	3.2
APF	6.6e-3	1.1e-1	7.1e-1	8.3e-1
GSMC	6.3e-3	1.0e-1	6.8e-1	7.9e-1
RM-SMC	6.3e-3	1.0e-1	5.7e-2	1.6e-1
LM-SMC	1.4e-3	2.6e-2	5.7e-2	8.4e-2

maximum point of the observation likelihood, $r(x_{n-1}) \equiv p(x_n|x_{n-1})p_2(y_n|x_n)/q_2(x_n|x_{n-1}, y_n)$ and denote the mean of $q_2(x_n|x_{n-1}, y_n)$ by $\theta(x_{n-1})$. Smith and Santos (2006) discussed the benefit of using $q_2(x_n|x_{n-1}, y_n)$ as the proposal when there are extreme outliers in the observations. One problem of $q_2(x_n|x_{n-1}, y_n)$ is that it does not have a tail heavier than $p(x_n|x_{n-1})p_1(y_n|x_n)$.

By taking logarithm on both sides of the observation equation, the stochastic volatility model is a special case of model (14) in Section 3.3. Therefore, its construction of component proposals can be applied here. It gives an auxiliary function $r(x_{n-1}) + c_{n-1}$ where $c_{n-1} = E_{\pi^*_{n-2}q_\alpha}[r(x_{n-1})]$ which can be estimated by $\sum_{j=1}^N r(x_{n-1}^{(j)})/N$, and the component proposals $q_2(x_n|x_{n-1}, y_n)$, $p(x_n|x_{n-1})$, normalized $(x_n - \theta(x_{n-1}))^+ q_2(x_n|x_{n-1}, y_n)$, normalized $(x_n - \theta(x_{n-1}))^- q_2(x_n|x_{n-1}, y_n)$, normalized $(x_n - \phi x_{n-1})^+ p(x_n|x_{n-1})$, and normalized $(x_n - \phi x_{n-1})^- p(x_n|x_{n-1})$. The tail requirement can be satisfied by letting the mixture proportion of $p(x_n|x_{n-1})$ larger than 0.

Although all component proposals can be sampled directly, for the purpose of illustration, we follow the discussion in Section 3.4 and set the mixture proportions of $q_2(x_n|x_{n-1}, y_n)$ and $p(x_n|x_{n-1})$ to be 0.5 for each, and the other proposals to be 0. With such mixture proportions, for different target functions, the sampling procedure remains the same and only the control variates need to be changed.

4.2.2 Results. Here, we use the parameter settings in Sandmann and Koopman (1998). The values of the autoregressive parameter ϕ are set to be 0.90, 0.95, and 0.98, which are compatible with the range from 0.9 to 0.995 of ϕ found in empirical studies. Then for each ϕ, the values of σ_η are selected so that the coefficient of variation of the volatility $h = \overline{\sigma}^2 \exp(x_n)$,

Table 4. Comparison of five methods in Example 2. $\overline{\text{MSE}}$ is reported. As a measure of computationally efficiency, the ratio of $\overline{\text{MSE}}$ multiplied with computing time between LM-SMC and the corresponding method is reported in the parenthesis. The theoretical posterior mean is calculated using Monte Carlo sample

CV = 10	$\phi = 0.90$, $\sigma_\eta = 0.675, \overline{\sigma} = 0.0165$	$\phi = 0.95$, $\sigma_\eta = 0.484, \overline{\sigma} = 0.0164$	$\phi = 0.98$, $\sigma_\eta = 0.308, \overline{\sigma} = 0.0166$	Time s
BF	8.0E–4(0.54)	4.3E–4(0.78)	2.2E–4(1.4)	53
APF	2.4E–2(0.012)	9.8E–3(0.023)	2.7E–4(0.80)	78
GSMC	5.1E–4(0.56)	3.6E–4(0.61)	2.6E–4(0.80)	81
RM-SMC	3.2E–4(0.58)	2.6E–4(0.54)	2.0E–4(0.67)	126
LM-SMC	1.2E–4(1.0)	9.2E–5(1.0)	8.7E–5(1.0)	193
CV = 1	$\phi = 0.90$, $\sigma_\eta = 0.363, \overline{\sigma} = 0.0252$	$\phi = 0.95$, $\sigma_\eta = 0.260, \overline{\sigma} = 0.0252$	$\phi = 0.98$, $\sigma_\eta = 0.166, \overline{\sigma} = 0.0253$	Time s
BF	2.7E–4(0.61)	2.2E–4(0.78)	2.4E–4(1.7)	53
APF	4.0E–4(0.28)	2.5E–4(0.45)	4.6E–4(0.60)	78
GSMC	2.3E–4(0.46)	1.9E–4(0.57)	2.2E–4(1.2)	81
RM-SMC	1.7E–4(0.41)	1.6E–4(0.44)	2.1E–4(0.82)	126
LM-SMC	4.5E–5(1.0)	4.6E–5(1.0)	1.1E–4(1.0)	193
CV = 0.1	$\phi = 0.90$, $\sigma_\eta = 0.135, \overline{\sigma} = 0.0293$	$\phi = 0.95$, $\sigma_\eta = 0.096, \overline{\sigma} = 0.0293$	$\phi = 0.98$, $\sigma_\eta = 0.061, \overline{\sigma} = 0.0295$	Time s
BF	4.3E–5(0.36)	5.2E–5(1.4)	3.4E–5(1.7)	53
APF	3.9E–5(0.27)	4.5E–5(1.1)	3.2E–5(1.2)	78
GSMC	4.3E–5(0.23)	5.2E–5(0.93)	3.8E–5(0.95)	81
RM-SMC	3.2E–5(0.20)	5.0E–5(0.63)	3.3E–5(0.71)	126
LM-SMC	4.2E–6(1.0)	2.0E–5(1.0)	1.5E–5(1.0)	193

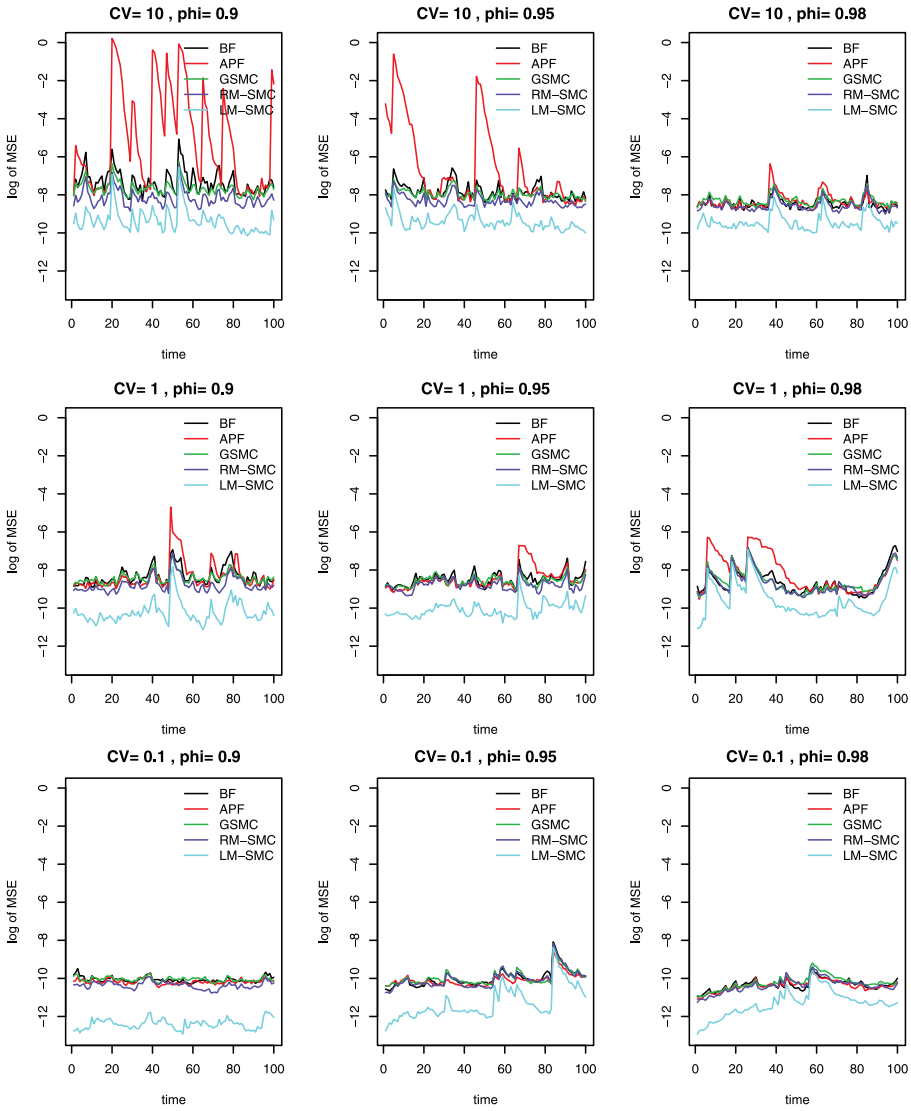

Figure 2. Trajectories of logarithm of MSE for the five estimators in all cases of Example 2.

defined as

$$\mathrm{CV} = \frac{\mathrm{var}[h]}{E[h]^2} = \exp\left(\frac{\sigma_\eta^2}{1-\phi^2}\right) - 1,$$

takes the values 10, 1, and 0.1. High value of CV indicates the relative strength of the volatility process, and low value of CV indicates the volatility is close to a constant. Finally, the average volatility level $\overline{\sigma}$ is selected such that

$$E[h] = \overline{\sigma}^2 \exp\left(\frac{\sigma_\eta^2}{2(1-\phi^2)}\right)$$

is equal to 0.0009. This value of $E[h]$ can be interpreted as an approximately 22% annualized variance if the simulated data are taken as weekly returns. For each setting, a series with length 100 is generated as observations. Results are listed in Table 4, and the trajectories of logarithm of MSE for all five methods are given in Figure 2.

In Figure 2, similar to Figure 1, the MSE of LM-SMC is always the smallest among all methods studied. In Table 4, the performance of LM-SMC also has a similar pattern as in Table 2. The larger ϕ is, that is, the more volatile the volatility process is, the greater improvement our new method achieves. For $\phi = 0.9$, the improvement of LM-SMC over the other methods ranges from 39% to 80%, not including the failed APF. For $\phi = 0.98$, compared with BF, although the improvement of LM-SMC on MSE is still over 50%, it does not perform better due to the extra computing time. In this setting, with $\phi = 0.98$, the volatility process is close to a random walk with small variance and is quite persistent. Therefore, the standard method already achieves good accuracy, and there is little room for improvement, in which case the new method may not be needed.

In the results, the mixture methods performs more robustly than APF. When CV = 10 and CV = 1, the APF performs worse than BF, and in two cases of CV = 10, its performances are extremely bad. A possible reason is that, although $p(y_n|x_n)/p_1(y_n|x_n)$ is bounded, the upper bound may still be large and the sample weights are skewed. In comparison, all three mixture methods are more stable. Therefore,

the protection provided by the mixture proposal outperforms the one provided by expanding the log-concave density.

Finally, it can be seen that LM-SMC reduces the MSE of RM-SMC from 48% to 87%. This means that in (10), the variance reduction brought by adding the control variates to the historical terms is significant.

5. SUMMARY

In this article, we propose a new SMC algorithm by using Tan's likelihood approach (Tan 2004) with both resampling and estimation and give practical guidelines of selecting control variates and sampling proposals which are critical for efficient implementation of the new algorithm. Compared to the direct use of multiple proposals and control variates, the new algorithm always has smaller asymptotic variance, which is proved in the established theoretical framework. The numerical studies show that, by including the information of target function and introducing heavy tailed proposal for protection, the new algorithm can be more efficient and stable than the BF and APF.

APPENDIX A: PROOF OF RESULTS

The new algorithm differs from the basic SMC method in the use of three elements: The proposal density which is the mixture proposal $q_{\boldsymbol{\alpha}}$, the addition of the auxiliary function $\eta(x_{0:n})$ in resampling, and the modified importance weights $v_n^{(j)}$. The extension of Douc and Moulines' (2008) proof to include mixture proposal $q_{\boldsymbol{\alpha}}$ is natural, and the extension to include auxiliary function can be referred to Douc, Moulines, and Olsson (2009). Therefore, the following proof is focused on including the new importance weights $v_n^{(j)}$ in the CLT. The theorem is proved by inductions. At time $t-1$, assume conditions (C1), (C2′), (C3)–(C6) are satisfied, A'_{t-1} and W'_{t-1} are proper, and the following consistency and the CLT hold:

$$\frac{1}{N}\sum_{j=1}^{N} h\left(\widetilde{x}_{1:t-1}^{(j)}\right) \xrightarrow{P} \widetilde{\mu}_{t-1}, \forall h \in L^1(\widetilde{\pi}_{t-1}^*), \tag{A.1}$$

$$\sqrt{N}\left(\frac{1}{N}\sum_{j=1}^{N} h\left(\widetilde{x}_{1:t-1}^{(j)}\right) - \widetilde{\mu}_{t-1}\right) \xrightarrow{P} \mathcal{L}N(0, \sigma_{2,t-1}^2(h)), \forall h \in A'_{t-1}, \tag{A.2}$$

where $\sigma_{2,t-1}^2(ch) = c\sigma_{2,t-1}^2(h)$ for all scalar c. Define $\sigma_{0,t}(h) = \sigma_{2,t-1}(E_{q_{\boldsymbol{\alpha}}}[h|x_{0:t-1}]) + E_{\pi_{n-1}^*}(\mathrm{var}_{q_{\boldsymbol{\alpha}}}[h|x_{0:t-1}])$. The following lemma gives convergence results for the weighted sample $\{(x_{1:t}^{(j)}, N^{-1})\}_{j=1}^N$.

Lemma A.1 (Propagation). Under the inductive hypotheses, the following convergence results hold:

$$\frac{1}{N}\sum_{j=1}^{N} h\left(x_{1:t}^{(j)}\right) \xrightarrow{P} \int h(x_{1:t})\widetilde{\pi}_{t-1}^*(x_{0:t-1})q_{\boldsymbol{\alpha}}(x_t|x_{0:t-1})dx_{0:t},$$

$$\forall h \in L^1(\pi_{t-1}^* q_{\boldsymbol{\alpha}}),$$

$$\sqrt{N}\left[\frac{1}{N}\sum_{j=1}^{N} h\left(x_{1:t}^{(j)}\right) - \int h(x_{1:t})\widetilde{\pi}_{t-1}^*(x_{0:t-1})q_{\boldsymbol{\alpha}}(x_t|x_{0:t-1})dx_{0:t}\right]$$

$$\xrightarrow{\mathcal{L}} N(0, \sigma_{0,t}^2(f)), \forall h \in \widetilde{A}'_{t-1},$$

where $\widetilde{A}'_{t-1} = \{h : \Theta_t \to \mathbb{R} | E_{q_{\boldsymbol{\alpha}}}[|h|] \in A'_{t-1}, E_{q_{\boldsymbol{\alpha}}}[h^2] \in W'_{t-1}\}$.

Proof. By applying Douc and Moulines (2008, Theorem 1 and 2) on $\{(x_{1:t}^{(j)}, N^{-1})\}_{j=1}^N$, Lemma 1 holds. □

In the next lemma, an expansion of the MLE weights $\widehat{\boldsymbol{\zeta}}_t$ is given.

Lemma A.2. Under the inductive hypotheses, it holds that

$$\widehat{\boldsymbol{\zeta}}_t = \mathrm{var}_{\widetilde{\pi}_{t-1}^* q_{\boldsymbol{\alpha}}}\left[\frac{\boldsymbol{g}}{q_\alpha}\right]^{-1} \frac{1}{N}\sum_{j=1}^{N} \frac{\boldsymbol{g}\left(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)}\right)}{q_\alpha\left(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)}\right)} + o_p\left(N^{-1/2}\right). \tag{A.3}$$

Then, $\widehat{\boldsymbol{\zeta}}_t \xrightarrow{P} 0$ and $\widehat{\boldsymbol{\zeta}}_t = O_p\left(N^{-1/2}\right)$.

Proof. By definition, $\sqrt{N}\widehat{\boldsymbol{\zeta}}_t$ is the maximizer of the concave function $\psi(\boldsymbol{s}) = \sum_{j=1}^N \log[q_{\boldsymbol{\alpha}}(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)}) + N^{-1/2}\boldsymbol{s}^T\boldsymbol{g}(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)})]$ the concavity of which can be seen by its Hessian matrix. The standard expansion theory for M-estimation with a convex criterion function cannot apply due to the dependence of $\{x_{0:t}^{(j)}\}$. Here, the arguments follow Hjort and Pollard (2011, basic corollary) and Li, Tan, and Chen (2013, lemma 4). By Hjort and Pollard (2011), if $\psi(\boldsymbol{s})$ can be represented as $\frac{1}{2}\boldsymbol{s}^T V_N \boldsymbol{s} + U_N^T\boldsymbol{s} + C_N + r_N(\boldsymbol{s})$ with $V_N \xrightarrow{P} V$, $-V$ is symmetric and positive definite, U_N is stochastic bounded, and $r_N(\boldsymbol{s}) = o_p(1)$ for each $\boldsymbol{s}$, then $\sqrt{N}\widehat{\boldsymbol{\zeta}}_t = -V^{-1}U_N + o_p(1)$. By Taylor expansion of $\psi(\boldsymbol{s})$ around 0, it can be represented in the above form with

$$V_N = -\frac{1}{N}\sum_{j=1}^{N} \frac{\boldsymbol{g}\left(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)}\right)\boldsymbol{g}^T\left(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)}\right)}{q_{\boldsymbol{\alpha}}^2\left(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)}\right)},$$

$$U_N = \frac{1}{\sqrt{N}}\sum_{j=1}^{N} \frac{\boldsymbol{g}\left(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)}\right)}{q_{\boldsymbol{\alpha}}\left(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)}\right)},$$

and

$$r_N(\boldsymbol{s}) = \frac{1}{N\sqrt{N}}\sum_{j=1}^{N}\sum_{i_1,i_2,i_3=1}^{d} \frac{g_{i_1}g_{i_2}g_{i_3}s_{i_1}s_{i_2}s_{i_3}}{q_{\boldsymbol{\xi}_{i_1,i_2,i_3}}^3\left(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)}\right)},$$

where g_i and s_i are coordinates of $\boldsymbol{g}(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)})$ and $\boldsymbol{s}$, respectively, and $\boldsymbol{\xi}_{i_1,i_2,i_3} - \boldsymbol{\alpha} = O(N^{-1/2})$. Then, with (C3), (C4), and Lemma 1, $V_N \xrightarrow{P} -\mathrm{var}_{\widetilde{\pi}_{t-1}^* q_{\boldsymbol{\alpha}}}[\boldsymbol{g}/q_{\boldsymbol{\alpha}}]$ and $r_N(\boldsymbol{s}) = o_p(1)$. Since $\int |g_i|dx_t \le 2$ and A'_{t-1} is proper, $E_{q_{\boldsymbol{\alpha}}}[|g_i|/q_{\boldsymbol{\alpha}}] \in A'_{t-1}$. Then with (C5) and Lemma 1, U_N is asymptotic normal. Therefore, the expansion (A.3) holds. Note that by (C6) and Lemma 1, U_N is consistent to 0, which concludes the proof. □

With the expansion in Lemma 2, the consistency for the weighting step and resampling step holds in the following lemma.

Lemma A.3. Under the inductive hypotheses, it holds that

$$\frac{\sum_{j=1}^N h\left(x_{1:t}^{(j)}\right)v_t^{(j)}}{\sum_{j=1}^N v_t^{(j)}} \xrightarrow{P} \mu_t, \forall h \in L^1(\pi_t^*), \tag{A.4}$$

$$\frac{1}{N}\sum_{j=1}^{N} h\left(\widetilde{x}_{1:t}^{(j)}\right) \xrightarrow{P} \widetilde{\mu}_t, \forall h \in L^1(\widetilde{\pi}_t^*). \tag{A.5}$$

Proof. The Taylor expansion of $N^{-1}\sum_{j=1}^N h(x_{1:t}^{(j)})v_t^{(j)}$ for $\widehat{\boldsymbol{\zeta}}_t$ around 0 gives that

$$\begin{aligned}
&\frac{1}{N}\sum_{j=1}^N h\left(x_{1:t}^{(j)}\right)v_t^{(j)}\\
&=\frac{1}{N}\sum_{j=1}^N \frac{h\left(x_{1:t}^{(j)}\right)\pi_t\left(x_{0:t}^{(j)}\right)}{\widetilde{\pi}_{t-1}\left(\widetilde{x}_{0:t-1}^{(j)}\right)\left[q_{\boldsymbol{\alpha}}\left(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)}\right)+\widehat{\boldsymbol{\zeta}}_t^T\boldsymbol{g}\left(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)}\right)\right]}\\
&=\frac{1}{N}\sum_{j=1}^N \frac{h\left(x_{1:t}^{(j)}\right)\pi_t\left(x_{0:t}^{(j)}\right)}{\widetilde{\pi}_{t-1}\left(\widetilde{x}_{0:t-1}^{(j)}\right)q_{\boldsymbol{\alpha}}\left(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)}\right)}\\
&\quad-\frac{1}{N}\sum_{j=1}^N \frac{h\left(x_{1:t}^{(j)}\right)\pi_t\left(x_{0:t}^{(j)}\right)\boldsymbol{g}\left(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)}\right)^T}{\widetilde{\pi}_{t-1}\left(\widetilde{x}_{0:t-1}^{(j)}\right)q_{\boldsymbol{\alpha}}\left(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)}\right)^2}\widehat{\boldsymbol{\zeta}}_t+o_p\left(\frac{1}{\sqrt{N}}\right).
\end{aligned}$$

Plugging in the expansion of $\widehat{\boldsymbol{\zeta}}_t$ of Lemma 2 gives that

$$\begin{aligned}
\frac{1}{N}\sum_{j=1}^N h\left(x_{1:t}^{(j)}\right)v_t^{(j)} &= \frac{1}{N}\sum_{j=1}^N \frac{h\left(x_{1:t}^{(j)}\right)\pi_t\left(x_{0:t}^{(j)}\right)}{\widetilde{\pi}_{t-1}\left(\widetilde{x}_{0:t-1}^{(j)}\right)q_{\boldsymbol{\alpha}}\left(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)}\right)}\\
&\quad-\frac{1}{N}\sum_{j=1}^N \frac{h\left(x_{1:t}^{(j)}\right)\pi_t\left(x_{0:t}^{(j)}\right)\boldsymbol{g}\left(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)}\right)^T}{\widetilde{\pi}_{t-1}\left(\widetilde{x}_{0:t-1}^{(j)}\right)q_{\boldsymbol{\alpha}}\left(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)}\right)^2}\\
&\quad\times\mathrm{var}\left[\frac{\boldsymbol{g}}{q_{\boldsymbol{\alpha}}}\right]^{-1}\frac{1}{N}\sum_{j=1}^N\frac{\boldsymbol{g}}{q_{\boldsymbol{\alpha}}}+o_p\left(\frac{1}{\sqrt{N}}\right)\\
&= \frac{1}{N}\sum_{j=1}^N \frac{h\left(x_{1:t}^{(j)}\right)\pi_t\left(x_{0:t}^{(j)}\right)}{\widetilde{\pi}_{t-1}\left(\widetilde{x}_{0:t-1}^{(j)}\right)q_{\boldsymbol{\alpha}}\left(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)}\right)}\\
&\quad-\boldsymbol{\tau}_{1t}^T\frac{1}{N}\sum_{i=1}^N\frac{\boldsymbol{g}\left(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)}\right)}{q_{\boldsymbol{\alpha}}\left(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)}\right)}+o_p\left(\frac{1}{\sqrt{N}}\right),
\end{aligned}\tag{A.6}$$

where

$$\boldsymbol{\tau}_{1t}=\mathrm{var}\left[\frac{\boldsymbol{g}}{q_{\boldsymbol{\alpha}}}\right]^{-1}\mathrm{cov}\left[\frac{h\pi_t}{\widetilde{\pi}_{t-1}q_{\boldsymbol{\alpha}}},\frac{\boldsymbol{g}}{q_{\boldsymbol{\alpha}}}\right],$$

and the second equation holds by Lemma 1. Then by Lemma 1 again, from (A.6) we have

$$\frac{1}{N}\sum_{j=1}^N h\left(x_{1:t}^{(j)}\right)v_t^{(j)}\xrightarrow{P}e_{t-1}^{-1}\int h(x_{1:t})\pi_t(x_{0:t})dx_{0:t}.$$

Similarly,

$$\frac{1}{N}\sum_{j=1}^N v_t^{(j)}\xrightarrow{P}e_{t-1}^{-1}c_t.\tag{A.7}$$

Therefore, (A.4) holds by Slutsky's theorem.

For (A.5), consider the weighted sample $\{(x_{0:t}^{(j)},\widetilde{v}_t^{(j)})\}_{j=1}^N$. Following the similar line as above, it holds that

$$\frac{\sum_{j=1}^N h\left(x_{1:t}^{(j)}\right)\widetilde{v}_t^{(j)}}{\sum_{j=1}^N\widetilde{v}_t^{(j)}}\xrightarrow{P}\widetilde{\mu}_t,\forall h\in L^1\left(\widetilde{\pi}_t^*\right).\tag{A.8}$$

Then with Douc and Moulines (2008, Theorem 3), (A.5) holds. □

Then, the CLT of the weighting and resampling steps can be established.

Proof of Theorem 2. Equation (8) has been given by Lemma 3. $\forall h\in A_n'$, for (9), it can be shown by applying Slutsky's theorem to the weak convergence of $N^{-1/2}\sum_{j=1}^N(h(x_{1:t}^{(j)})-\mu_t)v_t^{(j)}$, which has the expansion

$$e_{t-1}^{-1}c_t\frac{1}{\sqrt{N}}\sum_{j=1}^N\left(h_t(x_{0:t}^{(j)})-\boldsymbol{\beta}_t^T\boldsymbol{g}_n\left(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)}\right)\right)+o_p(1)\tag{A.9}$$

by the procedure similar to (A.6) and (A.7). Let $U_N^{(j)}=N^{-1/2}(h_t(x_{0:t}^{(j)})-\boldsymbol{\beta}_t^T\boldsymbol{g}_t(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)}))$. Equation (9) holds if $\sum_{j=1}^N U_N^{(j)}\xrightarrow{\mathcal{L}}N(0,\sigma_{1,t}^2(h))$. Write $\sum_{j=1}^N U_N^{(j)}$ as A_N+B_N, where

$$A_N=\sum_{j=1}^N E_{q_{\boldsymbol{\alpha}}}\left[U_N^{(j)}|\widetilde{x}_{0:t-1}^{(j)}\right],\ B_N=\sum_{j=1}^N\left(U_N^{(j)}-E_{q_{\boldsymbol{\alpha}}}\left[U_N^{(j)}|\widetilde{x}_{0:t-1}^{(j)}\right]\right).$$

For A_N, by condition (C6), $A_N=N^{-1/2}\sum_{j=1}^N E_{q_{\boldsymbol{\alpha}}}[h_t(x_{0:t}^{(j)})|\widetilde{x}_{0:t-1}^{(j)}]$. Condition (C2′) implies that $\mu_t\in A_n'$, and so $h-\mu_t\in A_n'$. By the definition of A_n', $E_{q_{\boldsymbol{\alpha}}}[h_t(x_{0:t})|x_{0:t-1}]\in A_{n-1}'$. Then, since $\int E_{q_{\boldsymbol{\alpha}}}[h_t(x_{0:t})|x_{0:t-1}]\widetilde{\pi}_{t-1}^*(x_{0:t-1})dx_{0:t-1}=0$, by the assumption (A.2), $A_N\xrightarrow{\mathcal{L}}N(0,\sigma_{2,t-1}^2(E_{q_{\boldsymbol{\alpha}}}[h_t(x_{0:t})]))$. For B_N, we prove that

$$\begin{aligned}
&E_{q_{\boldsymbol{\alpha}}}\left[\exp\{iuB_N\}|\{\widetilde{x}_{0:t-1}^{(j)}\}_{j=1}^N\right]\xrightarrow{P}\exp\left\{-(u^2/2)E_{\widetilde{\pi}_{n-1}^*}\left(\mathrm{var}_{q_{\boldsymbol{\alpha}}}[h_t(x_{0:t})\right.\right.\\
&\left.\left.-\boldsymbol{\beta}_t^T\boldsymbol{g}_t(x_t|x_{0:t-1})]\right)\right\}
\end{aligned}\tag{A.10}$$

by applying the Theorem A.3 in Douc and Moulines (2008). It is needed to check its two conditions. Since $h-\mu_t\in A_n'$, by the definition of W_n' and condition (C3), $E_{q_{\boldsymbol{\alpha}}}[(h_t-\boldsymbol{\beta}_t^T\boldsymbol{g}_t)^2]\in L^1(\widetilde{\pi}_{t-1}^*)$. Since $(E_{q_{\boldsymbol{\alpha}}}[h_t-\boldsymbol{\beta}_t^T\boldsymbol{g}_t])^2\le E_{q_{\boldsymbol{\alpha}}}[(h_t-\boldsymbol{\beta}_t^T\boldsymbol{g}_t)^2]$, $(E_{q_{\boldsymbol{\alpha}}}[h_t-\boldsymbol{\beta}_t^T\boldsymbol{g}_t])^2\in L^1(\widetilde{\pi}_{t-1}^*)$. Then by assumption (A.1),

$$\begin{aligned}
&\sum_{j=1}^N E_{q_{\boldsymbol{\alpha}}}\left[U_N^{(j)^2}|\{\widetilde{x}_{0:t-1}^{(j)}\}_{j=1}^N\right]=\frac{1}{N}\sum_{j=1}^N E_{q_{\boldsymbol{\alpha}}}\left[\left(h_t-\boldsymbol{\beta}_t^T\boldsymbol{g}_t\right)^2|\widetilde{x}_{0:t-1}^{(j)}\right]\\
&\quad\xrightarrow{P}E_{\widetilde{\pi}_{t-1}^*}\left(E_{q_{\boldsymbol{\alpha}}}\left[\left(h_t-\boldsymbol{\beta}_t^T\boldsymbol{g}_t\right)^2|\widetilde{x}_{0:t-1}\right]\right),\\
&\sum_{j=1}^N\left(E_{q_{\boldsymbol{\alpha}}}\left[U_N^{(j)}|\{\widetilde{x}_{0:t-1}^{(j)}\}_{j=1}^N\right]\right)^2=\frac{1}{N}\sum_{j=1}^N\left(E_{q_{\boldsymbol{\alpha}}}\left[h_t-\boldsymbol{\beta}_t^T\boldsymbol{g}_t|\widetilde{x}_{0:t-1}^{(j)}\right]\right)^2\\
&\quad\xrightarrow{P}E_{\widetilde{\pi}_{t-1}^*}\left(E_{q_{\boldsymbol{\alpha}}}\left[h_t-\boldsymbol{\beta}_t^T\boldsymbol{g}_t|\widetilde{x}_{0:t-1}\right]\right)^2,\\
&\text{then }\sum_{j=1}^N E_{q_{\boldsymbol{\alpha}}}\left[U_N^{(j)^2}|\{\widetilde{x}_{0:t-1}^{(j)}\}_{j=1}^N\right]-\left(E_{q_{\boldsymbol{\alpha}}}\left[U_N^{(j)}|\{\widetilde{x}_{0:t-1}^{(j)}\}_{j=1}^N\right]\right)^2\\
&\quad\xrightarrow{P}E_{\widetilde{\pi}_{n-1}^*}\left(\mathrm{var}_{q_{\boldsymbol{\alpha}}}\left[h_t-\boldsymbol{\beta}_t^T\boldsymbol{g}_t\right]\right).
\end{aligned}$$

Therefore the first condition holds. For the second condition, $\forall\epsilon>0$, consider a constant $C>0$. It is easy to see that

$$\begin{aligned}
&\sum_{j=1}^N E_{q_{\boldsymbol{\alpha}}}\left[U_N^{(j)^2}\mathbf{1}_{\{|U_N^{(j)}|\ge\epsilon\}}|\{\widetilde{x}_{0:t-1}^{(j)}\}_{j=1}^N\right]\\
&\le\frac{1}{N}\sum_{j=1}^N E_{q_{\boldsymbol{\alpha}}}\left[\left(h_t-\boldsymbol{\beta}_t^T\boldsymbol{g}_t\right)^2\mathbf{1}_{\{|h_t-\boldsymbol{\beta}_t^T\boldsymbol{g}_t|\ge C\}}|\widetilde{x}_{0:t-1}^{(j)}\right]\\
&\quad+\mathbf{1}_{\{C>\sqrt{N}\epsilon\}}\frac{1}{N}\sum_{j=1}^N E_{q_{\boldsymbol{\alpha}}}\left[\left(h_t-\boldsymbol{\beta}_t^T\boldsymbol{g}_t\right)^2|\widetilde{x}_{0:t-1}^{(j)}\right]\\
&\xrightarrow{P}E_{\widetilde{\pi}_{t-1}^*q_{\boldsymbol{\alpha}}}\left[(h_t-\boldsymbol{\beta}_t^T\boldsymbol{g}_t)^2\mathbf{1}_{\{|h_t-\boldsymbol{\beta}_t^T\boldsymbol{g}_t|\ge C\}}\right].
\end{aligned}$$

Since $E_{\widetilde{\pi}_{t-1}^*q_{\boldsymbol{\alpha}}}[(h_t-\boldsymbol{\beta}_t^T\boldsymbol{g}_t)^2]<\infty$ and the above convergence holds for any $C>0$,

$$\sum_{j=1}^N E_{q_{\boldsymbol{\alpha}}}\left[U_N^{(j)^2}\mathbf{1}_{\{|U_N^{(j)}|\ge\epsilon\}}|\left\{\widetilde{x}_{0:t-1}^{(j)}\right\}_{j=1}^N\right]\xrightarrow{P}0,$$

and the second condition holds. Then applying the Theorem A.3 in Douc and Moulines (2008), (A.10) holds. Since A_N is a function of

$\{\widetilde{x}_{0:t-1}^{(j)}\}_{j=1}^N$, it can be proved through the characteristic function that $A_N + B_N \xrightarrow{\mathfrak{L}} N(0, \sigma_{1,t}^2(h))$ and therefore (9) holds.

To finish the proof, the induction hypotheses need to hold at time t. It is easy to show that W_t is proper with proper W_{t-1} and condition (C2′), and A_t is proper with proper A_{t-1} and condition (C2′). By Lemma 2, (A.5) which is the correspondence of (A.1) holds. It is left to show that the correspondence of (A.2),

$$\sqrt{N}\left(\frac{1}{N}\sum_{j=1}^N h\left(\widetilde{x}_{1:t}^{(j)}\right) - \widetilde{\mu}_t\right) \xrightarrow{\mathfrak{L}} N\left(0, \sigma_{2,t}^2(h)\right), \forall h \in A_t', \quad \text{(A.11)}$$

holds. Write the LHS of (A.11) as $\widetilde{A}_N + \widetilde{B}_N$, where

$$\widetilde{A}_N = \frac{1}{\sqrt{N}}\sum_{j=1}^N E\left[h\left(\widetilde{x}_{1:t}^{(j)}\right) - \widetilde{\mu}_t \middle| \left\{x_{0:t}^{(j)}\right\}_{j=1}^N\right],$$

$$\widetilde{B}_N = \frac{1}{\sqrt{N}}\sum_{j=1}^N \left(h\left(\widetilde{x}_{1:t}^{(j)}\right) - E\left[h\left(\widetilde{x}_{1:t}^{(j)}\right) \middle| \left\{x_{0:t}^{(j)}\right\}\right]_{j=1}^N\right).$$

For $\widetilde{B}_N$, we prove that

$$E_{q_{\boldsymbol{\alpha}}}\left[\exp\{iuB_N\} \middle| \left\{x_{0:t}^{(j)}\right\}_{j=1}^N\right] \xrightarrow{P} \exp\left\{-\left(u^2/2\right)\mathrm{var}_{\widetilde{\pi}_t^*}[h]\right\} \quad \text{(A.12)}$$

similarly to the proof of (A.10). Let $\widetilde{U}_N^{(j)} = N^{-1/2}h(\widetilde{x}_{1:t})^{(j)}$. The two conditions of Theorem A.3 in Douc and Moulines (2008) need to be checked. For $h \in A_t', h^2 \in L^1(\widetilde{\pi}_n^*)$ and then $h \in L^1(\widetilde{\pi}_n^*)$. Then by (A.8),

$$\sum_{j=1}^N E\left[\widetilde{U}_N^{(j)2} \middle| \left\{x_{0:t}^{(j)}\right\}_{j=1}^N\right] = \frac{\sum_{j=1}^N h\left(x_{1:t}^{(j)}\right)^2 \widetilde{v}_t^{(j)}}{\sum_{j=1}^N \widetilde{v}_t^{(j)}}$$
$$\xrightarrow{P} \int h(x_{1:t})^2 \widetilde{\pi}_t^*(x_{0:t})dx_{0:t},$$

and

$$\sum_{j=1}^N \left(E\left[\widetilde{U}_N^{(j)} \middle| \left\{x_{0:t}^{(j)}\right\}_{j=1}^N\right]\right)^2 = \left(\frac{\sum_{j=1}^N h\left(x_{1:t}^{(j)}\right)\widetilde{v}_t^{(j)}}{\sum_{j=1}^N \widetilde{v}_t^{(j)}}\right)^2$$
$$\xrightarrow{P} \left(\int h(x_{1:t})\widetilde{\pi}_t^*(x_{0:t})dx_{0:t}\right)^2.$$

Then $\sum_{j=1}^N \{E[\widetilde{U}_N^{(j)2}|\{x_{0:t}^{(j)}\}_{j=1}^N] - (E[\widetilde{U}_N^{(j)}|\{x_{0:t}^{(j)}\}_{j=1}^N])^2\} \xrightarrow{P} Var_{\pi_t^*}[h]$ and the first condition holds. For the second condition, $\forall \epsilon > 0$, consider a constant $C > 0$. It is easy to see that

$$\sum_{j=1}^N E\left[U_N^{(j)2}\mathbf{1}_{\{|U_N^{(j)}|\geq\epsilon\}} \middle| \left\{\widetilde{x}_{0:t}^{(j)}\right\}_{j=1}^N\right] \leq \frac{\sum_{j=1}^N h\left(x_{1:t}^{(j)}\right)^2 \mathbf{1}_{\{|h|\geq C\}}\widetilde{v}_t^{(j)}}{\sum_{j=1}^N \widetilde{v}_t^{(j)}}$$
$$+\mathbf{1}_{\{C\geq\sqrt{N}\epsilon\}}\frac{\sum_{j=1}^N h\left(x_{1:t}^{(j)}\right)^2 \widetilde{v}_t^{(j)}}{\sum_{j=1}^N \widetilde{v}_t^{(j)}}$$
$$\xrightarrow{P} E_{\widetilde{\pi}_t^*}\left[h^2\mathbf{1}_{\{|h|\geq C\}}\right].$$

Since $E_{\widetilde{\pi}_t^*}[h^2] < \infty$ and the above convergence holds for any $C \geq 0$,

$$\sum_{j=1}^N E\left[U_N^{(j)2}\mathbf{1}_{\{|U_N^{(j)}|\geq\epsilon\}} \middle| \left\{\widetilde{x}_{0:t}^{(j)}\right\}_{j=1}^N\right] \xrightarrow{P} 0,$$

and the second condition holds. Then, (A.12) holds. For $\widetilde{A}_N$,

$$\widetilde{A}_N = \frac{N^{-1/2}\sum_{j=1}^N \left(h\left(\widetilde{x}_{1:t}^{(j)}\right) - \widetilde{\mu}_t\right)\widetilde{v}_t^{(j)}}{N^{-1}\sum_{j=1}^N \widetilde{v}_t^{(j)}},$$

where the numerator has the expansion $e_{t-1}^{-1}e_t N^{-1/2}\sum_{j=1}^N(\widetilde{h}_t(x_{0:t}^{(j)}) - \widetilde{\boldsymbol{\beta}}_t^T \boldsymbol{g}_n(x_t^{(j)}|\widetilde{x}_{0:t-1}^{(j)})) + o_p(1)$ similarly to (A.9) and the denominator $N^{-1}\sum_{j=1}^N \widetilde{v}_t^{(j)} \xrightarrow{P} e_{t-1}^{-1}e_t N^{-1/2}$ similarly to (A.7). Then, following the same arguments of proving convergence of (A.9), with $c_t, h_t, \boldsymbol{\beta}_t$, and π_t^* replacing by $e_t, \widetilde{h}_t, \widetilde{\boldsymbol{\beta}}_t$, and $\widetilde{\pi}_t^*$, respectively, for $h \in A_t'$, we have

$$\widetilde{A}_N \xrightarrow{\mathfrak{L}} N\left(0, \sigma_{2,t}^2(h) - \mathrm{var}_{\pi_t^*}[h]\right).$$

Since $\widetilde{A}_N$ is a function of $\{x_{0:t}^{(j)}\}_{j=1}^N$, it can be proved through the characteristic function that $\widetilde{A}_N + \widetilde{B}_N \xrightarrow{\mathfrak{L}} N(0, \sigma_{2,t}^2(h))$ and (A.11) holds. □

For applying Owen and Zhou's regression approach in the estimation of the generalized SMC method, which results in $\widehat{\mu}_{n,\mathrm{Reg}}$ in (7), its asymptotic variance is stated below. Define

$$\sigma_{1,n}'^2(h) = \sigma_{2,n-1}'^2\left(E_{q_{\boldsymbol{\alpha}}}[h_n(x_{0:n})]\right)$$
$$+E_{\widetilde{\pi}_{n-1}^*}\left(\mathrm{var}_{q_{\boldsymbol{\alpha}}}\left[h_n(x_{0:n}) - \boldsymbol{\beta}_n^T\boldsymbol{g}_n(x_n|x_{0:n-1})\right]\right),$$

$$\sigma_{2,n}'^2(h) = \sigma_{2,n-1}'^2\left(E_{q_{\boldsymbol{\alpha}}}\left[\widetilde{h}_n(x_{0:n})\right]\right)$$
$$+E_{\widetilde{\pi}_{n-1}^*}\left(\mathrm{var}_{q_{\boldsymbol{\alpha}}}\left[\widetilde{h}_n(x_{0:n})\right]\right) + \mathrm{var}_{\widetilde{\pi}_n^*}[h].$$

Proof of Corollary A.1. The corollary holds by noting that

$$\sigma_{1,n}^2(h) \leq \sum_{t=1}^n \int \frac{\left[\pi_n^*(x_{0:t})(\mu_t(x_{0:t}) - \mu_n)\right]^2}{\widetilde{\pi}_{t-1}^*(x_{0:t-1})q_{\boldsymbol{\alpha}}(x_t|x_{0:t-1})}dx_{0:t}$$
$$\leq \frac{1}{\alpha_k}\sum_{t=1}^n \int \frac{\left[\pi_n^*(x_{0:t})(\mu_t(x_{0:t}) - \mu_n)\right]^2}{\widetilde{\pi}_{t-1}^*(x_{0:t-1})q_k(x_t|x_{0:t-1})}dx_{0:t}$$

for any $\alpha_k > 0$. □

Proposition 1. Under the same assumption as Theorem 2, for any n, $\sigma_{1,n}'^2(h)$ is finite and the following convergence holds:

$$\sqrt{N}\left(\widehat{\mu}_{n,\mathrm{Reg}} - \mu_n\right) \xrightarrow{\mathfrak{L}} N\left(0, \sigma_{1,n}'^2(h)\right). \quad \text{(A.13)}$$

Specifically,

$$\sigma_{2,n}'^2(h) = \sum_{t=1}^{n-1}\int \frac{\pi_n^*(x_{0:t})(\mu_t(x_{0:t}) - \mu_n)}{\widetilde{\pi}_{t-1}^*(x_{0:t-1})q_{\boldsymbol{\alpha}}(x_t|x_{0:t-1})}dx_{0:t}$$
$$+\int \frac{\left[\pi_n^*(x_{0:n})(h(x_{1:n}) - \mu_n) - \boldsymbol{\beta}_{nn}^T\widetilde{\pi}_{n-1}^*(x_{0:n-1})\boldsymbol{g}(x_n|x_{0:n-1})\right]^2}{\widetilde{\pi}_{n-1}^*(x_{0:n-1})q_{\boldsymbol{\alpha}}(x_n|x_{0:n-1})}dx_{0:n}.$$

Proof. Let $\widetilde{w}_t^{(j)} = \eta(x_{0:t}^{(j)})w_t^{(j)}$. Note that the difference between LM-SMC and RM-SMC is in the resampling step, that the resampling in the former is according to $\{\widetilde{v}_t^{(j)}\}_{j=1}^N$ and in the later is according to $\{\widetilde{w}_t^{(j)}\}_{j=1}^N$. Therefore, the proof can follow the exact same line as the proof of Theorem 2, with difference only in arguments related to the resampling which are the results (A.5) and (A.11). For the proof of (A.5) and the part of $\widetilde{B}_N$ in the proof (A.11), all arguments can go through when $\widetilde{v}_t^{(j)}$ is replaced by $\widetilde{w}_t^{(j)}$ and therefore, the corresponding results still hold. For the part of $\widetilde{A}_N$ in the proof (A.11), when $\widetilde{v}_t^{(j)}$ is replaced by $\widetilde{w}_t^{(j)}$, the numerator of $\widetilde{A}_N$ is $e_{t-1}^{-1}e_t N^{-1/2}\sum_{j=1}^N \widetilde{h}_t(x_{0:t}^{(j)})$ and the denominator is $N^{-1}\sum_{j=1}^N \widetilde{w}_t^{(j)}$. Then, following the similar arguments of proving convergence of (A.9) without control variates $\boldsymbol{g}_n$, for $h \in A_t'$, we have

$$\widetilde{A}_N \xrightarrow{\mathfrak{L}} N\left(0, \sigma_{2,t}'^2(h) - \mathrm{var}_{\pi_t^*}[h]\right).$$

Then all inductive hypotheses hold, which concludes the proof. □

[Received February 2013. Revised December 2014.]

REFERENCES

Cappé, O., Godsill, S., and Moulines, E. (2007), "An Overview of Existing Methods and Recent Advances in Sequential Monte Carlo," in *Proceedings of the IEEE*, 95, 899–924. [298,300]

Carpenter, J., Clifford, P., and Fearnhead, P. (1999), "Improved Particle Filter for Nonlinear Problems," *IEE Proceedings. Radar, Sonar and Navigation*, 146, 2–7. [301,305]

Cérou, F., Del Moral, P., Furon, T., and Guyader, A. (2012), "Sequential Monte Carlo for Rare Event Estimation," *Statistics and Computing*, 22, 795–808. [298]

Chan, H., and Lai, T. (2011), "A Sequential Monte Carlo Approach to Computing Tail Probabilities in Stochastic Models," *The Annals of Applied Probability*, 21, 2315–2342. [298]

Chen, R. (2005), "Sequential Monte Carlo Methods and Their Applications," *IMS Lecture Notes Series, Markov Chain Monte Carlo*, 7, 147–182. [298]

Chen, R., and Liu, J. (2000), "Mixture Kalman Filters," *Journal of the Royal Statistical Society,* Series B, 62, 493–508. [298]

Chen, Y., Diaconis, P., Holmes, S., and Liu, J. (2005), "Sequential Monte Carlo Methods for Statistical Analysis of Tables," *Journal of the American Statistical Association*, 100, 109–120. [298]

Douc, R., and Cappé, O. (2005), "Comparison of Resampling Schemes for Particle Filtering," in *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis*, pp. 64–69. [305]

Douc, R., and Moulines, E. (2008), "Limit Theorems for Weighted Samples With Applications to Sequential Monte Carlo Methods," *Annals of Statistics*, 36, 2344–2376. [300,302,310,311,312]

Douc, R., Moulines, É., and Olsson, J. (2009), "Optimality of the Auxiliary Particle Filter," *Probability and Mathematical Statistics*, 29, 1–28. [301,303,310]

Doucet, A., Briers, M., and Sénécal, S. (2006), "Efficient Block Sampling Strategies for Sequential Monte Carlo Methods," *Journal of Computational and Graphical Statistics*, 15, 693–711. [298]

Doucet, A., Godsill, S., and Andrieu, C. (2000), "On Sequential Monte Carlo Sampling Methods for Bayesian Filtering," *Statistics and Computing*, 10, 197–208. [298]

Doucet, A., and Johansen, A. (2009), "A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later," *Handbook of Nonlinear Filtering*, pp. 656–704. [298]

Elinas, P., Sim, R., and Little, J. (2006), "$\sigma SLAM$: Stereo Vision SLAM Using the Rao-Blackwellised Particle Filter and a Novel Mixture Proposal Distribution," in *Proceedings IEEE International Conference on Robotics and Automation*, pp. 1564–1570. [299,301]

Fox, D., Thrun, S., Burgard, W., and Dellaert, F. (2001), "Particle Filters for Mobile Robot Localization," in *Sequential Monte Carlo Methods in Practice*, New York: Springer, pp. 401–428. [299]

Gilks, W., and Berzuini, C. (2001), "Following a Moving Target-Monte Carlo Inference for Dynamic Bayesian Models," *Journal of The Royal Statistical Society,* Series B, 63, 127–146. [298]

Gordon, N., Salmond, D., and Smith, A. (1993), "Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation," in *IEE Proceedings F Radar and Signal Processing*, (vol. 140), pp. 107–113. [298]

Guo, D., Wang, X., and Chen, R. (2005), "New Sequential Monte Carlo Methods for Nonlinear Dynamic Systems," *Statistics and Computing*, 15, 135–147. [298]

Hesterberg, T. (1995), "Weighted Average Importance Sampling and Defensive Mixture Distributions," *Technometrics*, 37, 185–194. [299,300]

Hjort, N., and Pollard, D. (2011), "Asymptotics for Minimisers of Convex Processes," arXiv Preprint, *arXiv:1107.3806*. [310]

Johansen, A., and Doucet, A. (2008), "A Note on Auxiliary Particle Filters," *Statistics & Probability Letters*, 78, 1498–1504. [300]

Kim, S., Shephard, N., and Chib, S. (1998), "Stochastic Volatility: Likelihood Inference and Comparison With ARCH Models," *The Review of Economic Studies*, 65, 361–393. [307]

Kong, A., McCullagh, P., Meng, X., Nicolae, D., and Tan, Z. (2003), "A Theory of Statistical Models for Monte Carlo Integration," *Journal of the Royal Statistical Society,* Series B, 65, 585–604. [300]

Li, W., Tan, Z., and Chen, R. (2013), "Two-Stage Importance Sampling With Mixture Proposals," *Journal of the American Statistical Association*, 108, 1350–1360. [299,301,304,310]

Lin, M., Chen, R., and Mykland, P. (2010), "On Generating Monte Carlo Samples of Continuous Diffusion Bridges," *Journal of American Statistical Association*, 105, 820–838. [301]

Lin, M., Zhang, J., Cheng, Q., and Chen, R. (2005), "Independent Particle Filters," *Journal of the American Statistical Association*, 100, 1412–1421. [298]

Liu, J. (2008), *Monte Carlo Strategies in Scientific Computing*, New York: Springer-Verlag. [299]

Liu, J., and Chen, R. (1995), "Blind Deconvolution via Sequential Imputations," *Journal of the American Statistical Association*, 90, 567–576. [298]

——— (1998), "Sequential Monte Carlo Methods for Dynamic Systems," *Journal of the American Statistical Association*, 93, 1032. [298,301]

Nocedal, J., and Wright, S. (1999), *Numerical Optimization*, New York: Springer-Verlag. [305]

Oh, M., and Berger, J. (1993), "Integration of Multimodal Functions by Monte Carlo Importance Sampling," *Journal of the American Statistical Association*, 88, 450–456. [299]

Owen, A., and Zhou, Y. (2000), "Safe and Effective Importance Sampling," *Journal of the American Statistical Association*, 95, 135–143. [299,300]

Pitt, M., and Shephard, N. (1999), "Filtering via Simulation: Auxiliary Particle Filters," *Journal of the American Statistical Association*, 94, 590–599. [298,301,305,307]

Robert, C., and Casella, G. (2004), *Monte Carlo Statistical Methods*, New York: Springer-Verlag. [299]

Saha, S., Mandal, P., Boers, Y., Driessen, H., and Bagchi, A. (2009), "Gaussian Proposal Density Using Moment Matching in SMC Methods," *Statistics and Computing*, 19, 203–208. [298]

Sandmann, G., and Koopman, S. (1998), "Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood," *Journal of Econometrics*, 87, 271–301. [307,308]

Shephard, N. (2005), *Stochastic Volatility: Selected Readings, Advanced Texts in Econometrics Series*. Oxford, UK: Oxford University Press. [298]

Singh, S., Kantas, N., Vo, B., Doucet, A., and Evans, R. (2007), "Simulation-Based Optimal Sensor Scheduling With Application to Observer Trajectory Planning," *Automatica*, 43, 817–830. [299]

Singh, S., Vo, B., Doucet, A., and Evans, R. (2004), "Variance Reduction for Monte Carlo Implementation of Adaptive Sensor Management," in *Proceedings of the International Conference on Information Fusion*, pp. 901–908. [299,301]

Smith, J., and Santos, A. (2006), "Second-Order Filter Distribution Approximations for Financial Time Series With Extreme Outliers," *Journal of Business and Economic Statistics*, 24, 329–337. [308]

Tan, Z. (2004), "On a Likelihood Approach for Monte Carlo Integration," *Journal of the American Statistical Association*, 99, 1027–1036. [298,299,300,301,303,304,310]

Thrun, S., Fox, D., Burgard, W., and Dellaert, F. (2001), "Robust Monte Carlo Localization for Mobile Robots," *Artificial Intelligence*, 128, 99–141. [301]

Veach, E., and Guibas, L. (1995), "Optimally Combining Sampling Techniques for Monte Carlo Rendering," in *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, pp. 419–428. [299]

Wang, X., Chen, R., and Guo, D. (2002), "Delayed-Pilot Sampling for Mixture Kalman Filter With Application in Fading Channels," *IEEE Transactions on Signal Processing*, 50, 241–254. [298]

Zhang, J., Chen, Y., Chen, R., and Liang, J. (2004), "Importance of Chirality and Reduced Flexibility of Protein Side Chains: A Study With Square and Tetrahedral Lattice Models," *Journal of Chemical Physics*, 121, 592–603. [301]